



UNIVERSITÀ
DEGLI STUDI DELLA
Tuscia

**DIPARTIMENTO DI SCIENZE AGRARIE E FORESTALI
(DAFNE)**

Corso di Dottorato di Ricerca in

Scienze delle Produzioni Vegetali e Animali - XXIX Ciclo

**APPLICAZIONE SPAZIALIZZATA DI METODI DI ASSIMILAZIONE DI
DATI OTTENUTI DA TELERILEVAMENTO IN MODELLI DI CRESCITA
DELLE COLTURE PER LA STIMA DELLA PRODUZIONE DI FRUMENTO
TENERO**

Tesi di dottorato di:

Dott. Paolo Cosmo, SILVESTRO

Coordinatore del corso

Prof.ssa Stefania MASCI

Tutore

Prof. Raffaele CASA

Co-tutore

Dott. Stefano Pignatti

A.A. 2016/17

Indice

Abstract	5
Riassunto	7
1. Introduzione	9
1.1. Stato dell'arte ed obiettivi	9
1.2. Panoramica sui contenuti della tesi	17
2. Concetti di base	20
2.1. Modelli di accrescimento delle colture	20
2.2. Analisi di sensibilità (Sensitivity Analysis)	24
2.3. Assimilazione dati	26
2.4. Principi di Telerilevamento	28
3. Aree di studio e descrizione dei dati	33
3.1. Area di studio e dataset climatici	33
3.2. Dati in situ e statistici	36
3.3. Dati da telerilevamento	38
3.3.1. Landsat 8	38
3.3.2. Huan Jing-1	39
3.3.3. Pre-elaborazione delle immagini	40
4. Stima di variabili biofisiche dal telerilevamento	42
4.1. Stato dell'arte delle metodologie di stima delle variabili biofisiche delle colture per mezzo di dati ottenuti da telerilevamento	42
4.2. Algoritmo di stima delle variabili biofisiche PROSAIL-ANN	45
4.2.1. Generazione del database di addestramento	46
4.2.2. Addestramento delle Reti Neurali Artificiali (ANN training)	48
4.2.3. Applicazione delle Reti Neurali Artificiali	49
4.2.4. Risultati e discussione	49
5. Modelli di accrescimento delle colture	53
5.1. Aquacrop	53
5.1.1. Variabili di Input del modello Aquacrop	55
5.2. Simple Algorithm For Yield (SAFY)	59
6. Analisi di sensibilità	62
6.1. Introduzione	62
6.2. Metodi di Global Sensitivity Analysis (GSA) utilizzati	64
6.2.1. Metodo di Morris	65
6.2.2. Extended Fourier Amplitude Sensitivity Test (EFAST)	68
6.2.3. Intervalli di variazione dei parametri	71
6.2.4. Indici per il confronto tra i modelli	76
6.3. Risultati della SA	77
6.3.1. Applicazione del metodo di Morris	77
6.3.2. Applicazione del metodo EFAST	81
6.3.3. Plasticità e complessità di Aquacrop e SAFY	83
6.4. Discussione sulla GSA dei modelli Aquacrop e SAFY	84
7. Metodi di assimilazione dati in modelli di accrescimento delle colture	90
7.1. Metodo di assimilazione EnKF usato per SAFY	90
7.2. Metodo di assimilazione PSO usato per Aquacrop	94
7.3. Applicazione puntuale e spazializzata dei metodi di assimilazione	97
8. Risultati: applicazione dei metodi di assimilazione	100
8.1. Applicazione puntuale dei metodi di assimilazione	100

8.2.	Applicazione spazializzata dei metodi di assimilazione	103
8.3.	Discussione dei risultati	106
9.	Conclusioni	109
	Ringraziamenti	115
	Bibliografia	116
	Allegato A	125
	Ciclo Colturale 2012-2013	126
	Ciclo Colturale 2013-2014	130
	Ciclo Colturale 2014-2015	137
	Allegato B	141
	Articoli sottoposti a riviste con impact factor	141
	Allegato C	142
	Articoli pubblicati (atti di convegni).....	142

Abstract

The use of remote sensing data has been shown to be very useful for field management, crop estimation and yield prediction, both for field scale applications, e.g. in precision agriculture, and at the regional scale. The development of new technologies and the availability of more detailed satellite data allow the increasing diffusion of these techniques. The aim of this work is to develop and compare methods for estimating the yield of winter wheat, at the field and regional scale, using a limited number of images acquired from remote sensing. The methodology adopted consists in the estimation of crop biophysical variables from data acquired from multispectral sensors and their assimilation into a crop growth model, in order to improve the yield estimation. The sensors used are the OLI on board of the US satellite Landsat 8 and the CCD on board of the Chinese satellite HJ1A/B, two sensors with a spatial resolution of 30 m and similar technical characteristics. Two growth crop models are used: Aquacrop and the Simplex Algorithm for yield (SAFY). The first is a water-driven crop model, with a high accuracy for estimating yield response to water, but driven by a large number of input variables and parameters, while the second is a radiation-driven simple model, using a reduced number of input variables and parameters. These two models are compared, by quantifying their complexity and plasticity, through a Global Sensitivity Analysis (GSA) for the climatic and environmental data of the test site. A combination of two different GSA techniques are used: Morris and the Extended Fourier Amplitude Sensitivity Test (EFAST), in order to exploit the fast computation of Morris to identify all the non-influential parameters and the high accuracy of EFAST for ranking the most important parameters. Because of the differences between the two crop models, two different assimilation methods were applied, a parameter calibration method based on the Particle Swarm Optimization (PSO) for Aquacrop and an updating assimilation method based on the Ensemble Kalman Filter (EnKF) for SAFY. Leaf area index (LAI) and canopy cover (CC) are estimated from satellite data using a method based on the training of Artificial Neural Network (ANN) on the simulations of PROSAIL, a radiative transfer model, allowing to estimate also the error. The variables were then assimilated into the crop models, both in point and spatialized applications, respectively at the field and regional scale, by using data collected in the rural area of Yangling (Shaanxi Province, China) between 2013 and 2015. Results were validated

by utilizing yield data both from ground measurements and statistical survey. The point application of the EnKF coupled with SAFY allowed to obtain a relative root mean squared error (RRMSE) between observed and simulated yield of 18%, and the PSO method used with Aquacrop allowed to obtain a RRMSE of 24 %.

This study shows the possibility to estimate winter wheat production coupling RS data and crop growth models, highlighting the strengths and weaknesses of two different crop growth model and the relative assimilation procedures.

Riassunto

L'applicazione del telerilevamento in agricoltura ha dimostrato di essere molto utile per la gestione dei campi, la stima delle colture e la previsione della produzione, sia per applicazioni a scala di campo, es. in agricoltura di precisione, che per applicazioni a scala regionale. Lo sviluppo di nuove tecnologie e la crescente disponibilità di dati satellitari a basso costo, consentono una più maggiore diffusione di queste tecniche.

L'obiettivo di questo lavoro è stato quello di sviluppare un metodo per la stima della produzione di frumento tenero, sia su scala puntuale di campo che su scala regionale, usando un numero limitato di immagini satellitari acquisite da telerilevamento. La metodologia prevede la stima di variabili biofisiche delle colture da dati acquisiti da sensori multispettrali e l'assimilazione di questi dati in un modello di accrescimento della coltura, in modo da migliorare la stima della produzione. I sensori usati sono stati OLI, a bordo del satellite statunitense Landsat 8 e il CCD a bordo dei satelliti della costellazione cinese Huang-Jing1, due sensori con una risoluzione spaziale di 30 m e caratteristiche tecniche simili. Sono stati utilizzati due modelli di accrescimento delle colture: Aquacrop e Simplex Algorithm for Yield (SAFY). Il primo è un modello che mette in relazione la produttività con lo stato idrico della coltura, con un'elevata accuratezza nella stima della risposta della coltura all'acqua, ma dipendente da un grande numero di variabili di input e parametri, mentre il secondo è basato sulla conversione della radiazione solare assorbita in biomassa, ed usa un numero ridotto di variabili di input e parametri. Questi due modelli sono stati confrontati quantificando la loro complessità e plasticità per mezzo di un'analisi di sensibilità globale (GSA) per i dataset climatici e ambientali di siti test. È stata usata una combinazione di differenti tecniche di GSA: il metodo di Morris e l'Extended Fourier Amplitude Sensitivity Test (EFAST). La rapidità di calcolo di Morris è stata utilizzata per identificare tutti i parametri non influenti, e l'elevata accuratezza di EFAST per classificare i parametri in ordine di influenza sull'output. A causa delle differenze tra i 2 modelli di accrescimento delle colture, sono state applicati due diversi metodi di assimilazione, un metodo di ricalibrazione dei parametri basato sul Particle Swarm Optimization applicato al modello Aquacrop e un metodo di 'updating' (aggiornamento) basato sul Ensemble Kalman Filter (EnKF) applicato al modello SAFY. L'indice di area fogliare (LAI) e la percentuale di copertura del suolo da parte

della coltura (CC) sono stati stimati dalle immagini satellitari usando un metodo basato sull'addestramento di una rete neurale artificiale (ANN) per mezzo di simulazioni effettuate con PROSAIL, un modello di trasferimento radiativo, consentendo così di quantificare anche l'errore sulle stime. Queste procedure sono state testate in applicazioni puntuali su scala di campo e spazializzate su scala regionale, usando dati raccolti nell'area rurale di Yangling (Provincia di Shaanxi, Cina) tra il 2013 ed il 2015. I risultati sono stati validati usando dati di produzione sia raccolti da misure in campo che da forniti da enti ufficiali di rilevamento statistico.

L'applicazione puntuale del metodo EnKF accoppiato con il modello SAFY ha permesso di ottenere un errore relativo (RRMSE) tra produzione di granella osservata e simulata del 18%, mentre il metodo PSO, utilizzato con Aquacrop ha consentito di ottenere un RMSE del 24%.

Questo studio mostra la possibilità di stimare la produzione di frumento tenero combinando dati satellitari e modelli di accrescimento della coltura, evidenziando alcuni punti di forza e debolezza dei due differenti modelli di accrescimento e dei relativi metodi di assimilazione.

1. Introduzione

1.1. Stato dell'arte ed obiettivi

Più di un quarto della superficie terrestre globale è coperta da agroecosistemi (Coleman et Hendrix, 1988). Generalmente si definisce agroecosistema un ecosistema gestito dall'uomo con la finalità di produrre cibo o altro materiale utile all'uomo (Elliott et Edward, 1989).

La crescente domanda di prodotti agricoli degli ultimi 50 anni ha incentivato a ricercare metodi che aumentassero la produzione, comportando un cambiamento significativo dei tradizionali metodi di gestione degli agroecosistemi (Tilman, 1999). Questi cambiamenti consistono principalmente nell'uso intensivo di prodotti agrochimici e macchinari agricoli, nell'espansione delle aree d'irrigazione e nella diffusione di nuove varietà. Cambiamenti nelle strategie di gestione sono stati adottati in molte parti del mondo e, di conseguenza, in quelle aree la produzione di cibo è aumentata più rapidamente della popolazione umana (Dorigo et al., 2007).

Da un lato le nuove strategie di gestione hanno contribuito sostanzialmente ad aumentare il benessere umano e hanno sostenuto lo sviluppo economico di molte regioni in tutto il mondo, dall'altro hanno talvolta comportato un impatto significativo sull'ambiente (Gardner, 2005). L'erosione dei suoli agricoli, l'inquinamento e l'eccessivo sfruttamento dell'acqua, la perdita di biodiversità e la diffusione di resistenze delle infestanti e dei patogeni, sono alcuni dei problemi che possono essere imputati ad un'inopportuna gestione degli agroecosistemi (Tilman, 1999).

La conoscenza dei processi responsabili della degradazione degli agroecosistemi induce lo sviluppo di nuove strategie di gestione capaci di interagire con la loro complessità in modo da massimizzare le produzioni minimizzando gli impatti ambientali. In questo contesto trova largo spazio lo sviluppo di modelli che descrivono adeguatamente le risposte degli agroecosistemi a cambiamenti di strategie di gestione oppure di condizioni climatiche e ambientali.

La sviluppo di modelli che descrivono agroecosistemi è iniziata negli anni '60 con lo sviluppo e la diffusione dei computer (Passioura, 1996). Da quel momento sono stati sviluppati molti modelli che simulano l'accrescimento delle colture, per diverse tipologie di colture, operanti a differenti scale spaziali e temporali e con per una grande diversità di applicazioni (Tab. 1). Una rassegna dettagliata sui modelli di accrescimento delle colture è fornita da Rauff e Bello (2015). Nel presente lavoro sono

stati utilizzati 2 differenti modelli di accrescimento delle colture, descritti più approfonditamente nel capitolo 5.

Tab. 1.1: Potenzialità dei modelli di sistemi di accrescimento delle colture (adattato da Dorigo et al., 2007).

Uso primario del modello	Applicazioni principali
Strumento di ricerca	Sintetizzare la comprensione della ricerca Integrare conoscenze interdisciplinari Documentare esperimenti Assistere in miglioramenti genetici Analisi di produzione e di fattori rilevanti
Gestione dei sistemi colturali	Assistere nella gestione delle colture Assistere nella gestione dell'irrigazione e della fertilizzazione Agricoltura di Precisione
Analisi delle politiche	Assistenza per una gestione ridotta di pesticidi e fertilizzanti Previsione di Produzione Valutare i cambiamenti climatici

I modelli di accrescimento delle colture sono stati largamente usati per descrivere l'impatto delle condizioni climatiche e delle strategie di gestione a scala di campo per poi essere applicati in modo distribuito su scala regionale. I problemi principali con i modelli possono essere: un'eccessiva semplificazione dei sistemi naturali, una parametrizzazione e un'elevata incertezza ed accuratezza negli output. Questi problemi sono particolarmente evidenti su scala regionale, dove le informazioni sono raccolte da punti sparsi sull'area d'interesse (de Wit et al, 2005). Le condizioni al contorno (suolo, gestione, ecc.) sono spesso poco conosciute ed i parametri sono stimati da pochi dati sperimentali.

Una soluzione a questi problemi è l'impiego del telerilevamento (RS, dal termine inglese Remote Sensing), il cui contributo offre informazioni spazializzate sulle variabili biofisiche e biochimiche delle colture. L'uso combinato dei modelli di accrescimento delle colture con tecniche che, per mezzo del telerilevamento, determinano le variabili di stato di tali modelli ha dimostrato di migliorare le prestazioni dei modelli, particolarmente su scala regionale (Launay et Guerif, 2005). Il monitoraggio delle colture si è dimostrato molto efficace per migliorare la gestione dei campi, nella stima della biomassa e nella previsione della produzione (Dong et al.,

2016). L'uso del telerilevamento con lo scopo di monitorare la crescita delle colture sta aumentando con lo sviluppo di nuove tecnologie e con la sempre più semplice reperibilità d'immagini di buona qualità a costi sempre più accessibili, consentendo di acquisire informazioni in modo sempre più efficace (Mulla, 2013). In modo particolare il telerilevamento da satellite è sempre più usato nel monitoraggio delle colture e nell'acquisizione delle variabili che ne descrivono la crescita, poiché consente di monitorare in modo rapido vaste zone e al contempo, laddove la risoluzione del sensore usato lo consenta, dare informazioni su aree più piccole a scala di campo (Atzberger, 2013). L'uso di questa tecnologia si è rivelato molto efficace nella stima della produzione delle colture (Kalpana et al., 2003; Lobell, 2013), anche attraverso procedure che combinano il monitoraggio da telerilevamento delle variabili biofisiche delle colture con i modelli di accrescimento.

Tra le variabili biofisiche delle colture più usate nel campo del RS combinato con i modelli di accrescimento ci sono senz'altro: l'indice di area fogliare, o Leaf Area Index (LAI), la percentuale di copertura del suolo, o Canopy Cover (CC), la frazione di radiazione fotosinteticamente attiva assorbita dalla coltura (fAPAR) e la concentrazione di clorofilla nella vegetazione (Cab).

Una review dettagliata dei metodi per ricavare le variabili di stato delle colture da RS è stata fatta da Dorigo et al. (2007). Tra i metodi riportati in letteratura, di particolare interesse è quello basato sulle reti neurali artificiali (ANN, dall'inglese Artificial Neural Network) (Bacour et al., 2006), che consente di calcolare sia il LAI che il CC in modo indipendente da calibrazioni (Baret et al., 2006). La strategia di questo metodo è quella di addestrare una rete neurale artificiale (ANN) usando le simulazioni fatte con PROSAIL, un modello di trasferimento radiativo che accoppia i modelli PROSPECT e SAIL (Jacquemoud et al., 2009). Molti studi hanno dimostrato l'efficacia del metodo ANN, usando sensori sia multispettrali (Baret et al., 2007) che iperspettrali (Verger et al., 2011). Questo metodo, basato sulle reti neurali artificiali (Bacour et al., 2006), è stato impiegato in questa ricerca per stimare i valori di LAI e CC utilizzando dati ottenuti tramite sensori satellitari multispettrali.

Le variabili biofisiche ricavate da RS possono essere assimilate nei modelli di accrescimento delle colture per stimare la produzione, usando differenti approcci (vedi paragrafo 2.3). Seguendo la classificazione dei metodi di assimilazione fatta da Dorigo et al. (2007), è possibile distinguere due strategie: calibrazione e aggiornamento (updating). La teoria alla base del metodo di assimilazione per calibrazione si basa

sull'uso di algoritmi di ottimizzazione con la ricalibrazione dei parametri in modo da ottenere un andamento della variabile di stato simulata (es. il LAI) il più vicino possibile all'andamento osservato di tale variabile. Nel metodo di assimilazione per updating invece, il valore della variabile di stato simulata è rimpiazzato con quello stimato da RS. Vi sono nella letteratura diversi esempi di applicazione delle tecniche di assimilazione in ambito agricolo. Uno studio condotto da Doraiswamy et al. (2003) mostra un metodo per integrare simulazioni di accrescimento delle colture con dati RS per monitorare la crescita e la produzione di cereali su scala regionale. Tale metodo consiste nel convertire il LAI simulato dal modello EPIC in riflettanza, per le lunghezze d'onda delle bande del visibile e del vicino infrarosso (NIR) del sensore Landsat TM, per mezzo del modello di trasferimento radiativo SAIL (Scattered by Arbitrary Inclined Leaves). I valori di NDVI calcolati usando la riflettanza simulata da SAIL sono poi confrontati con i valori di NDVI derivati dalle immagini satellitari. Il modello di accrescimento delle colture EPIC è ricalibrato affinché la differenza tra NDVI simulato e misurato non fosse superiore al 20% del NDVI simulato.

Uno studio condotto da Dente et al. (2008) ha assimilato il LAI ricavato tramite ENVIOSAT – ASAR (sensore radar in banda C con risoluzione spaziale di 300 m.) e MERIS (sensore multispettrale, visibile-infrarosso con risoluzione spaziale di circa 260 m.) nel modello CERES-wheat, per mezzo di un metodo di assimilazione per calibrazione basato su un algoritmo variazionale (Dente et al. 2008), col fine di stimare su scala regionale la produzione di frumento tenero. I risultati di questo studio mostrano che tramite tale metodologia è possibile ottenere una stima della produzione con un'accuratezza compresa tra 0.36 e 0.42 t·ha⁻¹, che corrisponde ad un errore relativo (RRMSE) di circa il 12%.

Un altro esempio di assimilazione per calibrazione è lo studio presentato da Launay et Guerif (2005), in cui i parametri del modello SUCROS sono ricalibrati minimizzando la differenza tra il LAI stimato tramite il sensore SPOT e quello simulato dal modello. I risultati presentati in questo lavoro mostrano che questo metodo, applicato su un'area vasta ed eterogenea, migliora le stime di produzione rispetto allo stesso modello usato senza l'assimilazione. In questo metodo non tutti i parametri sono stati ricalibrati, pertanto la calibrazione dei parametri che non variano è un punto fondamentale. L'alto numero di parametri e le limitate informazioni su di essi ottenibili *in situ* complicano la calibrazione ed aumentano l'incertezza dei risultati (Launay et Guerif, 2005). Il problema della calibrazione dei parametri è molto comune

nei metodi di assimilazione, come ha evidenziato lo studio condotto da De Wit et al. (2012). In tale studio sono stati utilizzati scenari con differenti condizioni climatiche e ambientali. Il Green Area Index (GAI) osservato dal sensore MODIS è stato assimilato nel modello WOFOST (van Diepen et al., 1989) usando un approccio di calibrazione basato sull'ottimizzazione dei parametri, minimizzando il Weighted Mean Absolute Error tra GAI simulato dal modello e quello stimato tramite MODIS. Per semplificare l'algoritmo sono stati ottimizzati esclusivamente i parametri che più influenzano il GAI, ciò ha evidenziato la necessità di calcolare la sensibilità dei parametri per ogni scenario utilizzato nello studio (De Wit et al., 2012).

Anche l'approccio di assimilazione "updating" è stato testato in molti studi (e.g. Wu et al, 2012; Ines et al., 2013; Huang et al, 2013; Jiang et al. 2014; Dong et al. 2016), in cui è stata provata l'efficacia dei metodi di assimilazione nel migliorare l'accuratezza dei modelli di accrescimento delle colture nello stimare la produzione.

Wu et al. (2012), ad esempio, ha testato l'Ensemble Kalman Filter (EnKF) per assimilare nel modello di accrescimento delle colture WOFOST il LAI derivato da MODIS, in modo da determinare su scala regionale la produzione di frumento. Questo studio ha mostrato il potenziale del metodo di assimilazione EnKF nella stima della produzione su scala regionale.

Ines et al. (2013) ha esplorato le potenzialità dell'EnKF assimilando simultaneamente due variabili di stato, l'umidità del suolo ricavata da dati RS e il LAI, dimostrando la possibilità di prevedere le perdite di produzione monitorando gli effetti dello stress idrico sulla coltura.

Il potenziale dell'EnKF è stato comparato con un altro metodo di assimilazione per updating, il POD4VAR, un nuovo algoritmo quadrimensionale variazionale (Jiang et al., 2014). In questo lavoro gli esperimenti sono stati condotti sia su scala regionale sia su scala di campo, assimilando nel modello CERES il LAI stimato da dati acquisiti da RS, per mezzo del sensore CCD a bordo della costellazione cinese di satelliti Huanjin (HJ). Il POD4VAR ha il vantaggio di avere un'alta efficienza computazionale nelle operazioni di monitoraggio della produzione di cereali su scala regionale, comparato con il metodo EnKF. La limitazione è che richiede un maggiore numero di osservazioni ed è maggiormente influenzato dall'incertezza sulle misure rispetto al metodo EnKF. Un miglioramento del metodo EnKF è stato testato da Huang et al. (2013), il metodo Ensemble Square Root Filter (EnSRF). Il principale vantaggio dell'EnSRF è che può analizzare i valori delle variabili di stato, come il LAI e il Leaf

Nitrogen Accumulation (LNA), per ogni immagine telerilevata e successivamente aggiornare il processo senza la necessità di aggiungere ulteriori errori sulle misure. Questo algoritmo dà un alto peso all'errore del modello e non dà molta importanza agli errori sulle misure: in questo caso le misure corrispondono alle stime fatte da telerilevamento.

L'efficienza dei metodi di assimilazione per stimare la produzione è strettamente legata al numero di osservazioni disponibili (Dorigo et al, 2007). Molti studi in letteratura (e.g. al Wu et al., 2012; de Wit et al., 2012; Ines et al., 2013; Huang et al., 2013) mostrano che ad un alto numero di osservazioni nel corso della stagione, corrisponde una migliore accuratezza delle stime, infatti il numero di assimilazioni è generalmente superiore a 6, e in alcuni casi superiore a 20 (Wu et al., 2012; Ines et al., 2013, Jiang et al., 2014). I metodi che richiedono un elevato numero di immagini trovano maggiore spazio in applicazioni a scala regionale, poiché sono a disposizione numerosi sensori (e.g. i satelliti MODIS e SPOT) che forniscono gratuitamente immagini a bassa risoluzione spaziale (inferiore a 250 m) con un'alta risoluzione temporale (frequenza 1-3 giorni). Per applicazioni a scala di campo è necessario disporre di immagini con risoluzione spaziale di almeno 30 m. Le opportunità di ottenere un elevato numero di dati ad alta risoluzione (sia temporale che spaziale) sono tuttora limitate dalla frequenza di passaggio dei satelliti e dalla presenza di nuvole, pertanto è importante sperimentare metodologie di assimilazione a scala di campo utilizzabili in caso di limitata disponibilità di immagini ad alta risoluzione.

Dong et al. (2016), ad esempio, ha provato a risolvere questo problema usando una tecnica di data fusion, combinando i dati ottenuti da MODIS e da Landsat 8, il primo a maggiore risoluzione temporale ed il secondo a maggiore risoluzione spaziale, per stimare il LAI a scala di campo ed assimilare i dati nel modello SAFY (Simple Algorithm For Yield) sviluppato da Duchemin et al. (2008). Il dataset di "fusione" ha la risoluzione temporale di MODIS e quella spaziale di OLI (il sensore a bordo di Landsat 8). Tale lavoro dimostra il potenziale dell'uso di algoritmi di fusione, sia su scala regionale che di campo, per potenziare il monitoraggio dell'accrescimento delle colture la stima della produzione quando il numero di immagini ad alta risoluzione è ridotto (Dong et al., 2016). Tale problematica non è stata riscontrata nei metodi di assimilazione per calibrazione, i quali generalmente utilizzano tra le 3 e le 6 immagini (Doraiswamy et al., 2003; Launay et Guerif, 2005; De Wit et al., 2012).

Gli studi precedentemente descritti usano differenti modelli di accrescimento delle colture a seconda dell'obiettivo che si intende perseguire. Generalmente, i modelli che descrivono più dettagliatamente il processo di accrescimento delle colture hanno un'accuratezza di stima maggiore, ma al contempo risultano essere più complessi, più difficili da integrare con altri processi (come i metodi di assimilazione) e con un costo computazionale più alto. Inoltre l'elevato numero di variabili di input e parametri rende più complessa la calibrazione. La stima dei parametri che pesano maggiormente sull'output del modello, problema comune a tutti i metodi di assimilazione, è particolarmente problematica in applicazioni a scala regionale. L'individuazione dei parametri che maggiormente influenzano il modello è di fondamentale importanza per la comprensione del modello, per semplificare e migliorare la calibrazione e per l'applicazione di qualsiasi metodo di assimilazione. Pertanto, particolare attenzione è stata dedicata a questo aspetto nel presente lavoro (Capitolo 6).

L'assimilazione dei dati consente l'applicazione spazializzata dei modelli di accrescimento delle colture pur non disponendo di informazioni accurate su tutti i parametri e variabili di input. L'assimilazione consente inoltre di migliorare l'accuratezza, definita come l'abilità del modello nel riprodurre i dati misurati di riferimento, come ad esempio la produzione, la biomassa, il LAI o la percentuale di copertura (Fox, 1981).

La scelta del modello da usare, in accordo con l'obiettivo da raggiungere, non può considerare esclusivamente l'accuratezza del modello, ma deve necessariamente prendere in considerazione anche altre proprietà (Confalonieri et al., 2012).

Confalonieri et al. (2012) hanno comparato diversi modelli di accrescimento delle colture usando, oltre all'accuratezza, due indici che descrivono differenti criteri di valutazione dei modelli: complessità e plasticità. La complessità è la parsimonia del modello nel rappresentare il sistema biofisico, e quantifica la rilevanza dei parametri del modello in funzione dell'output (Confalonieri et al., 2009). La plasticità è la tendenza del modello a cambiare il suo comportamento al variare delle condizioni in cui è applicato (Confalonieri et al. 2012). L'accuratezza, la complessità e la plasticità forniscono un'esaustiva descrizione dei modelli, consentendo di scegliere il modello che meglio soddisfa le necessità del lavoro da svolgere.

Alla luce di quanto riscontrato dall'analisi della bibliografia finora descritta, la ricerca effettuata in questo lavoro ha come principale obiettivo lo sviluppo di una metodologia basata sulla combinazione delle stime delle grandezze biofisiche

effettuate da telerilevamento con i modelli di accrescimento delle colture al fine di migliorare la stima della produzione per applicazioni puntuali e spazializzate.

Il presente studio vuole proseguire l'analisi, incontrata nella letteratura sopradescritta, delle metodologie legate all'assimilazione dei dati ottenuti da telerilevamento nei modelli di accrescimento delle colture. Si prefigge l'obiettivo di approfondire sia l'utilizzo di metodi di assimilazione per calibrazione sia di quelli per aggiornamento, analizzando accoppiamenti "modello-metodo di assimilazione" non riscontrati finora in letteratura. L'impiego di due modelli con differenti caratteristiche ha avuto l'obiettivo di metterne in risalto i punti di forza e di debolezza dei due modelli, cercando di capire quale fosse più adatto all'impiego nei metodi di assimilazione.

Sono stati utilizzati due modelli differenti: Aquacrop (Steduto et al., 2009), che include un elevato numero di parametri, ma che consente di ottenere accurate stime della produzione (Jin et al., 2014), e SAFY, un semplice modello di stima della produzione colturale, che ha una minore accuratezza del precedente, ma include un ridotto numero di parametri (Duchemin et al. 2008).

Questi modelli sono stati preventivamente analizzati usando metodi di analisi di sensibilità in modo da identificare i parametri ai quali risulta più sensibile la produzione colturale. L'analisi di sensibilità ha permesso di quantificare la plasticità e complessità dei due modelli, indici utilizzati per confrontarne il comportamento dei per alcuni scenari specifici. L'analisi di sensibilità, il confronto tra i modelli SAFY e Aquacrop e le metodologie usate per compierlo sono dettagliatamente descritte nel capitolo 6.

Ne è risultato che il modello SAFY, pur avendo un'accuratezza inferiore ad Aquacrop, per merito della sua semplicità si presta meglio ad interagire con i metodi di assimilazione. Il modello Aquacrop si è dimostrato più accurato, ma meno adatto ad interagire con dati esterni per mezzo di metodi di assimilazione. Nonostante la difficoltà di interazione di Aquacrop con i metodi di assimilazione si è proseguito nel suo impiego proprio in virtù della sua accuratezza.

Per ogni modello è stato utilizzato un differente metodo di assimilazione: per SAFY è stato utilizzato un metodo di assimilazione updating basato sull'Ensemble Kalman Filter, per Aquacrop è stato usato un metodo di assimilazione per calibrazione basato sul Particle Swarm Optimization (PSO) (Eberharth and Kennedy, 1995). In questo modo è stato possibile un confronto tra metodi di assimilazione per updating e per calibrazione utilizzando gli stessi scenari (condizioni climatiche ed ambientali) e

gli stessi dati acquisiti sia da RS che *in situ*, con l'intento di mettere in risalto i vantaggi e gli svantaggi delle due metodologie di assimilazione. L'insieme dei dati satellitari è stato acquisito usando due sensori simili: OLI, a bordo del satellite Landsat 8 e CCD a bordo della costellazione Huanjin (HJ). Tale dataset è stato utilizzato per stimare il LAI (variabile di stato usata nel modello SAFY) e il CC (variabile usata in Aquacrop), mediante un metodo basato su reti neurali artificiali (Bacour et al., 2006), sfruttando la capacità di questo metodo di stimare con buona accuratezza sia LAI che CC.

Le stime di produzione su scala di campo sono state validate con misure di effettuate nel sito di Yangling (Cina) tra il 2013 ed il 2015, mentre i risultati su scala regionale sono stati validati solo per il 2013 ed il 2014, usando informazioni sulla produzione fornite da due istituti ufficiali di statistica: il National Bureau of Statistics della Repubblica Popolare Cinese (che ha fornito i dati del 2013) e lo Shaanxi Provincial Bureau of Statistics (che ha fornito i dati del 2014).

Il lavoro svolto in questo Dottorato di Ricerca ed illustrato nella presente tesi è stato effettuato nell'ambito del programma ESA-MOST Dragon3, di cooperazione scientifica tra la European Space Agency (ESA) ed il Ministry of Science and Technology (MOST) della Repubblica Popolare Cinese. Tale programma si focalizza sull'utilizzo dei dati di osservazione della Terra forniti da ESA e dal National Remote Sensing Center of China (NRSCC) per favorire lo sviluppo di applicazioni in ambito terrestre, oceanico ed atmosferico. Più nello specifico il lavoro è inserito nell'ambito del progetto "Farmland Drought" svolto in collaborazione tra l'Università della Tuscia, l'Istituto di Metodologie per l'Analisi Ambientale del Consiglio Nazionale delle Ricerche (IMAA – CNR) ed il National Engineering Research Center for Information Technology in Agriculture (NERCITA) di Pechino.

1.2. Panoramica sui contenuti della tesi

La tesi è composta di sette capitoli principali, più le conclusioni finali.

Il capitolo 1 fornisce un'introduzione sulle motivazioni alla base dell'uso dei modelli nella gestione degli agroecosistemi, descrivendone in generale le caratteristiche e le principali applicazioni. Introduce inoltre l'uso combinato di dati acquisiti da telerilevamento con i modelli di accrescimento delle colture al fine di migliorare la stima della produzione, fornendo una panoramica degli studi presenti in letteratura e delle tecniche di assimilazione. Conclude presentando in che modo questa tesi prosegue lo studio nell'ambito dell'assimilazione dei dati ottenuti da

telerilevamento nei modelli di accrescimento delle colture e quali obiettivi si prefigge di raggiungere.

Nel capitolo 2 vengono introdotte le nozioni di base sugli argomenti trattati.

Il capitolo è suddiviso in 4 sezioni, nella prima delle quali sono introdotti i modelli di accrescimento delle colture ed i principali indici utilizzati per quantificare la concordanza tra le osservazioni *in situ* e le simulazioni del modello. Nella seconda sezione è fornita una panoramica generale sull'importanza dell'analisi di sensibilità (SA) e sui principali metodi di SA. Nella terza sezione è discussa l'importanza dell'assimilazione dei dati ed una macro classificazione dei metodi di assimilazione. Infine nella quarta sezione sono esposti i principi basilari del telerilevamento ed i principali indici di vegetazione, utilizzati in questo lavoro.

Nel capitolo 3 sono descritte le aree di studio sperimentali in cui sono stati raccolti i dati *in situ* per la validazione puntuale dei metodi di assimilazione sviluppati. Sono inoltre descritte le modalità di raccolta dei dati e le caratteristiche dei dati statistici reperiti. Sono inoltre presentati i dati RS utilizzati e brevemente descritte le caratteristiche dei sensori impiegati per l'acquisizione dei dati.

Il capitolo 4 è dedicato alla descrizione della stima del LAI e del CC utilizzando dati RS. Introduce le varie metodologie esistenti per stimare da RS differenti grandezze biofisiche della vegetazione, esplicitando le motivazioni che hanno portato alla scelta del metodo PROSAIL-ANN. È successivamente descritta la metodologia dell'algoritmo utilizzato, ed infine sono presentati e discussi i risultati dell'applicazione di tale metodi ai dati RS a disposizione, validati con i dati *in situ* descritti nel capitolo 3.

Il capitolo 5 presenta una descrizione dettagliata dei modelli di accrescimento delle colture utilizzati, Aquacrop e SAFY.

Nel capitolo 6 sono riportate le metodologie utilizzate per lo studio di sensibilità applicato ai modelli descritti nel capitolo 5. Inizialmente è riportata un'introduzione alla SA, nella quale è esplicitata l'importanza di applicare tale analisi, fornendo una panoramica generale sulle metodologie esistenti e sugli studi riscontrati in letteratura. Successivamente sono descritti i metodi di Morris ed EFAST, utilizzati in questo studio, presentando una descrizione accurata di tutti i parametri analizzati nella SA, che compongono i due modelli presi in esame. Sono inoltre descritti gli indici di plasticità e complessità, utilizzati per effettuare il confronto tra i due modelli. Infine

sono presentati e discussi i risultati della SA effettuata sugli scenari specifici presi in esame.

Nel capitolo 7 sono descritti i metodi di assimilazione dati in modelli di accrescimento delle colture sviluppati in questa ricerca. Il metodo di assimilazione SAFY-ENKF è stato da me sviluppato con l'aiuto del Prof. Casa, utilizzando la piattaforma Matlab. Il metodo di assimilazione Aquacrop-PSO è stato sviluppato dal partner cinese al progetto Dragon3, sviluppo al quale ho potuto contribuire parzialmente durante il mio periodo di permanenza presso la sede centrale di Beijing del NERCITA, avvenuto tra maggio e luglio del 2016.

Nel capitolo 8 sono presentati i risultati dei metodi di assimilazione per entrambi i metodi, sia per applicazioni puntuali che spazializzate, le prime validate per mezzo dei dati *in situ* raccolti durante le campagne sperimentali di misura, le seconde confrontate con i dati statistici forniti da enti ufficiali.

Infine il capitolo 9 conclude questa tesi, riassumendo e discutendo i principali risultati ottenuti, presentando le nuove problematiche insorte durante lo sviluppo e l'applicazione di questi metodi e suggerendo potenziali miglioramenti per future ricerche relazionate alle tematiche trattate.

2. Concetti di base

2.1. Modelli di accrescimento delle colture

I modelli di accrescimento delle colture sono modelli matematici che descrivono il funzionamento e lo sviluppo di una coltura e delle interazioni con lo scenario in cui la coltura è coltivata. Per scenario si intende l'insieme delle caratteristiche meteorologiche ed ambientali (ad esempio del suolo) in cui è inserita la coltura. In termini matematici i modelli di accrescimento delle colture possono essere visti in due modi differenti e complementari: come sistemi dinamici in funzioni del tempo o come modelli di risposta, funzioni di variabili esplicative (Wallach, 2006a).

La forma generale di un sistema dinamico è:

$$\begin{aligned} U_1(t + \Delta T) &= U_1(t) + g_1[U(t), X(t); \theta] \\ &\vdots \\ U_s(t + \Delta T) &= U_s(t) + g_s[U(t), X(t); \theta] \end{aligned} \tag{2.1}$$

dove:

t: tempo

Δt : incremento di tempo

$U(t) = [U_1(t), \dots, U_s(t)]$: vettore delle variabili di stato al tempo t

$X(t)$: vettore delle variabili esplicative al tempo t

θ : vettore dei parametri

$g[\dots]$: funzione generica

Per i modelli di accrescimento delle colture Δt corrisponde generalmente ad un giorno, le variabili di stato sono generalmente variabili biofisiche come Leaf Area Index (LAI) o la biomassa, le variabili esplicative (input) in genere si riferiscono alle variabili che descrivono lo scenario, come le caratteristiche del suolo, le variabili climatiche, le condizioni iniziali o le variabili di gestione (e.g. date di semina, irrigazioni ecc.). Le variabili di stato occupano un ruolo fondamentale nei modelli che usano un sistema dinamico. Esse determinano le equazioni da includere nel sistema. Di conseguenza definiscono quali processi descrivono l'evoluzione nel tempo della coltura in esame. Ogni variabile di stato corrisponde ad una caratteristica della coltura (generalmente una variabile biofisica della coltura) e la simulazione è indipendente da misure o osservazioni esterne. Pertanto, partendo da una condizione iniziale nota è

possibile determinare il valore delle variabili di stato della coltura ad ogni momento t del ciclo (dove generalmente t corrisponde ad un giorno). Le variabili esplicative descrivono caratteristiche che variano in modo indipendente dal sistema in esame. Le variabili meteorologiche, ad esempio, sono variabili esplicative in quanto la loro variazione ad ogni istante t non dipende da equazioni appartenenti al modello di accrescimento della coltura, ma da misure quotidiane esterne o da modelli climatici esterni al sistema in esame. Allo stesso modo alcune caratteristiche della coltura possono essere inserite nel sistema come variabili esplicative, in funzione di equazioni esterne al sistema, con il fine di migliorare la stima delle variabili di stato senza aumentare eccessivamente la complessità del modello. I parametri sono valori indipendenti dal tempo che possono descrivere sia caratteristiche reali delle colture, che semplicemente rappresentare coefficienti nelle equazioni, utili a mettere in relazione grandezze differenti. La principale differenza con le variabili è che essi non sono funzioni del tempo, generalmente non variano durante l'esecuzione del modello. In fase di sviluppo del modello una caratteristica che descrive lo stato di crescita della coltura può essere inserita nel sistema come variabile di stato o come variabile esplicativa a seconda dell'obiettivo che si vuole perseguire. L'aumento del numero di variabili diminuisce l'approssimazione con cui il modello descrive la realtà, però aumenta anche la complessità del modello introducendo nuove equazioni e nuovi parametri da calibrare, inserendo di fatto ulteriori possibilità di errore. Pertanto nello sviluppo di un modello si deve tener conto del giusto equilibrio tra la complessità e l'accuratezza con cui il modello simula lo sviluppo di una coltura.

I modelli di accrescimento delle colture possono essere visti anche come modelli di risposta, integrando le equazioni differenziali dei modelli con sistemi dinamici. Il risultato di questa integrazione è di eliminare i valori intermedi delle variabili di stato (Wallach, 2006a). Le variabili di stato a ogni tempo t sono in questo modo funzioni esclusivamente delle variabili esplicative per tutti i tempi da $t = 0$ a $t = T - \Delta t$. Le variabili di stato, successivamente all'integrazione, possono essere così rappresentate:

$$U_i(T) = f_{i,T}[X(0), X(\Delta t), X(2\Delta t), \dots, X(T - \Delta t); \theta] \quad \text{con } i = 1, \dots, S \quad (2.2)$$

In generale in un modello ci sono un numero limitato di risultati (variabili di output) che sono di interesse primario. Ci si può riferire a questi come a variabili di

risposta. Possono essere variabili di stato ad un particolare tempo o funzioni delle variabili di stato. Le variabili di risposta sono anche definiti output del modello, un esempio molto comune nei modelli di accrescimento delle colture è la produzione finale. Spesso questa variabile viene calcolata a partire dalla biomassa (che è una variabile di stato) attraverso proporzionalità diretta regolata da un parametro (harvest index) compreso tra 0 e 1. Rappresentando con Y la variabile di risposta, o output del modello, è possibile semplificare l'equazione (2.2) tramite l'equazione (2.3):

$$Y = f(X; \theta) + \varepsilon \quad (2.3)$$

dove X rappresenta il vettore delle variabili esplicative per tutti i tempi t compresi tra 0 e il tempo finale di interesse che definisce l'output e θ è il vettore dei parametri. L'equazione 2.3 differisce dalla 2.2 per la presenza di un termine ε . Si tratta di una variabile casuale che rappresenta e sintetizza gli errori del modello. L'errore sorge sia dal fatto che le variabili esplicative non spiegano tutta la variabilità nelle variabili di risposta, che dalla possibilità di errore nelle equazioni che semplificano la realtà, o dall'incertezza nelle variabili esplicative utilizzate. Tale rappresentazione del modello è molto utile, ad esempio, nell'applicazione di metodi di assimilazione, in particolare del metodo basato su Ensemble Kalman Filter, in quanto permette di sintetizzare in modo semplice ed immediato l'errore sul modello, valore fondamentale in questo tipo di applicazione.

Lo scopo del modello è di simulare lo sviluppo di una coltura, rappresentando attraverso algoritmi matematici lo sviluppo fenologico della pianta e le interazioni tra l'ambiente e la vegetazione che determinano la sua crescita. Al fine di rendere il modello usufruibile è necessario effettuare semplificazioni e approssimazioni della realtà, privilegiando alcuni processi biofisici e trascurandone altri. La funzione principale di un modello di accrescimento delle colture è mettere in relazione gli input (variabili esplicative e parametri) con gli output (variabili di stato e di risposta). Tuttavia, le equazioni dinamiche che lo compongono non descrivono la realtà dei processi, a causa delle semplificazioni fatte nello svilupparlo. È pertanto fondamentale valutare l'attendibilità e l'accuratezza del modello confrontando le simulazioni con osservazioni reali, mediante la validazione. Ciò consente inoltre di quantificare l'errore del modello (Wallach, 2006b). La valutazione di accuratezza e affidabilità è fondamentale perché aiuta a chiarire i vantaggi dell'uso di un modello, le condizioni

in cui il modello può essere applicato ed il livello di qualità che può ritenersi accettabile. Inoltre è indispensabile per poter applicare tecniche per migliorare le stime del modello, come ad esempio i metodi di assimilazione. Tale stima di accuratezza è effettuata confrontando i risultati del modello con misure ed osservazioni indipendenti, in questo modo il modello è validato con dati *in situ*. Il metodo più immediato per confrontare i dati misurati è certamente grafico, confrontando i valori calcolati con quelli misurati è possibile osservare in modo qualitativo l'accuratezza del modello dalla distanza tra valori misurati e simulati. Per ottenere una valutazione quantitativa dell'errore di stima del modello è necessario però utilizzare indici che esprimano la differenza tra simulazioni e misure in campo, distinguendo l'effetto dovuto ad errori sistematici e casuali. Di seguito sono indicati solo alcuni di questi indici, utili a comprendere le validazioni effettuate in questa tesi, un approfondimento sui metodi di valutazione di un modello e i relativi indici può essere trovato in Mayer and Butler (1993), Mesplé et al. (1996), Rykiel (1996).

L'elemento basilare per quantificare la concordanza tra osservazioni e modello è la differenza tra i due valori:

$$D_i = Y_i - \hat{Y}_i \quad (2.4)$$

dove:

Y_i : valore misurato in campo alla situazione i

$\hat{Y}_i$: valore simulato dal modello alla situazione i

Il valore Y può essere un qualsiasi output del modello come ad esempio la produzione, il LAI o la biomassa ad una data specifica, ecc. La distanza D_i può essere rappresentata in un grafico dei residui. Un modo molto semplice e molto diffuso per riassumere i valori D_i è l'errore quadratico medio (Mean Square Error, MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N (D_i)^2 \quad (2.5)$$

Molto spesso all'indice MSE si preferisce la radice quadrata dell'errore quadratico medio (Root Mean Square Error, RMSE):

$$RMSE = \sqrt{MSE} \quad (2.6)$$

Il vantaggio di usare la RMSE è che ha la stessa unità di misura di Y , pertanto risulta più immediata la comprensione dell'ordine di grandezza dell'errore.

Una variante della RMSE è la radice dell'errore quadratico medio relativo (RRMSE):

$$RRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(Y_i - \hat{Y}_i)^2}{Y_i}} \quad (2.7)$$

Questo indice è maggiormente usato per comparare errori basati su data set differenti, in quanto è espresso in percentuale essendo un rapporto tra RMSE e la media dei valori di Y_i . Inoltre, essendo una grandezza adimensionale presenta il vantaggio di non tener conto dell'unità di misura di partenza. RRMSE è l'indice utilizzato in questa tesi per quantificare l'accuratezza dei modelli esaminati.

2.2. Analisi di sensibilità (Sensitivity Analysis)

I modelli matematici che simulano la crescita delle colture introducono semplificazioni e approssimazioni al fine di schematizzare processi generalmente molto complessi e concatenati tra loro. Tali semplificazioni introducono un certo livello di incertezza nei risultati, pertanto è importante stabilire la relazione che esiste tra gli input e gli output del modello, ovvero come variano i valori di output al variare dei valori degli input e dei parametri. Lo strumento adatto a studiare questo aspetto è la Sensitivity Analysis (SA). Questa può essere definita come lo studio del modo in cui la variazione dei fattori influenzi gli output del modello (Saltelli et al., 2004). Gli obiettivi che si vogliono raggiungere applicando la SA sono: controllare che gli output del modello si comportano atteso al variare degli input; identificare in che modo i parametri influenzano l'output ed evidenziare quali parametri risultano più influenti, quali meno influenti e quali trascurabili; identificare quali parametri necessitano di essere stimati con più accuratezza; individuare e quantificare gli effetti delle interazioni tra i parametri; determinare possibili semplificazioni del modello; identificare le variabili di input da misurare con la massima accuratezza (Monod et al., 2006). La SA risulta molto utile in fase di progettazione e sviluppo del modello, in quanto consente di migliorarne le prestazioni, ma è molto utile anche in fase di utilizzo in quanto consente di semplificare e comprendere meglio il funzionamento del modello, di studiare il comportamento del modello al variare delle condizioni al contorno (come condizioni climatiche e ambientali) e di semplificarne la calibrazione. Nel caso di applicazione di metodi di assimilazione, l'individuazione dei parametri che maggiormente influenzano l'output è un passo irrinunciabile, pertanto risulta indispensabile uno studio di sensibilità dei parametri e delle variabili di input. Si distinguono due differenti classi di metodi di SA: locali e globali (Saltelli et al., 2008).

I metodi di SA locali esaminano l'impatto locale dei fattori (per fattori si intende l'insieme dei parametri e delle variabili esplicative) sugli output del modello. Si effettua calcolando le derivate parziali delle variabili di output rispetto ai fattori di input (Monod et al., 2006). Ogni fattore (parametro o variabile esplicativa) è fatto variare intorno ad un valore nominale, in modo da calcolare numericamente le derivate delle variabili di output. Durante la variazione di un singolo fattore, tutti gli altri sono tenuti costanti. Questo approccio è fattibile quando l'intorno attorno al valore nominale è abbastanza piccolo, in modo da considerare lineare la relazione tra input e output. I metodi di SA locali sono caratterizzati da un costo computazionale ridotto, ma sono inadeguati a valutare gli effetti delle variazioni simultanee di tutti i fattori di input (Saltelli et al., 2008).

I metodi di SA globali valutano gli effetti che la variazione simultanea di tutti i fattori ha sugli output del modello in un ampio intervallo di variazione (Pianosi et Wagener, 2014). La stima della sensibilità di ogni singolo parametro o variabile esplicativa è fatta variando allo stesso tempo tutti i fattori di input, e misurandone l'effetto contemporaneamente sulla variabile di output presa in esame. L'inconveniente di questi metodi è che hanno tempi di calcolo più lunghi dei precedenti e nel caso di un numero troppo elevato di parametri risulta impossibile la convergenza dell'algoritmo. Un'accurata panoramica dei metodi di SA è fatta da Cariboni et al. (2007), in cui sono messe in evidenza le differenze tra metodi locali e globali. Cariboni et al. (2007) inoltre, mostrano l'inadeguatezza dei metodi locali per i modelli di accrescimento delle colture, perché la loro complessità impone di conoscere le interazioni tra le variazioni dei parametri, pertanto in questa tesi sono stati applicati metodi di SA globali.

I metodi di SA globali sono suddivisi in tre gruppi: metodi di screening, come ad esempio il metodo di Morris (Morris, 1991), metodi basati sulla varianza, come ad esempio SOBOL (Sobol, 1993), il Fourier Amplitude Sensitivity Test (FAST, Cukier et al, 1978) o il metodo Extended FAST (EFAST; Saltelli et al., 1999) ed infine metodi basati sulla regressione, che si distinguono principalmente per la metodologia di campionamento dei fattori, come ad esempio il Latin Hyper Cube Sampling o il Random e il Quasi Random LpTau (Confalonieri et al., 2010a).

Confalonieri et al. (2010a) fa un confronto tra le tecniche di SA sopracitate, applicandole su un modello di accrescimento del riso WARM (Water Accounting Rice Model). Viene dimostrata la concordanza tra le classifiche dei parametri della coltura,

ordinati in ordine decrescente dal più rilevante al meno rilevante, inoltre viene valutata l'accuratezza e l'efficienza di ogni metodo.

2.3. Assimilazione dati

I principali problemi legati ai modelli che descrivono gli agroecosistemi, e più in generale i modelli matematici, sono la possibilità di cadere in una descrizione troppo semplificata dei sistemi naturali, un'inaccurata parametrizzazione, un'elevata incertezza nei dati di input e di conseguenza una riduzione delle prestazioni del modello (Dorigo et al., 2007). L'assimilazione dati è uno strumento che consente di riallineare l'andamento della simulazione di un dato fenomeno con l'andamento osservato delle variabili di stato, avendo a disposizione una limitata quantità di dati misurati o ulteriormente simulati. In sostanza il modello viene modificato introducendo dati esterni col fine di migliorare l'accuratezza della stima degli output (Makowski et al., 2006). L'obiettivo generale dell'assimilazione dati è caratterizzare lo stato di un sistema combinando informazioni da varie fonti, come modelli matematici ed osservazioni nello spazio e nel tempo delle variabili d'interesse. Tali metodi prevedono la scelta di una o più variabile guida (*driving variable*), esse sono le grandezze misurate che andranno ad interagire con quelle simulate, il punto di contatto tra modello e misure (Delécolle et al., 1992). È possibile dividere i metodi di assimilazioni in 3 categorie, largamente descritte in letteratura (Delécolle et al., 1992; Makowski et al., 2006; Dorigo et al., 2007; Li et al., 2015):

- *Calibration* (Calibrazione): con i metodi di calibrazione, i valori dei parametri del modello sono modificati in modo da ottenere una concordanza ottimale tra le variabili di stato, simulate e osservate, che caratterizzano il comportamento del sistema (Fig. 2.1a). In questo tipo di metodi i parametri vengono ricalibrati in modo da minimizzare la differenza tra le osservazioni e le simulazioni mediante metodi di ottimizzazione numerica. La calibrazione dei parametri del modello deve essere comunque effettuata, successivamente vengono individuati i parametri più influenti e solo questi vengono ricalibrati. Una ricalibrazione di tutti i parametri non è fattibile perché avrebbe un costo computazionale non sostenibile dalle attuali tecnologie ed introdurrebbe ulteriori errori. I metodi di assimilazione per calibrazione hanno il vantaggio di non dover modificare il modello che descrive il sistema, possono essere quindi utilizzati per tutti i modelli il cui codice non sia accessibile. In questa tesi, ad esempio, è stato

utilizzato un metodo di assimilazione per calibrazione per assimilare dati ottenuti da telerilevamento nel modello Aquacrop, attraverso il metodo Particle Swarm Optimization (PSO). Esempi di applicazioni di questo tipo di metodi sono forniti da Doraiswamy et al. (2003), Launay et Guerif (2005), De Wit et al. (2012), Li et al. (2015).

- *Forcing* (forzatura): nel metodo di forzatura il valore di una variabile di stato simulato dal modello viene rimpiazzato con un dato osservato (Fig. 2.1 b). L'uso diretto dei dati osservati per descrivere una variabile di stato richiede la disponibilità di osservazioni a ogni time step del modello (generalmente giornaliero nei modelli di accrescimento delle colture, più raramente settimanale). Tale caratteristica ne limita fortemente l'utilizzo nei modelli che simulano agroecosistemi, a causa della difficoltà di ottenere osservazioni con una frequenza giornaliera, specialmente se tali osservazioni vengono effettuate da telerilevamento, dove la frequenza di passaggio su uno stesso punto è compresa, generalmente tra i 5 e i 20 giorni. Per derivare la variabile di stato al time step del modello e riempire lo spazio tra due osservazioni sono impiegate diverse tecniche di interpolazione, come interpolazione lineare, fast Fourier transformations (Roerink et al., 2010) ed approcci wavelet.
- *Updating* (aggiornamento): il metodo di updating consiste nell'aggiornare una o più variabili di stato ogni volta che è disponibile un'osservazione (Fig. 2.1 c). Questo metodo è basato sull'assunzione che una variabile di stato meglio simulata in un istante t migliorerà l'accuratezza della simulazione di tale variabile in tutti gli istanti successivi. In letteratura sono presenti numerose applicazioni di metodi di assimilazione per updating a modelli di agroecosistemi (Wu et al, 2012; Ines et al., 2013; Huang et al, 2013; Jiang et al. 2014; Dong et al. 2016). Negli ultimi anni i metodi per updating sono stati maggiormente applicati nell'ambito di modellistica di sistemi agroecologici e idrologici a causa della loro versatilità nel combinare simulazioni e osservazioni. Il principale vantaggio di tali sistemi è che essi non assimilano il valore vero misurato perché anche quest'ultimo è affetto da un errore non trascurabile, ma assimilano un valore che tiene conto sia degli errori sulla misura sia di quelli sui modelli.

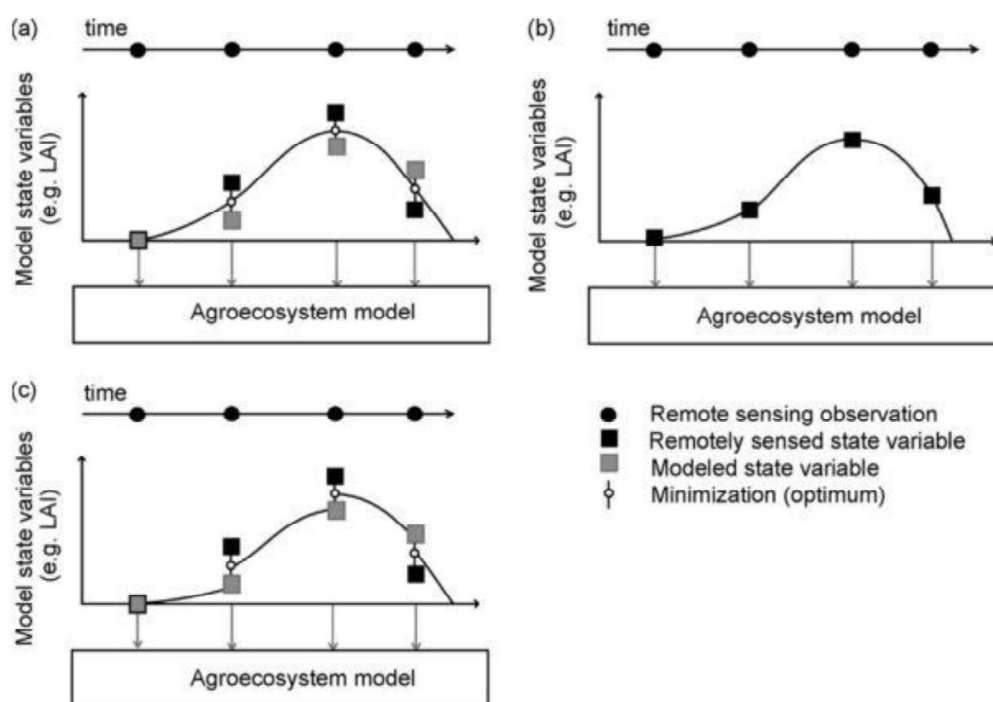


Fig. 2.1: Rappresentazione schematica di differenti metodi per l'assimilazione di variabili di stato in modelli di agroecosistemi (in questo caso dati ottenuti da Remote Sensing, ma adattabile a qualsiasi tipo di dato osservato): (a) Calibration, (b) Forcing, (c) Updating (da Dorigo et al., 2007)

Numerosi algoritmi basati sulla strategia di assimilazione updating sono stati sviluppati, di seguito (Capitolo 7) è stato dettagliato esclusivamente l'algoritmo basato sul Ensemble Kalman Filter, in quanto utilizzato in questo studio. Per ulteriori dettagli sui metodi di assimilazione updating è possibile consultare, ad esempio, Makowski et al. (2006), Law et al. (2011), Montzka et al. (2012).

2.4. Principi di Telerilevamento

Il telerilevamento, o Remote Sensing, è una disciplina tecnico scientifica il cui scopo è l'acquisizione di informazioni circa un oggetto, un fenomeno o un'area geografica pur non essendo fisicamente nelle vicinanze del target interessato. È utilizzata in innumerevoli campi: geografico, scienze della terra, militare, commerciale, economico fino anche ad applicazioni umanitarie. In questo paragrafo si è tentato di riassumerne i principi basi con particolare riferimento alle applicazioni per il monitoraggio ambientale.

Il termine Remote Sensing ha avuto molte definizioni nel corso del tempo, seguendo l'evoluzione tecnologica degli strumenti che ne permettono l'impiego. Una delle definizioni che meglio riassume il concetto di Telerilevamento è "La scienza che acquisisce informazioni sulla superficie terrestre senza venire a contatto con essa"

(Fischer et al., 1975), ovvero “la scienza di ricavare informazioni circa la superficie della Terra da immagini acquisite a distanza. Essa si basa su misure dell’energia elettromagnetica riflessa o emessa dalle caratteristiche di interesse” (Campbell, 2002). L’energia elettromagnetica a cui fa riferimento Campbell (2002) nel definire il Telerilevamento è quella trasportata dalla radiazione elettromagnetica. La radiazione elettromagnetica può essere vista come un’onda che si propaga nel vuoto alla velocità della luce (la stessa luce solare è una radiazione elettromagnetica). Nella sua forma di onda può essere descritta da due caratteristiche: la lunghezza d’onda (λ), ovvero la distanza tra le creste di due onde sequenziali, o dalla frequenza (ν), che è il numero di creste che passa per un punto fissato durante un numero di intervalli stabilito. Nell’ambito del telerilevamento va sempre più affermandosi la convenzione di descrivere la radiazione elettromagnetica rispetto alla lunghezza d’onda.

Tutte le possibili lunghezze d’onda delle radiazioni elettromagnetiche compongono quello che è definito “lo spettro elettromagnetico”, dalle lunghezze di 10^{-12} m delle radiazioni gamma alle 10^3 delle onde radio (Fig. 2.2). La radiazione elettromagnetica è classificata, sulla base della lunghezza d’onda, in regioni o bande (radiation type in Fig. 2.2). Lunghezze d’onda che si rivelano di grande interesse nel telerilevamento appartengono alla banda ottica, la quale si estende tra i 0.3 e i 15 μ m. A queste lunghezze d’onda l’energia può essere riflessa o rifratta da materiali come specchi e lenti.

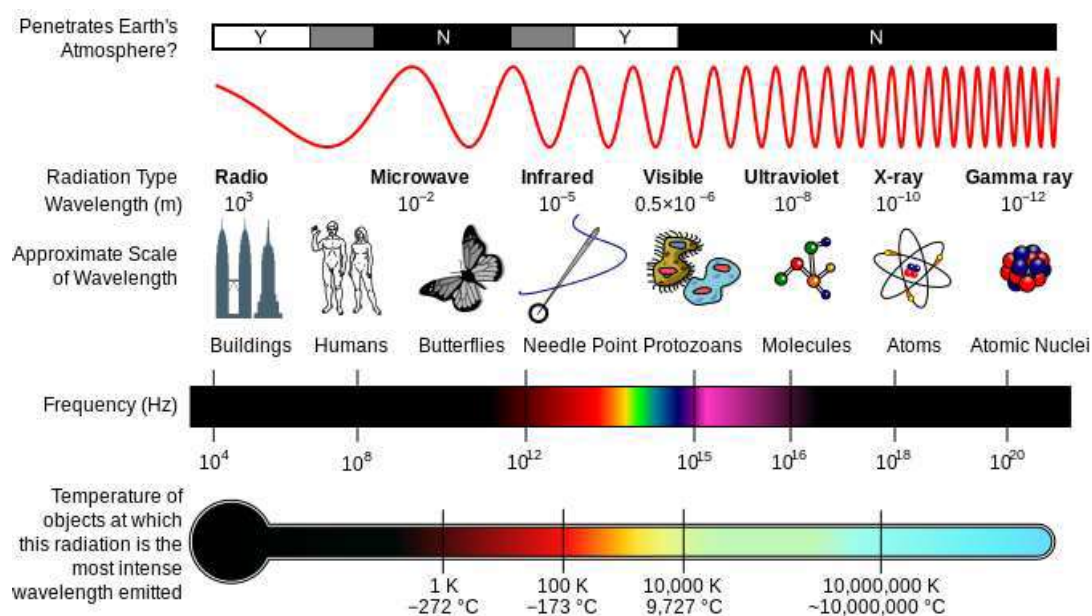


Fig. 2.2: Diagramma dello spettro elettromagnetico nel quale è mostrato il tipo, la lunghezza d'onda (con relativi esempi comparativi di dimensione), la frequenza, l'emissione di temperatura di un corpo nero (riadattata da www.nasa.org)

La parte dello spettro di maggior interesse per il telerilevamento è la cosiddetta parte riflettiva dello spettro, in cui la luce è principalmente originaria dal sole e riflessa dalla superficie terrestre. Tale parte è ulteriormente divisibile in 3 sezioni: ultravioletto (0.2-0.4 μm), visibile (0.4-0.7 μm), infrarosso riflettivo (0.7-3 μm). L'occhio umano è sensibile alle radiazioni comprese fra 0.38 e 0.72 μm , questa è la banda spettrale indicata come visibile, che comprende i tre "colori fondamentali": blu (0.38 - 0.48 μm), verde (0.48 - 0.60 μm) e rosso (0.60 - 0.72 μm). Le lunghezze d'onda inferiori a 0.3 μm non sono interessanti per il telerilevamento, mentre si vanno sempre più sviluppando sistemi che utilizzano bande appartenenti alla regione delle microonde (0.3-133 cm). In questa sezione ci limiteremo a introdurre i sensori che operano nella regione dell'ottico, per ulteriori informazioni più dettagliate sul telerilevamento si possono consultare numerosi testi come ad esempio Fischer et al. (1975) o Campbell (2002).

Per questa tesi sono stati utilizzati esclusivamente sensori passivi che operano nella regione dell'ottico. Tali sensori sono detti passivi perché rilevano la radiazione emessa dal sole e riflessa dai target sulla terra (siano essi superficie terrestre, specchi d'acqua, aerosol o nubi). Il sistema sole-target-sensore è mostrato in Fig. 2.3, la qualità e le caratteristiche del segnale misurato dipendono dalle caratteristiche del sensore (numero di bande, risoluzione, ecc.).

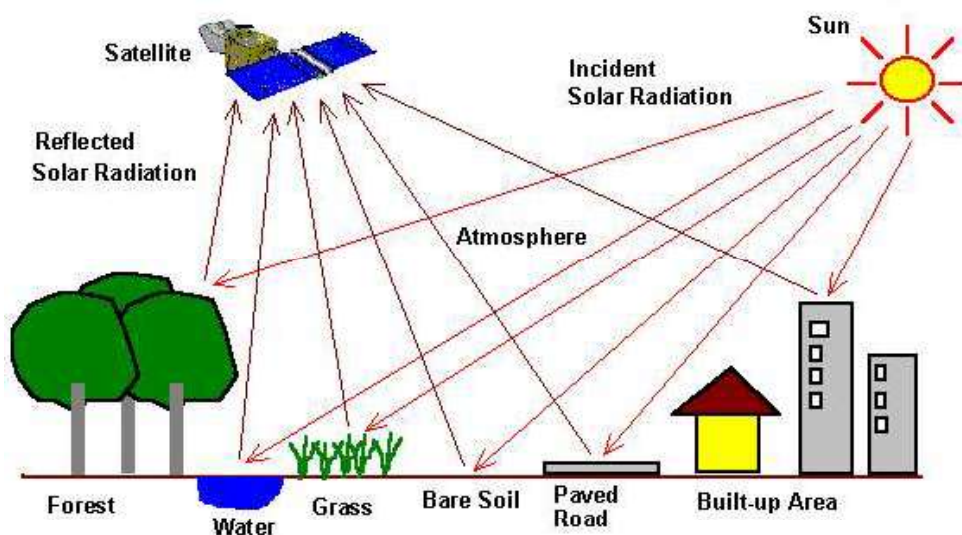


Fig. 2.3: Rappresentazione stilizzata del funzionamento di un sensore ottico passivo a bordo di un satellite (da <http://www.crisp.nus.edu.sg/~research/tutorial/optical.htm>)

Per le applicazioni del telerilevamento al monitoraggio ambientale, e più in particolare all'osservazione della vegetazione, bisogna specificare che il segnale misurato dal sensore è principalmente quello riflesso dalla parte alta del manto vegetale ("canopy"). Il segnale è condizionato da numerosi aspetti quali le condizioni

d'illuminazione, che dipendono dalla stagione in cui è acquisita l'immagine, l'ora, il giorno dell'anno, dalla posizione della scena (latitudine e longitudine) rispetto al sole, dalle condizioni atmosferiche, quali ad esempio la presenza di nubi, il contenuto di acqua o l'assorbimento e lo scattering dovuto alla presenza di aerosol. Le caratteristiche fisiche della vegetazione come la forma, la struttura, la geometria e la composizione possono influire sulle proprietà della radiazione riflessa ed emessa. Anche le proprietà del sensore come, l'altitudine e la geometria dell'osservazione del satellite, hanno un'influenza diretta sulla radiazione misurata (Veloso, 2014). La posizione del sensore e del sole rispetto al punto osservato sulla terra è convenzionalmente rappresentata da 4 angoli: angolo zenitale di osservazione (view zenith angle, L_z), angolo zenitale d'illuminazione (sun zenith angle, Z_s), angolo azimutale di osservazione (view azimuth angle, γ) ed angolo azimutale d'illuminazione (sun azimuth angle, γ_s) (Fig. 2.4).

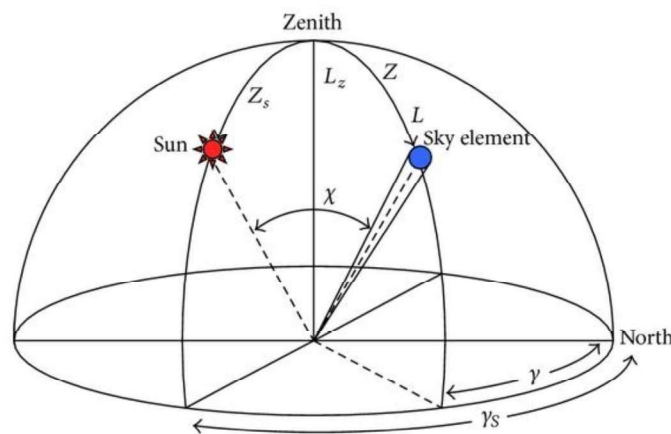


Fig. 2.4: Geometria di osservazione ed illuminazione nel telerilevamento per un sensore passivo. Sono indicati l'angolo zenitale del sole (Z_s), l'angolo zenitale di osservazione (L_z), l'angolo azimutale di osservazione (γ) e del sole (γ_s) (da www.hindawi.com)

Per meglio comprendere le misure della radiazione fatte dai vari sensori è necessario comprendere la definizione di riflettanza. La riflettanza è il rapporto tra la quantità totale di radiazione riflessa dalla superficie e la quantità totale di radiazione incidente sulla superficie. È pertanto una grandezza adimensionale compresa tra 0 e 1. Lo spettro della riflettanza o firma spettrale di un elemento è il grafico della frazione di radiazione riflessa in funzione della lunghezza d'onda della radiazione emessa ed è distinto per ogni elemento considerato (Fig. 2.5).

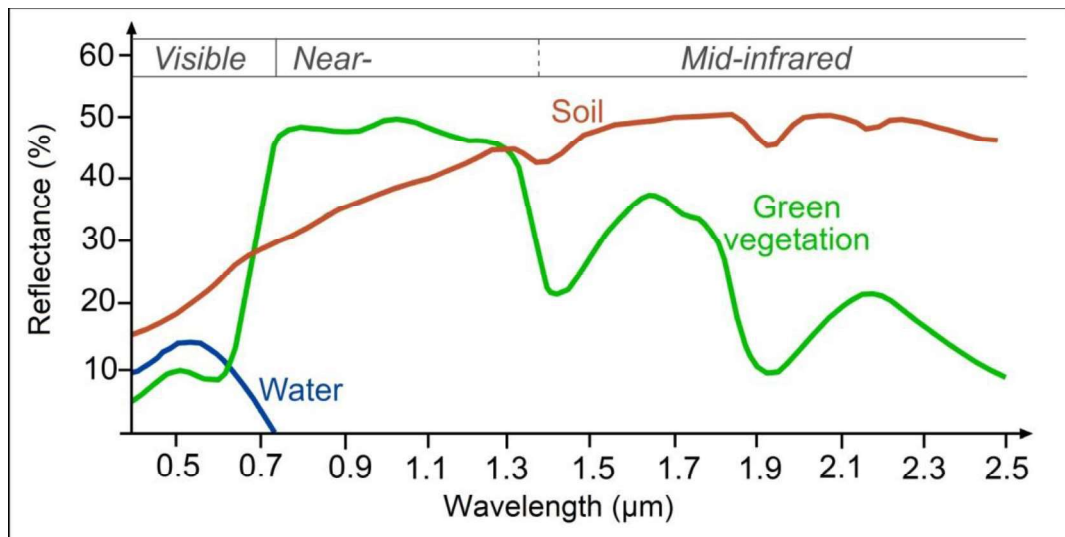


Fig. 2.5: Esempi di firme spettrali di tre distinti elementi: acqua, suolo e vegetazione verde (da <http://www.seos-project.eu>).

Le risposte spettrali della vegetazione mettono in relazione i dati ottenuti da telerilevamento mediante sensori ottici con le variabili biofisiche e biochimiche delle colture, è quindi attraverso di esse che si possono ricavare informazioni ad esempio circa lo stato di crescita delle colture. Lo stato di accrescimento della vegetazione è rappresentato da diversi indici, i quali riassumono caratteristiche biofisiche della vegetazione. Due sono gli indici di interesse in questo studio: il Leaf Area Index e il Canopy Cover.

Il Leaf Area Index (LAI) può essere definito come la metà del totale dell'area sviluppata di foglie verdi per unità di area superficiale orizzontale al suolo (Chen et Black, 1992). Nel telerilevamento, a causa dell'impossibilità di distinguere le foglie dalle altre parti foto-sinteticamente attive sarebbe più proprio definire il Green Area Index, in quanto non fa riferimento alle sole foglie ma più in generale a tutte le parti verdi della pianta (steli, rami e gli organi riproduttivi). Seppur formalmente differenti il LAI e il GAI sono molto spesso considerati equivalenti (Veloso, 2014).

Il Canopy Cover (CC) o la frazione di copertura del suolo da parte della vegetazione (Fraction of green vegetation Cover, fCover) è più semplicemente la percentuale di suolo coperta da vegetazione verde.

3. Aree di studio e descrizione dei dati

3.1. Area di studio e dataset climatici

L'area di studio in cui si è svolto il lavoro di analisi dell'accrescimento del frumento tenero è localizzata nella zona rurale di Yangling, una città situata nel centro della Cina, nella regione dello Shaanxi (34.27°N , 108.09°E). La sperimentazione dei metodi di assimilazione sviluppati in questo studio è stata interamente svolta in quest'area, pertanto tutti i dati, sia acquisiti da telerilevamento che *in situ* fanno riferimento a questa area di studio. L'analisi di sensibilità ed un'attività preliminare di calibrazione, è stata svolta anche utilizzando dati raccolti a Xiaotanshan, nelle vicinanze di Pechino (44.17°N , 116.43°E).



Figura 3.1: Localizzazione dei siti sperimentali

L'area sperimentale di Yangling è caratterizzata da un clima semiarido, con una media di precipitazioni annue di circa 570 mm, per lo più concentrate tra maggio e settembre e quasi completamente assenti nei restanti mesi. Le temperature variano tra i -4°C in inverno e i 35°C in estate e raggiungono valori sotto lo 0 pochi giorni all'anno tra dicembre e febbraio. Il dataset climatico utilizzato in questo studio per la zona di Yangling è riferito agli anni 2013, 2014 e 2015 (Figura 3.2 b).

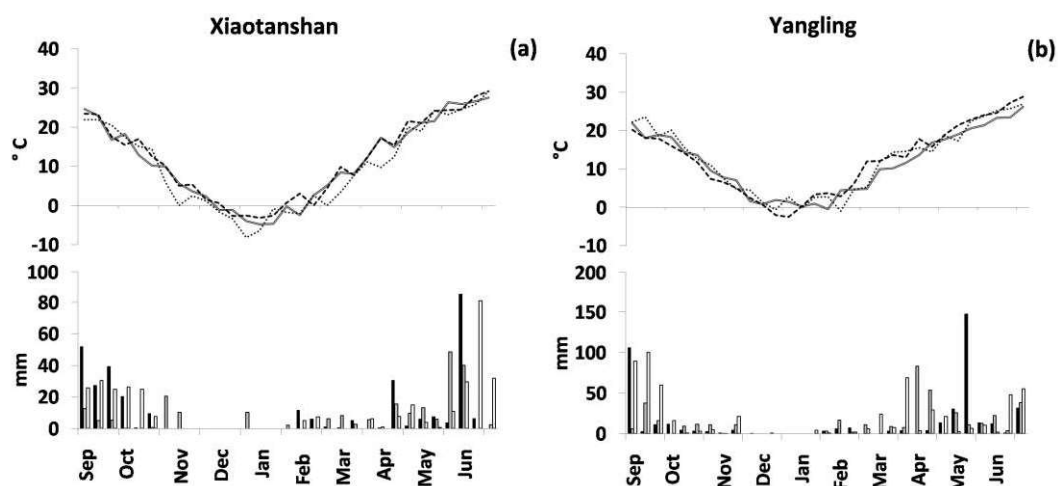


Figura 3.2: (a) Temperature (in alto) e precipitazioni (in basso) per il sito di Xiaotanshan per gli anni 2008-2009 (linea tratteggiata e barre nere), 2009-2010 (linea punteggiata e barre grigie), 2010-2011 (linea continua e barre bianche); (b) Temperature (in alto) e precipitazioni (in basso) per il sito di Yangling per gli anni 2012-2013 (linea tratteggiata e barre nere), 2013-2014 (linea punteggiata e barre grigie), 2014-2015 (linea continua e barre bianche).

Il ciclo colturale dell'anno 2012-2013 presenta un picco di -5.9°C raggiunto in inverno e un massimo di 39.3°C raggiunto in primavera, entrambi fuori dall'intervallo delle temperature medie caratteristiche di questa zona, ciononostante non risulta essere, nella media complessiva, né l'anno più freddo né il più caldo tra quelli analizzati. Le precipitazioni sono per lo più concentrate nei giorni subito successivi alla semina (inizio di ottobre) e nei mesi finali del ciclo (aprile e maggio), molto vicine alle medie stagionali caratteristiche. Il ciclo colturale 2013-2014 è caratterizzato da temperature mediamente più alte, sia in inverno che in primavera, con temperature comprese tra i -2°C e i 35.2°C , è quindi il ciclo colturale mediamente più caldo tra quelli considerati. Le precipitazioni nel periodo di crescita del frumento sono inferiori alle medie stagionali caratteristiche di questa zona, risultando il ciclo colturale con meno precipitazioni tra quelli considerati. Il 2014-2015 è l'anno più freddo tra i tre presi in esame, con temperature tra i -7.5°C e 29.7°C , risulta invece che la concentrazione di pioggia durante il periodo di crescita della coltura sia il più abbondante tra i tre considerati.

Il sito sperimentale di Xiaotanshan, utilizzato esclusivamente per l'analisi di sensibilità, è caratterizzato da un tipico clima continentale, con temperature medie invernali intorno ai -4°C e massime in estate intorno ai 27°C . Le precipitazioni annue sono mediamente intorno ai 650 mm di pioggia, concentrate particolarmente nei mesi estivi e sono particolarmente scarse nel periodo di crescita del frumento tenero. Anche le temperature medie sono più basse per Xiaotanshan, in particolare nei mesi invernali, quando vi sono lunghi periodi al di sotto dello zero. Il dataset climatico utilizzato per

l'area di Xiaotanshan è riferito ai cicli colturali di frumento duro avvenuti tra gli anni 2008 e 2011 (Figura 3.2 a). Osservando il periodo di pertinenza del ciclo di crescita del frumento duro si nota che il 2008-2009 è caratterizzato da una precipitazione complessiva di circa 222 mm di pioggia, temperature comprese tra i -9.3°C e i 39.6°C. Tra i 3 periodi presi in esame nell'area di Xiaotanshan questo risulta essere il più piovoso e il più caldo. Il 2009-2010 è invece il più freddo e il meno piovoso, con temperature medie invernali intorno a -1.7°C, e picchi fino a -11.6°C. Il 2010-2011 risulta invece pienamente in linea con le medie stagionali, con circa 190 mm di pioggia (riferita esclusivamente al periodo ottobre-maggio) e temperature comprese tra i -7°C e i 35°C.

Il dataset delle variabili meteorologiche, raccolto per entrambe le aeree ed utilizzato come input nei modelli, consiste in informazioni giornaliere su: precipitazioni, temperature (massime, medie, minime), umidità relativa e velocità del vento. L'evapotraspirazione è stata calcolata usando il programma ET₀ Calculator (Raes et al, 2012), in accordo con le indicazioni FAO (Allen et al, 1998). La radiazione solare è stata calcolata usando la relazione proposta da Bristow and Campbell (1984):

$$EstRad_i = tt_i \cdot PotRad_i \quad (3.8)$$

dove: EstRad_i è la radiazione giornaliera stimata (MJ/m²), PotRad_i è la radiazione potenziale giornaliera e tt_i è la trasmissività giornaliera, espresso dalla seguente equazione:

$$tt_i = \tau \left[1 - \exp\left(\frac{-b \cdot \Delta T_i^c}{month\Delta T}\right) \right] \quad (3.9)$$

con:

τ = trasmissività del cielo sereno

$$\Delta T = T_{max_i} - \frac{T_{min_i} + T_{min_{i+1}}}{2} \quad (10.3)$$

Tmax_i = temperatura massima giornaliera (°C)

Tmin_i = temperatura minima giornaliera (°C)

monthΔT = escursione termica mensile ΔT

b e c = parametri empirici

Il valore dei parametri usati in questo lavoro sono quelli di riferimento indicati in letteratura (Bristow and Campbell, 1984): $\tau=0.7$, $b=0.11$, $c=2$

Per quantificare il grado di stress idrico nei due siti è stato utilizzato il “normalized synthetic agrometeorological indicator” (SAM) proposto da Confalonieri et al. (2010b). Tale indice, compreso tra -1 e 1, è espresso dalla seguente equazione:

$$SAM = \frac{Rain - ET_0}{Rain + ET_0} \quad (3.11)$$

dove Rain e ET_0 rappresentano rispettivamente la precipitazione totale e l'evapotraspirazione di riferimento misurate nel periodo d'interesse (in questo studio tra ottobre e maggio, periodo di crescita del frumento tenero nelle regioni osservate). In Tab. 1 sono mostrati i valori di SAM per Xiaotanshan (tra il 2008 e il 2011) e Yangling (tra il 2012 e il 2015), evidenziando che entrambi i siti sono soggetti a deficit idrico (valori di SAM inferiori a 0), in particolare che Xiaotanshan risulta essere generalmente più secca che Yangling. In entrambe le zone viene praticata l'irrigazione per il frumento.

Tab. 3.2: Normalized Synthetic Agrometereological Indicar (SAM) per i siti sperimentali di Xiaotanshan e di Yangling, calcolati tra il 1 Ottobre e il 31 Maggio per ogni periodo analizzato

Sito	Anno			
	2009	2010	2011	Media
Xiaotanshan	-0.4697	-0.4582	-0.3525	-0.4268
	2013	2014	2015	
Yiangling	-0.2092	-0.362	-0.1617	-0.2443

3.2. Dati in situ e statistici

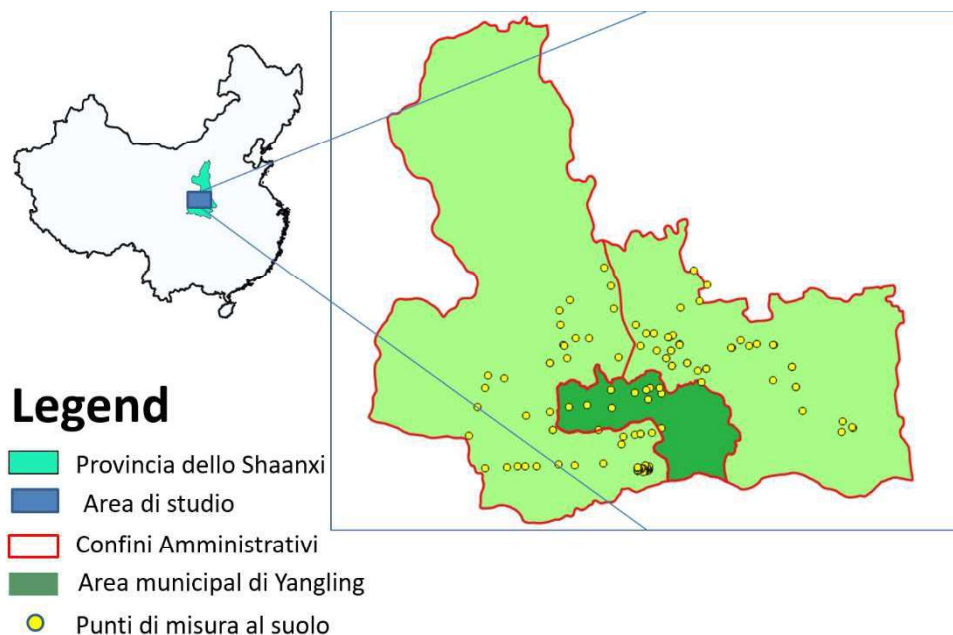


Figura 3.3: Area di studio situata nell'area rurale di Yangling (Shaanxi, Cina), con in evidenza i punti di misura a terra dislocati in tre aree amministrative differenti (confini in rosso).

I rilevamenti effettuati in campo sono consistiti in misure di LAI, biomassa e produzione finale, effettuate per ogni stagione colturale (2013, 2014 e 2015) tra marzo e giugno da personale del National Engineering Research Centre for Information Technology in Agriculture (NERCITA) di Pechino. In Tab. 3.2 è riportata una sintesi del dataset dei dati raccolti in campo.

Tab. 3.3: Date e tipologie di misure in campo di LAI (L), biomassa (b) e produzione finale (y) effettuate nel sito di Yangling tra il 2013 e il 2015. n.pts è il numero di punti di campionamento.

Anno	n. punti	Marzo	Aprile	Maggio	Giugno
2013	49	30 L, b	27 L, b		1 b,y
2014	35	5, 28 L, b	22, 27 L, b	16 L, b	9 b,y
2015	28	27 L, b	25 L,b		5 b,y

Le misure effettuate a terra (Fig. 3.3) sono state svolte in differenti campi di frumento tenero, su un'area di circa 1200 km². I campi presi in esame variano di anno in anno sia per posizione che per numero di campionamenti (Tab. 3.2). La biomassa è stata determinata raccogliendo un campione di piante in una superficie di 0.25 m² posizionata in modo casuale nelle vicinanze del punto di interesse. Tutti i campioni sono essiccati in stufa a 70°C fino ad un peso costante ed infine è stato registrato il peso secco. Il Leaf Area Index (LAI) è stato stimato usando il LAI-2000 Plant Canopy Analyzer (LI-COR Inc., Lincoln, NE, USA) e successivamente è stato convertito in Canopy Cover (CC) usando l'equazione proposta da Nielsen et al. (2012):

$$CC = 94.00(1 - e^{-0.43LAI})^{0.52} \quad (3.12)$$

Le informazioni inerenti alle proprietà del suolo, la varietà di frumento tenero e le date di irrigazione non sono disponibili. Le date di semina sono comprese nella prima quindicina di ottobre.

Le misure in campo sono state effettuate in un'area che appartiene a 3 amministrazioni municipali differenti, di cui solo quella di Yangling è pienamente compresa nell'area sperimentale di interesse (Fig. 3.3), le altre due appartengono ad aree più vaste non prese completamente in considerazione in questo studio. Pertanto è stato possibile ottenere i dati statistici forniti dagli enti ufficiali solo per l'area amministrativa di Yangling, quindi dell'estensione iniziale su cui erano distribuiti tutti i campi osservati, si è potuto ottenere i dati esatti esclusivamente di una superficie di circa 120 km². I dati statistici di produzione totale di frumento tenero sono stati forniti

dal National Bureau of Statistics of the People's Republic of China, per il 2013 e dallo Shaanxi Provincial Bureau of Statistics, per il 2014.

3.3. Dati da telerilevamento

In questo studio è stato scelto di utilizzare sensori multispettrali perché allo stato attuale della ricerca scientifica e dello sviluppo tecnologico sono quelli che meglio si prestano alla determinazione delle grandezze biofisiche delle colture. Tali sensori risentono molto degli effetti atmosferici, i quali causano errori nella acquisizione dei dati fino ad impedirne l'acquisizione in caso di copertura da nubi.

Il monitoraggio delle coltivazioni necessita di una serie temporale continua di immagini, per questo studio in particolare è stato necessario monitorare l'area di interesse nel periodo di crescita della coltura più importante, compreso tra la levata e la maturazione, che nel caso specifico dell'area geografica osservata era tra marzo e maggio. Per garantire la corretta acquisizione dei dati necessari all'assimilazioni sono stati utilizzati i sensori a bordo di 3 differenti satelliti, l'americano Landsat 8 e i satelliti cinesi HJ1A e HJ1B. Tali satelliti, pur avendo caratteristiche orbitali molto diverse tra loro, montano sensori multispettrali con molte caratteristiche in comune. In questo modo si è potuto disporre di un numero sufficiente di dati ed allo stesso tempo si è potuto stimare le grandezze biofisiche con sensori simili tra di loro, in modo che i risultati e i relativi errori di misura, fossero confrontabili.

3.3.1. Landsat 8

Landsat 8 è un satellite americano per l'osservazione della terra lanciato nel febbraio del 2013. È posto in un'orbita eliosincrona ad una quota di 705 km con inclinazione di 98.2°, impiega circa 98.9 minuti per effettuare un giro della Terra e copre l'intero globo circa ogni 16 giorni, ad eccezione delle alte latitudini polari (http://landsat.usgs.gov/about_ldcm.php). È dotato di 2 sensori: Thermal Infrared Sensor (TIRS) e Operational Land Imager (OLI) che acquisiscono immagini al suolo su una superficie di 170 km per 185 km. Il primo è un sensore a due bande, una operante tra i 10.6 e gli 11.9 μm , l'altra tra gli 11.5 e i 12.51 μm , con una risoluzione di 100 m. Il secondo è un sensore multispettrale composto da nove bande, inclusa una banda pancromatica, operante tra gli 0.43 e i 1.38 μm (Tab. 3.4), con una risoluzione di 30 m, ad eccezione della pancromatica per la quale è di 15 m.

Tab. 3.4: Landsat 8 OLI, caratteristiche delle bande

Banda	Lunghezza d'onda (μm)	Risoluzione spaziale (m)
Band 1 - Coastal aerosol	0.43 - 0.45	30
Band 2 – Blue	0.45 - 0.51	30
Band 3 – Green	0.53 - 0.59	30
Band 4 – Red	0.64 - 0.67	30
Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
Band 6 - SWIR 1	1.57 - 1.65	30
Band 7 - SWIR 2	2.11 - 2.29	30
Band 8 - Panchromatic	0.50 - 0.68	15
Band 9 – Cirrus	1.36 - 1.38	30

3.3.2. Huan Jing-1

La costellazione di minisatelliti HJ-1 è inserita nell'ambito di un programma cinese per l'osservazione della terra del National Committee for Disaster Reduction and State Environmental Protection Administration (NDRCC/SEPA). È composta da 3 minisatelliti, HJ-1A, HJ-1B e HJ-1C, i primi due con a bordo sensori ottici, il terzo con un sistema SAR in banda S. I satelliti HJ-1A e HJ-1B sono stati lanciati nel settembre del 2008 su un'orbita complanare eliosincrona, con fase di 180° , ad una quota di 650 km, con un'inclinazione di 97.95° . I due satelliti, agendo in combinazione, permettono di ottenere immagini su una stessa zona con una frequenza di 2 giorni. Il satellite HJ-1A ha a bordo due camere ottiche CCD (WVC) ed una iperspettrale (HSI), anche HJ-1B è dotato delle camere WVC e HSI, in più è dotato di uno scanner multispettrale infrarosso (IRMSS). Il sensore CCD è dotato di 4 bande, con lunghezza d'onda che vanno da 0.43 a 0.90 μm ed una risoluzione spaziale di 30 m (Tab. 3.5).

Tab. 3.5: HJ 1 CCD, caratteristiche delle bande.

Bande	Lunghezza d'onda (μm)	Risoluzione spaziale (m)
Band 1 - Blu	0.43 - 0.52	30
Band 2 - Green	0.52- 0.6	30
Band 3 - Red	0.63 - 0.69	30
Band 4 - NIR	0.76 - 0.9	30

Le caratteristiche dei due sensori sono molto simili. La risoluzione al suolo è per entrambi 30 m e anche le Risposte Spettrali Relative (RSR) di ogni banda sono

abbastanza simili (Fig. 3.4). Le bande nello spettro visibile dei sensori CCD e OLI coprono approssimativamente le stesse lunghezze d'onda, con andamenti della RSR simili, invece la risposta operante nella regione del NIR ha un comportamento diverso per i due sensori, una banda più larga (0.76-09 μm) e un picco per lunghezze d'onda di circa 0.8 μm per la CCD, una larghezza di banda più stretta (0.85 – 0.88 μm) ed un andamento più uniforme per la OLI. Nel complesso la qualità dei dati ottenuti dai 2 sensori risulta comunque paragonabile (Zhao et al. 2013).

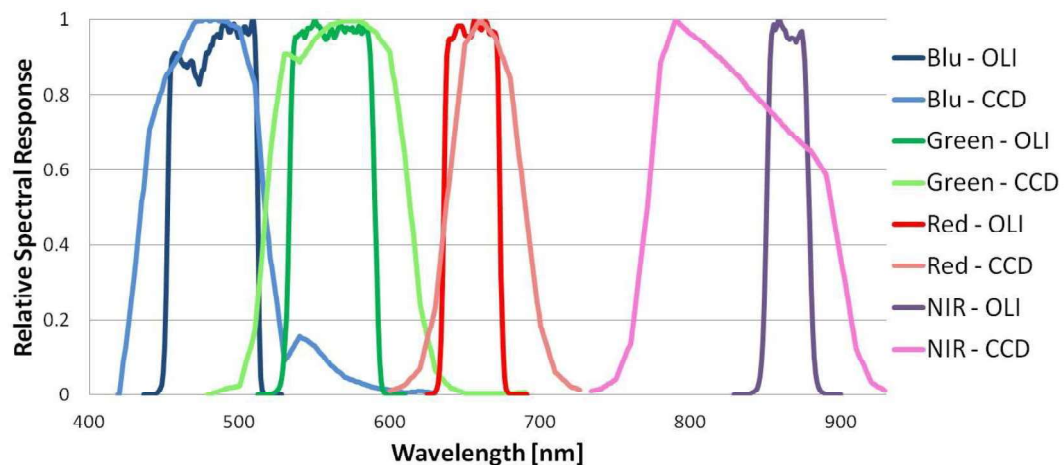


Figura 3.4: Confronto tra le Risposte Spettrali Relative (RSR) di OLI (Landsat 8) e CCD (HJ 1). I sensori CCD utilizzati nella costellazione HJ 1 sono complessivamente 4 con caratteristiche molto simili ma non identici. In questo grafico è mostrato a titolo di esempio la RSR esclusivamente per il sensore CCD1 a bordo di HJ 1A.

3.3.3. Pre-elaborazione delle immagini

Per pre-elaborazione si intendono tutte le operazioni che servono a rendere l'immagine fruibile per l'utente finale. In questo caso è necessario ottenere delle immagini georeferenziate, corrette dai disturbi atmosferici e con il dato finale in riflettanza, pertanto sono necessari processi di calibrazione radiometrica, correzione atmosferica e georeferenziazione, per ottenere le immagini adatte all'applicazione del algoritmo per la stima delle variabili biofisiche.

Le immagini Landsat sono fornite con correzione geografica ed atmosferica dal United States Geological Survey (USGS) degli USA. I dati delle immagini sono già forniti in riflettanza, pertanto non è stato necessario nessun lavoro di pre-elaborazione.

Per le immagini HJ è necessaria una correzione geografica, una calibrazione radiometrica e una correzione atmosferica. La georeferenziazione è stata fatta identificando su ogni immagini una serie di punti di cui erano note le coordinate. Successivamente l'immagine è stata “distorta” attorno ai punti noti usando il tool di ENVI “Warp from GCPs: Image-to-Map” (ENVI User's Guide, 2004). La

calibrazione radiometrica consiste nella conversione del dato dell'immagine, fornito in Digital Number (DN) nella radianza spettrale del sensore L ($\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$), usando la seguente equazione:

$$L = \frac{DN}{g} + L0 \quad (3.13)$$

dove g ($\text{W}^{-1} \text{m}^2 \text{sr} \mu\text{m}$) è il fattore di calibrazione e $L0$ ($\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$) è il parametro offset di calibrazione. Sia g che $L0$ sono differenti per ogni banda delle immagini, il loro valore è fornito nei metadata delle rispettive immagini. Non disponendo di dati sull'atmosfera raccolti in loco, la correzione atmosferica è stata eseguita usando il modulo Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH), incluso nel software ENVI. L'altitudine del sensore è stata impostata ad un valore medio di 650 km con elevazione del sito di interesse di 500 m (quota media dell'area rurale di Yangling). È stato utilizzato un modello atmosferico Winter Middle-Latitude e un modello Rural Aerosol, impostando la visibilità iniziale tra i 20 e i 30 km. Tale procedimento è stato ripetuto per tutte le immagini utilizzate in questo studio.

4. Stima di variabili biofisiche dal telerilevamento

4.1. Stato dell'arte delle metodologie di stima delle variabili biofisiche delle colture per mezzo di dati ottenuti da telerilevamento

L'uso del telerilevamento per la stima delle variabili biofisiche delle colture è un campo di studio molto vasto che racchiude tecniche e metodologie che hanno lo scopo di mettere in relazione i dati ottenuti misurando le radiazioni elettromagnetiche riflesse dalla superficie terrestre con lo stato di crescita della vegetazione, rappresentato da variabili biofisiche. Le tecniche e le metodologie impiegate sono svariate, legate principalmente alle caratteristiche dei sensori utilizzati e della variabile biofisiche che si intende stimare. In questa sezione la descrizione di questo argomento è stata limitata ai metodi e alle tecniche che impiegano sensori, per lo più multispettrali, che operano a lunghezze d'onda comprese tra la regione del visibile e del vicino infrarosso (0.4 e 2.5 μm) e che hanno come scopo la stima di grandezze biofisiche legate al canopy della vegetazione, quali Leaf Area Index e Canopy Cover.

Nell'intervallo di lunghezze d'onda compreso tra 0.4 e 2.5 μm , la quantità di energia elettromagnetica riflessa (radianza) sul manto della vegetazione (canopy) è influenzata principalmente dalle caratteristiche di 4 componenti: la radiazione sorgente, la vegetazione e il sensore. Nei casi qui analizzati la radiazione sorgente è sempre la radiazione solare, le cui caratteristiche sono note in funzione dell'ora di acquisizione e della posizione geografica, pertanto è semplicemente deducibile fornendo queste due informazioni. Le caratteristiche che descrivono la vegetazione sono riconducibili a tre categorie: proprietà ottiche degli elementi che caratterizzano la vegetazione, la disposizione di questi elementi nel canopy della vegetazione e le proprietà ottiche del suolo. Tali caratteristiche vengono rappresentate in modi differenti da parametri che variano a seconda del modello e del metodo che si vuole utilizzare. Il sensore è caratterizzato da parametri che descrivono la sua posizione al momento dell'acquisizione e da caratteristiche tecniche proprie del sensore come il numero di bande, la larghezza di ogni banda e la risposta spettrale di ogni banda (Dorigo et al., 2007).

Le proprietà spettrali degli elementi che caratterizzano la vegetazione, come foglie, steli e frutti, possono essere considerati i principali determinanti della

riflettanza del canopy e influenzano la forma complessiva della firma spettrale. (Jacquemoud et Baret, 1990).

Vari algoritmi sono stati sviluppati per ricavare le variabili biofisiche e biochimiche utilizzando i dati di riflettanza acquisiti da telerilevamento e possono essere suddivisi in due categorie: Approcci statistici/empirici o metodi fisici. I primi sono basati su metodi che cercano una relazione statistica tra le firme spettrali e le proprietà biofisiche o biochimiche del canopy, i secondi sui principi di propagazione delle radiazioni (Dorigo et al., 2007).

Tra i metodi statistici è largamente diffuso l'impiego di indici di vegetazione, calcolati come rapporto delle misure di riflettanza ad alcune lunghezze d'onda (generalmente rosso e vicino infrarosso). Una prima descrizione dettagliata sugli indici di vegetazione impiegati nel telerilevamento è stata data da Bannari et al. (1995), in cui vengono accuratamente mostrati i vantaggi e gli svantaggi dei più utilizzati indici di vegetazione impiegati fino al 1995. Tra questi è presente l'indice tuttora più utilizzato, il Normalized Difference Vegetation Index (NDVI; Rouse et al., 1973). Un'integrazione a tale lavoro è proposta da Basso et al. (2004), il quale descrive i principi biofisici degli indici di vegetazione focalizzando l'attenzione sulle applicazioni del telerilevamento per una gestione ottimale delle colture e del suolo. Più recentemente sono stati prodotti diversi studi che analizzano le relazioni tra NDVI, Biomassa e Leaf Area Index per diverse tipologie di colture, tra i quali citiamo: Glenn et al. (2008), Rinaldi et al., 2010, Viña et al. (2011), Boegh et al. (2013), Kang et al. (2016).

Tra i metodi statistici per la stima delle grandezze biofisiche da RS sono inclusi anche i metodi di regressione che collegano le informazioni spettrali con le variabili biofisiche e biochimiche misurate. Si citano a titolo di esempio: la Single Regression, dove la funzione di regressione è espressa da una semplice equazione lineare, con la limitazione di raggiungere una saturazione a livelli alti di densità di canopy (Baret et Guyot, 1991); la Stepwise Multiple Linear Regression, un'estensione della regressione lineare a funzioni contenenti due o più variabili indipendenti, usata particolarmente per sensori iperspettrali (Jacquemoud et al., 1995); la Partial Least Squares regression, una estensione multivariata del modello di regressione multiplo, usato in situazioni in cui si ha a disposizione un numero limitato di osservazioni (Huang et al. 2004).

Le tecniche di regressione menzionate in precedenza utilizzano tutte una relazione lineare tra le variabili biofisiche della vegetazione e firme spettrali misurate. Tra le relazioni non lineari quelle che hanno trovato un buon utilizzo nel campo del

RS sono certamente le Reti Neurali Artificiali (ANN, dall'inglese Artificial Neural Network; Combal et al., 2003 Atzberger et al., 2004). Si tratta di una tecnica sofisticata capace di modellare funzioni non lineari estremamente complesse. Tali metodi sono particolarmente efficaci perché capaci di auto-addestrarsi, in altre parole di calcolare i coefficienti e i parametri che ne descrivono l'architettura, partendo da un dataset detto appunto di addestramento, inizialmente composto da misure di campo (Combal et al., 2003).

I metodi che ricavano le grandezze biofisiche seguendo un approccio fisico consistono nell'invertire un modello di riflettanza del canopy per la stima di alcune proprietà della vegetazione. I modelli di riflettanza del canopy simulano le interazioni tra radiazione solare e gli elementi che costituiscono il canopy e il suolo usando le leggi fisiche, stabiliscono quindi una relazione tra il bersaglio della radiazione (la vegetazione e il suolo) e la firma spettrale. Il problema che si vuole risolvere è l'inversione di tali modelli non lineari, stabilire quindi una relazione tra le firme spettrali e il bersaglio. Differenti tecniche sono state usate per elaborare un metodo di inversione.

Le tecniche Iterative Optimization, ad esempio, le quali cercano la soluzione che minimizzi la differenza tra le firme spettrali misurate e quelle simulate. Questo tipo di metodo è utile quando si dispone di misure in campo di riflettanza, in letteratura si incontrano diversi lavori nei quali sono stati testati diversi tipi di vegetazione (Jacquemoud et al., 1995; Qiu et al., 1998; Bacour et al., 2002; Fang et al., 2003).

Altro metodo basato sull'approccio fisico è il Look Up Table (LUT). In questo metodo il modello di riflettanza del canopy è usato direttamente per generare un elevato numero di spettri possibili, dipendenti dalle diverse combinazioni dei parametri di input che descrivono le caratteristiche della vegetazione e del suolo (Kötz et al., 2004). L'applicazione del metodo LUT risulta essere più rapido del Iterative Optimization, anche se il tempo di computazione può aumentare significativamente se il numero di simulazioni è eccessivamente alto.

I metodi basati sulle Artificial Neural Network (ANN), elencati in precedenza tra i metodi di stima delle variabili biofisiche con approccio statistico, sono stati utilizzati anche come metodo di inversione dei modelli di trasferimento radiativo delle colture, sono infatti considerati come il principale tra i metodi ibridi, cioè che combinano gli approcci statistici e fisici (Liang, 2003). Come nell'approccio LUT, un modello di riflettanza del canopy è usato in modo diretto per costruire un dataset sintetico che

rappresenti una grande varietà di configurazioni suolo-vegetazione, la differenza è che mentre nei metodi con approccio LUT è ricercato lo spettro simulato che più si avvicina a quello misurato, nei metodi ANN la distanza tra le variabili del canopy viene minimizzata. I metodi basati sulle ANN, una volta effettuato l'addestramento della rete, sono più rapidi dei metodi Iterative Optimization e LUT, inoltre sono meno sensibili alle incertezze del modello (Combal et al., 2003). Inoltre tale metodo consente di stimare contemporaneamente e in modo indipendente variabili diverse, come LAI e CC (Baret et al., 2006).

Di particolare interesse è il metodo sviluppato da Bacour et al. (2006), nel quale le Reti Neurali Artificiali (ANN) sono utilizzate per invertire il modello di trasferimento radiativo PROSAIL, nato dall'accoppiamento di due modelli preesistenti, i modelli Prospect e SAIL (Jacquemoud et al., 2009), dimostrato efficace nella stima del LAI e del CC da molti studi presenti in letteratura (e.g. Baret et al. 2007; Verger et al., 2011).

Per il raggiungimento degli obiettivi di questa tesi era necessario stimare il LAI ed il CC utilizzando dati acquisiti tramite sensori multispettrali (4 bande, dal blu al vicino infrarosso). Il metodo basato sull'addestramento di una rete neurale tramite le simulazioni fatte dal modello di trasferimento radiativo PROSAIL, è stato ritenuto il più idoneo al raggiungimento degli obiettivi, in quanto, dall'analisi bibliografica, era quello che meglio stimava il CC e il LAI non disponendo di misure in campo per la calibrazione di metodi empirici. Questo metodo è largamente impiegato nella stima delle variabili biofisiche per mezzo di sensori multispettrali (Baret et Buis, 2008). Inoltre, consente di prendere in esame, a costi computazionali ragionevoli, un dataset composto da un elevato numero di simulazioni, utilizzato sia per l'addestramento della rete che stima il LAI sia per quella che stima il CC.

Di seguito tale metodo è descritto più dettagliatamente, evidenziando le impostazioni di input utilizzate in questo studio. I risultati ottenuti sono stati validati con misure in campo, in parte per fornire un'ulteriore valutazione di questo metodo, in parte per ottenere una stima, seppur approssimativa, dell'errore presente in tali stime, dato che questa informazione è richiesta dall'algoritmo di assimilazione utilizzato.

4.2. Algoritmo di stima delle variabili biofisiche PROSAIL-ANN

Il Leaf Area Index (LAI) e il Canopy Cover (CC) sono stati calcolati usando l'algoritmo sviluppato da Bacour et al. (2006). Questo algoritmo, da qui in poi

chiamato PROSAIL-ANN, converte le riflettanze misurate da sensori multispettrali in variabili biofisiche e biochimiche della vegetazione, in questo caso specifico LAI e CC.

L'algoritmo PROSAIL-ANN utilizza un dataset di simulazioni effettuate dal modello di trasferimento radiativo PROSAIL, modello che accoppia i preesistenti modelli PROSPECT e SAIL, per addestrare una serie di reti neurali, una per ogni variabile biofisica ricercata. Tale procedimento può essere suddiviso in tre step:

1. generazione del database di simulazioni da usare per l'addestramento (training database, parte rossa in Fig. 4.1);
2. addestramento delle reti neurali (parte blu in Fig. 4.1)
3. applicazione delle reti per la stima di LAI e CC (parte verde in Fig. 4).

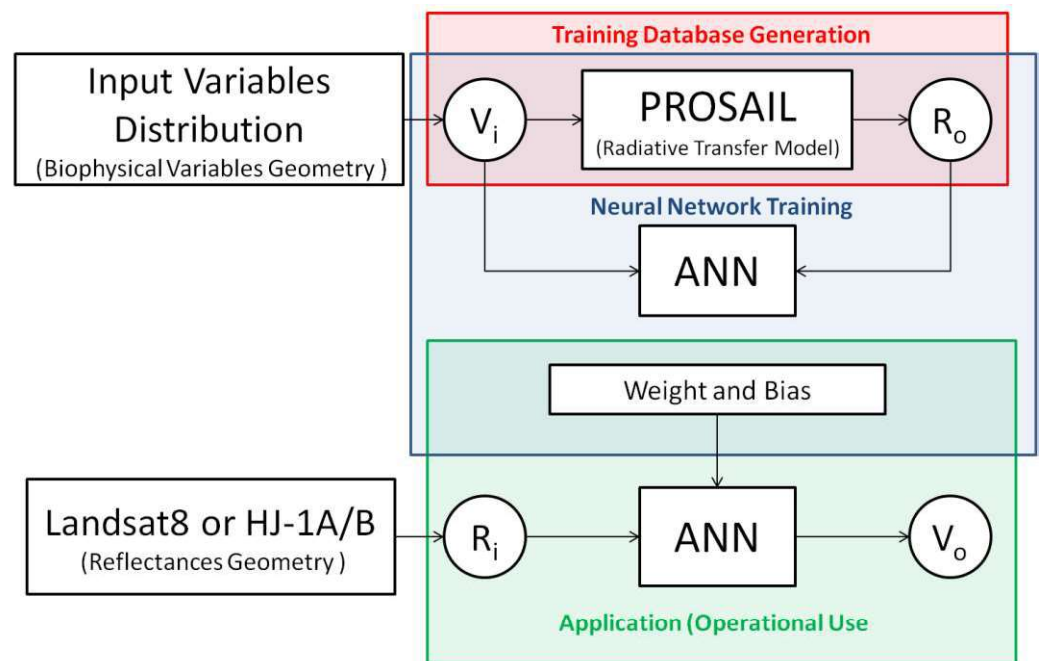


Fig. 4.6: Schema del funzionamento dell'algoritmo PROSAIL-ANN. V_i : Variabili di input (LAI e CC in questo studio), R_o : Riflettanze simulate, output del modello PROSAIL, R_i : Riflettanze misurate da satellite, input della fase di applicazione delle reti, V_o : Variabili di output (LAI e CC stimati) (tratto da Veloso, 2014).

4.2.1. Generazione del database di addestramento

Il database di addestramento è composto dalle simulazioni delle firme spettrali di un'ampia varietà di scenari. Ogni scenario è composto da una combinazione di parametri che descrivono le caratteristiche geometriche e biofisiche della vegetazione e del suolo, al quale il modello PROSAIL associa una firma spettrale.

PROSAIL è una combinazione dei modelli PROSPECT (Jacquemoud et al., 2009) e SAIL (Verhoef, 1984). Il primo simula le riflettanze e le trasmittanze delle foglie,

che sono usate come input nel modello SAIL. PROSPECT richiede parametri di input che descrivono le proprietà delle foglie: parametro della struttura del mesofillo (N), contenuto di clorofilla (Cab), contenuto di acqua (Cw), contenuto di pigmento marrone (Cbp) e contenuto di materia secca (Cdm). Il modello SAIL fornisce le riflettanze direzionali a livello di manto vegetale (canopy), e permette di calcolare variabili derivate come la *frazione di radiazione fotosinteticamente attiva assorbita* (fAPAR) ed il CC. I parametri di input che descrivono la struttura del canopy della vegetazione sono il LAI e l'inclinazione media delle foglie (ALA). Infine il parametro che descrive la riflettanza del suolo di sfondo è il coefficiente di luminosità (Bs). Gli input riguardanti la configurazione geometrica del satellite includono: posizione geografica del bersaglio, ora e giorno dell'anno dell'acquisizione, angolo di zenit e di azimut sia del sole che del satellite, riconducibili a tre parametri (Fig. 4.2): angolo di zenit del sole (θ_s), angolo di zenit del satellite (θ_v) e l'angolo di azimuth relativo ($\Delta\phi$). Queste informazioni sono differenti per ogni immagine, per questo è stato necessario eseguire un addestramento per ogni immagine utilizzata.

Per tutti i parametri sopracitati è stato definito un intervallo di variazione. Tale intervallo di variazione è stato suddiviso in sotto intervalli, definiti dal numero di classi assegnato a ogni parametro. In questo modo ogni scenario (combinazione delle caratteristiche del suolo e della vegetazione) è rappresentato dalla combinazione dei valori appartenenti ad ogni classe designata. I valori dei parametri che delimitano ogni classe sono stati calcolati secondo una distribuzione gaussiana, definendo per ogni parametro: il limite minimo (min), quello massimo (Max), la moda (Mode) e la deviazione standard (Std). Tali valori sono stati impostati seguendo le indicazioni fornite dallo studio di Bacour et al. (2006) per il frumento tenero, eccetto per la moda del LAI. Per quest'ultimo si è usato come riferimento osservazioni di LAI fatte in campo, e si è impostato la moda di LAI a 1.5 per le immagini acquisite tra febbraio e la prima settimana di marzo, a 2.5 per le immagini acquisite tra marzo e l'inizio di aprile e infine a 4 per le immagini acquisite tra aprile e maggio.

Il numero di simulazioni totale è rappresentato dal prodotto dei numeri di classe (N_C) di ogni parametro, che in questo caso specifico è di 98304.

Tab. 6: Definizione dei parametri di input del modello di trasferimento radiativo PROSAIL, usato per generare il Training Database. La distribuzione usata per ogni parametro è la "Truncated Gaussian Function" definita da: limite inferiore (min) e superiore (Max), moda (Mode) e deviazione standard (Std). N_C è il numero di classi in cui ogni distribuzione è suddivisa.

	Variable	min	Max	Mode	Std	N_C
--	----------	-----	-----	------	-----	-----

Canopy	LAI	0.0	10.0	1.5 / 2.5 / 4	4.0	6
	Average Leaf Inclination (ALA)	30	80	60	20	4
Leaf	N	1.20	2.20	1.50	0.30	4
	Chlorophyll Content (Cab, µg.m⁻²)	20	90	45	30	4
	Dry matter content (Cdm, g.m⁻²)	0.0030	0.0110	0.0050	0.0050	4
	Relative water content (Cw)	0.60	0.85	0.75	0.08	4
	Brown Pigment index (Cbp)	0.00	2.00	0.00	0.30	4
Soil	Brightness Coefficient (Bs)	0.50	3.50	1.20	2.00	4

4.2.2. Addestramento delle Reti Neurali Artificiali (ANN training)

Le reti neurali artificiali sono caratterizzate dalle funzioni di trasferimento, rappresentate come tipo di neurone e dalla loro architettura. Scopo dell'addestramento è di stabilire una relazione tra gli input (le riflettanze) e gli output (LAI e CC) calcolando dei coefficienti moltiplicativi e addizionali (Weight and Bias) che caratterizzano le funzioni di regressione prese in esame. In questo lavoro è stato applicato un metodo di addestramento detto back-propagation (Sathyanarayana, 2014), con due strati nascosti. Il primo strato è composto di cinque neuroni con una funzione di trasferimento a sigmoide (Han et al., 1995) mentre il secondo strato è composto di un solo neurone con una funzione di trasferimento lineare (Debes et al., 2005). La combinazione della funzione di trasferimento sigmoide con la lineare è in grado di riprodurre qualsiasi tipo di funzione (Beale et al., 2013).

Prima dell'addestramento con le reti neurali, le riflettanze generate dalle simulazioni con PROSAIL sono ricampionate alle bande che definiscono le caratteristiche del sensore che si utilizza attraverso le risposte spettrali relative di ogni singola banda. Viene inoltre aggiunto alle riflettanze simulate dal modello un errore. Questo errore è rappresentato da quattro parametri, due moltiplicativi e due addizionali e includono gli errori dovuti al rumore strumentale, alla calibrazione radiometrica e alla correzione atmosferica. Esso è espresso dalla seguente equazione:

$$\rho_{train}(\lambda) = \rho_{sim}(\lambda) \cdot [1 + N(0, \sigma_{rel}(\lambda)) + N(0, \sigma_{rel}(all))] + N(0, \sigma_{abs}(\lambda)) + N(0, \sigma_{abs}(all)) \quad (4.14)$$

dove $\rho_{sim}(\lambda)$ e $\rho_{train}(\lambda)$ sono rispettivamente i valori di riflettanze per la banda generica λ (simulate da PROSAIL) e il valore di riflettanza compreso di errore, $N(0, \sigma)$ è una distribuzione normale con media 0 e deviazione standard σ , σ_{rel} e σ_{abs} sono le incertezze assoluta e relative applicata alla singola banda (λ) o a tutte le bande (all) (Verger et al. 2011). I valori per le incertezze assoluta e relative usate in questo lavoro sono: $\sigma_{abs}(\lambda)$

$= \sigma_{\text{abs}}(\text{all}) = 0.05$ e $\sigma_{\text{rel}}(\lambda) = \sigma_{\text{rel}}(\text{all}) = 5\%$. Con riferimento alla Figura 4.2, le riflettanze $\rho_1, \rho_2, \rho_3, \rho_4$ sono quelle comprendenti l'errore.

In questo studio sono stati impiegati due diversi sensori, il CCD a bordo dei satelliti HJ1A e HJ1B e OLI a bordo di Landsat8. Per il sensore OLI sono state considerate solo 4 bande (blu, verde, rosso e NIR), perché molto simili alle 4 bande in cui lavora il sensore CCD, in questo modo i due sensori hanno caratteristiche molto simili ed è stato possibile considerare la stessa incertezza sulle misure.

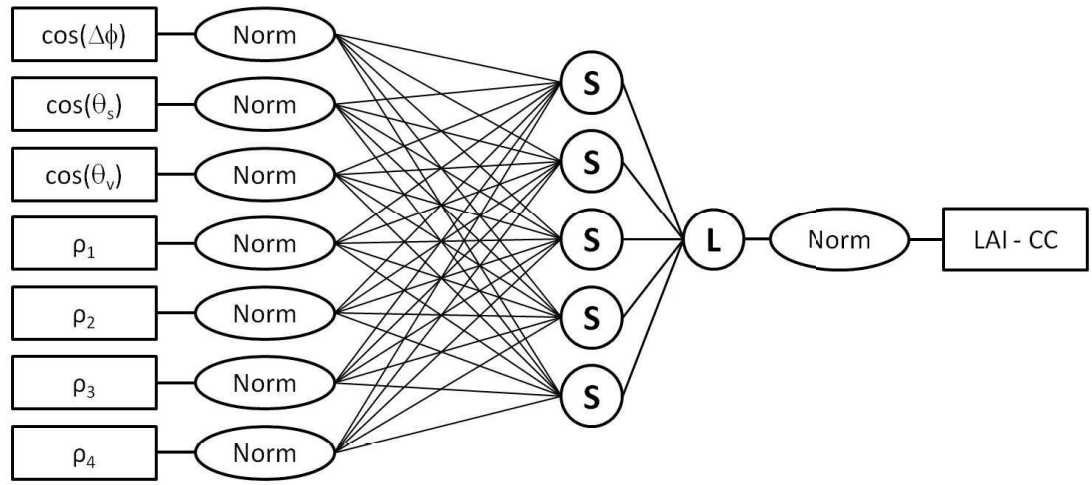


Fig. 4.7: Architettura della rete neurale artificiale sviluppata per la stima di LAI e CC considerando le bande del blu, del verde, del rosso e del NIR (pn) dei sensori a bordo dei satelliti HJ1A/B e Landsat8, e i tre angoli che definiscono la geometria delle osservazioni (angolo di zenit del sole θ_s , angolo di zenit del satellite θ_v e l'angolo di azimuth relativo $\Delta\phi$). La rete è composta da due strati nascosti, il primo composto da 5 neuroni con funzione sigmoide (S), il secondo composto da un neurone con funzione lineare (L).

4.2.3. Applicazione delle Reti Neurali Artificiali

Una volta effettuato l'addestramento delle reti neurali e calcolati Weight e Bias delle funzioni di trasferimento, la relazione tra le riflettanze misurate tramite RS e le variabili biofisiche è stabilita. Le immagini acquisite e corrette come descritto nel paragrafo 3.3.3 vengono utilizzate come input delle reti neurali ottenendo così una serie temporale di mappe di LAI e di CC. Tali mappe hanno la risoluzione spaziale a 30 m, la stessa delle immagini acquisite da telerilevamento.

4.2.4. Risultati e discussione

I valori di LAI e CC ottenuti tramite l'applicazione dell'algoritmo PROSAIL-ANN ai dati acquisiti tramite i sensori a bordo dei satelliti Landsat8 e HJ1A/B sono stati confrontati con le misure di LAI e CC acquisite come descritte nel paragrafo 3.2. Il confronto è stato limitato alle immagini acquisite in date vicine alle misure in campo.

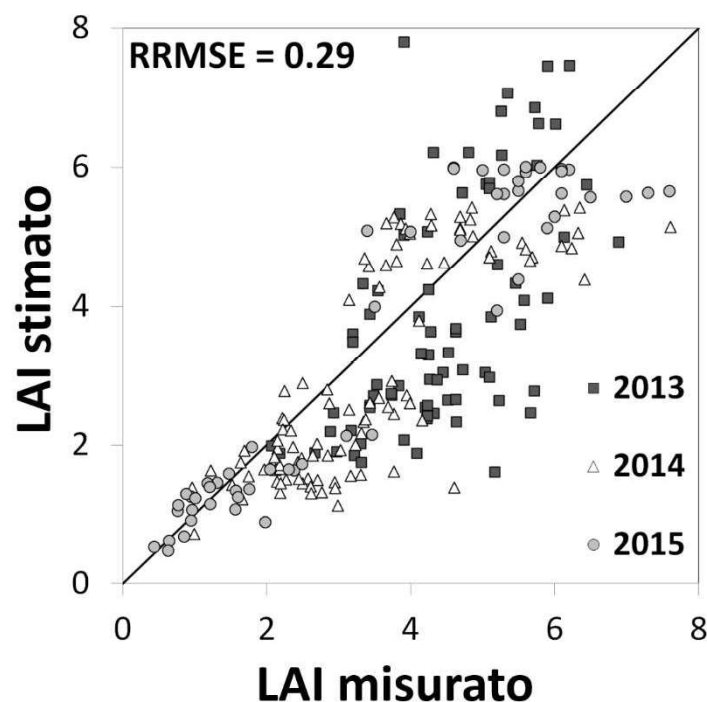


Fig. 4.8: Validazione dei risultati confrontando il LAI ottenuto utilizzando come input dell'algoritmo PROSAIL-ANN i dati acquisiti dai sensori a bordo dei satelliti HJ1A, HJ1B e Landsat8 con il LAI misurato in situ durante le campagne di misura effettuate a Yangling tra il 2013 e il 2015. È mostrata anche la retta 1:1.

In Fig. 4.3 è mostrato il confronto tra il LAI stimato combinando dati RS con l'algoritmo PROSAIL-ANN e LAI misurato durante le campagne di misura. Il Relative Root Mean Square Error (RRMSE), per i 3 anni è del 29%. L'errore è maggiore per valori di LAI elevati. Infatti, per LAI compresi tra 0 e 2 la differenza tra stime e misure è ridotta, mentre per LAI compresi tra 3 e 4 si ottiene un errore maggiore, per valori maggiori di 4 l'errore continua ad essere elevato ma non quanto nell'intervallo precedente. Questi risultati indicano che l'algoritmo tende a sottostimare il LAI, in particolar modo per il 2013, anno in cui il RRMSE risulta essere più alto che negli altri anni esaminati. I valori RMSE, RRMSE e r^2 complessivi e per ogni singolo anno sono mostrati in Tab. 4.2, dove è evidente che i risultati peggiori sono ottenuti per il 2013.

Tab. 4.7: RMSE, RRMSE e r^2 di LAI (in alto) e CC (in basso), per ogni anno (prime tre righe) e complessivi (ultima riga, in grassetto).

LAI			
	RMSE	RRMSE	r^2
2013	1.48	32%	0.38
2014	1.03	30%	0.65
2015	0.75	23%	0.87
	1.09	29%	0.61

Tutti gli anni	1.15	30%	0.64
CC			
	RMSE	RRMSE	r^2
2013	7.27	10%	0.51
2014	8.05	12%	0.59
2015	5.06	9%	0.89
Tutti gli anni	7.17	11%	0.72

In Fig. 4.4 sono mostrati i risultati ottenuti per la stima del CC. In questo caso il RRMSE è del 10%, più basso quindi che per il LAI. La saturazione di cui soffriva il LAI stimato tramite l'algoritmo PROSAIL-ANN non avviene nel caso del CC stimato tramite lo stesso algoritmo, anche se anche in questo caso, seppur con un errore più basso, l'algoritmo tenda a sottostimare tale variabile biofisica. Anche in questo caso (come mostrato in Tab. 4.2), il 2015 risulta l'anno con il valore di RRMSE più basso mentre il 2013 quello con il valore più alto.

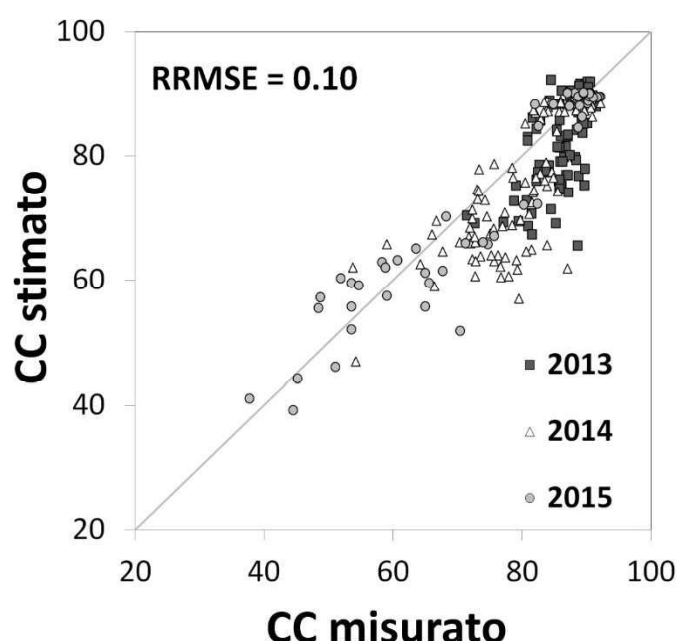


Fig. 4.9: Risultati di validazione per il CC ricavato dall'applicazione dei dati HJ1A/B e Landsat8 all'algoritmo PROSAIL-ANN usando misure di campo effettuate sull'area rurale di Yangling tra il 2013 e il 2016. Le misure di campo di LAI sono state convertite in CC usando l'equazione 3.5. È mostrata anche la retta 1:1.

Questa distribuzione dell'errore è dovuta in parte alle caratteristiche dell'algoritmo PROSAIL-ANN, per il quale le stime di LAI del frumento tenero peggiorano all'aumentare del valore di LAI (Verger et al., 2011). Un'altra causa di errore è da ricercare nella qualità delle immagini acquisite. L'area di studio è

fortemente caratterizzata dalla presenza di nuvole e foschia, per questo la probabilità di acquisire immagini di buona qualità è molto bassa, specialmente nel periodo primaverile quando le precipitazioni iniziano ad essere frequenti. La correzione atmosferica per queste immagini è fondamentale in quanto essa influenza fortemente la riflettanza e conseguentemente la stima del LAI e del CC. Nel caso del CC il valore di RRMSE tra misure e stime è più piccolo che per il LAI, ciò è dovuto al fatto che il CC è meno soggetto a saturazione ad alti valori. Il periodo di osservazione è compreso tra marzo e maggio, quando la variazione di LAI è tra 0.5 e 7, corrispondente a una variazione di CC tra il 40% e il 91%. Già dalla fine di marzo il LAI è superiore a 3, corrispondente a un CC di circa 80 %, per questo la gran parte delle misure varia tra 80% e 91%. In questo intervallo di variazione di valori così piccolo, l'algoritmo PROSAIL-ANN stima il CC con accuratezza maggiore.

In conclusione, le variabili biofisiche delle colture, più specificamente LAI e CC, da utilizzare in seguito nelle procedure di assimilazione, sono state stimate per mezzo di dati acquisiti tramite RS. Il LAI è stimato con un RRMSE piuttosto elevato, circa del 29%, mentre il CC risulta meglio stimato tramite questo algoritmo, consentendo un RRMSE del 10%, nonostante la qualità non ottimale delle immagini di input, confermando di essere tra le variabili biofisiche delle colture meglio stimate da RS (Verger et al., 2012).

5. Modelli di accrescimento delle colture

In questo studio è stato scelto di usare due diversi modelli di accrescimento delle colture per rispondere alle esigenze dei metodi di assimilazione per calibrazione e per aggiornamento. I modelli utilizzati sono Aquacrop (Steduto et al., 2008) e il Simple Algorithm For Yield (SAFY; Duchemin et al., 2008), per i quali di seguito è fornita una breve descrizione.

5.1. Aquacrop

Aquacrop è un modello sviluppato dalla FAO (Food and Agriculture Organization), per aggiornare la procedura di stima della risposta delle colture alla disponibilità idrica. È un modello detto infatti water-driven, proprio perché è l'acqua traspirata dalla coltura la variabile chiave che “guida” le variabili biofisiche che descrivono la crescita della vegetazione.

L'equazione di partenza da cui tale modello si sviluppa è quella proposta da Doorenbos et Kassam (1979) per mettere in relazione la produzione con l'evapotraspirazione:

$$\left(\frac{Y_x - Y_a}{Y_x} \right) = k_y \left(\frac{ET_x - ET_a}{ET_x} \right) \quad (5.15)$$

dove Y_x e Y_a sono la produzione massima e attuale, ET_x e ET_a sono l'evapotraspirazione massima e attuale, infine k_y è il fattore di proporzionalità tra le perdite relative di produzione e la riduzione relativa di evapotraspirazione.

Nel modello Aquacrop si è voluto migliorare l'equazione (4.1) separando la ET_a in evaporazione dal suolo (E_s) e in traspirazione dalle piante (T_a), e scomponendo la produzione finale (Y) in biomassa (B) e harvest index (HI). Tali cambiamenti conducono alla formulazione della seguente equazione, nucleo del modello Aquacrop:

$$B = WP \cdot \sum T_a \quad (5.16)$$

dove T_a è la traspirazione della coltura espressa in mm e WP è un parametro di produttività dell'acqua, espresso in kg di biomassa per m² e per mm di acqua traspirata nel periodo in cui la biomassa è prodotta. Altro cambiamento sostanziale introdotto dall'equazione (5.2) è la possibilità di utilizzare dati giornalieri, un periodo più vicino alla scala di tempo delle risposte delle colture al deficit idrico, rispetto alla scala

stagionale utilizzata nell'equazione (5.1). In Fig. 5.1 è mostrato schematicamente l'evoluzione di Aquacrop dall'equazione (5.1).

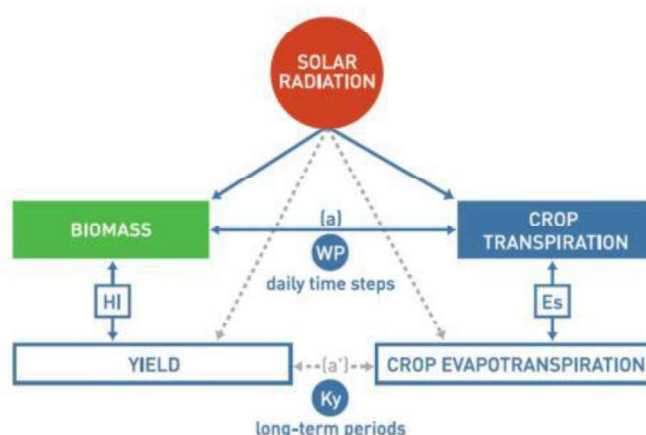


Fig. 5. 10: Evoluzione di Aquacrop dall'equazione (5.1), basata sull'introduzione di 2 fasi intermedie: la separazione dell'evaporazione del suolo (E_s) dalla traspirazione delle piante (T_a) e la separazione della produzione totale (Y) in Biomassa (B) e Harvest Index (HI). La relazione a' collega Y con l'evapotraspirazione delle colture per mezzo del parametro k_y , in accordo l'equazione (5.1), e per applicazione generalmente su scala temporale stagionale. La relazione a lega B a T_a per mezzo di WP , in accordo con l'equazione (5.2), utilizzando un intervallo di tempo giornaliero (da Steduto et al., 2008).

L'equazione (5.2) è il nucleo del modello Aquacrop ed è inserita in un insieme di componenti aggizionali che vanno a definire l'algoritmo che simula la crescita delle colture a partire da informazioni di input climatiche, sulla coltura e sull'ambiente in cui la coltura è inserita (come il suolo o la gestione della coltivazione). Come per tutti i modelli di accrescimento delle colture, Aquacrop ha una struttura che spazia su tutto il sistema suolo-pianta-atmosfera. Esso include: il suolo, con il suo bilancio idrico; la pianta, con i suoi processi accrescimento, sviluppo e produzione; l'atmosfera, con il suo regime termico, le precipitazioni, richiesta di evaporazione e concentrazione di anidride carbonica. In aggiunta a questo sono stati inseriti alcuni aspetti sulla gestione della coltura, come la fertilizzazione e l'irrigazione, i quali influenzano il bilancio idrico, lo sviluppo della vegetazione e la produzione finale.

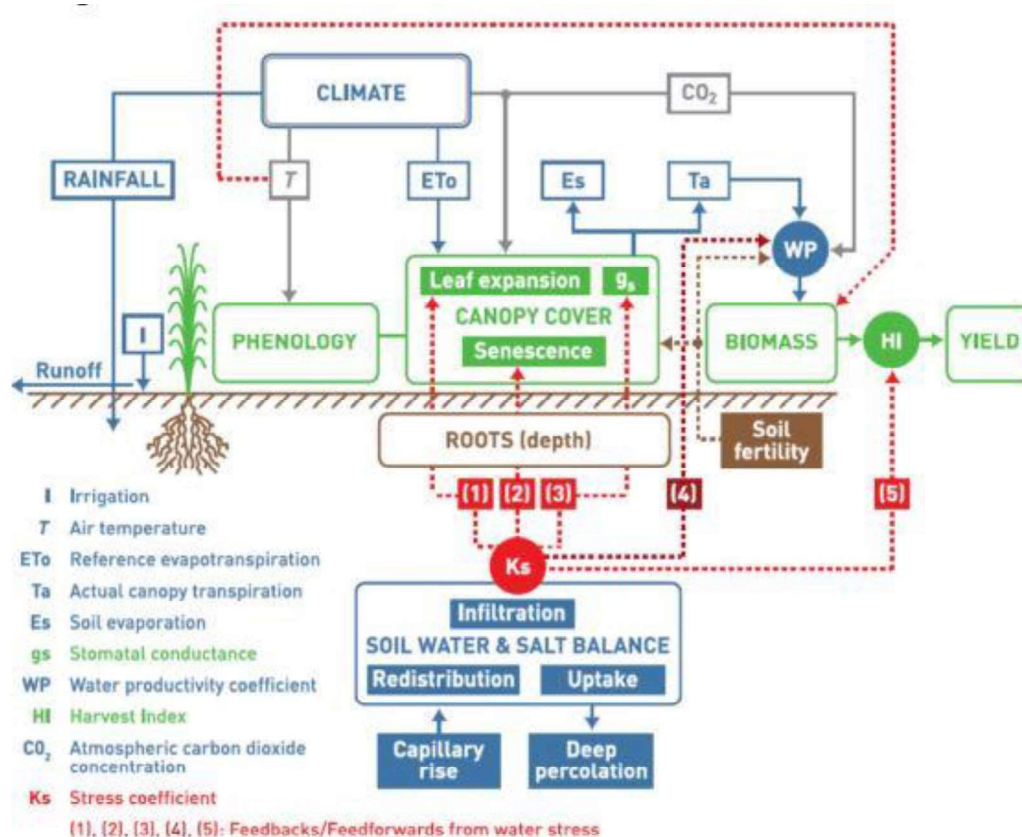


Fig. 5.11: Schematizzazione dei processi che compongono il modello Aquacrop, nel quale sono differenziate per colore le componenti che compongono il sistema suolo (marrone) – atmosfera (blu) – pianta (verde). (Da Steduto et al., 2008)

In Fig. 5.2 è mostrata una schematizzazione del modello Aquacrop, suddiviso nelle componenti che caratterizzano il processo di accrescimento delle colture. Una descrizione più dettagliata dei processi che compongono il modello Aquacrop è fornita da Steduto et al. (2009) e Raes et al. (2009), di seguito è fornita una descrizione introduttiva degli input necessari all'uso del modello.

5.1.1. Variabili di Input del modello Aquacrop

Le caratteristiche dell'atmosfera in cui è inserita la coltura sono descritte dalle componenti climatiche di Aquacrop. Esse sono riconducibili a cinque variabili di input: temperatura dell'aria giornaliera massima e minima, precipitazione giornaliera (in mm), richiesta giornaliera di evaporazione dell'atmosfera espressa dall'evapotraspirazione di riferimento (ET₀) e la media annuale di concentrazione di anidride carbonica nell'atmosfera. I primi quattro sono ricavati da stazioni meteorologiche di terra nelle vicinanze della regione di interesse, la concentrazione di anidride carbonica è ricavata dai dati forniti dal Mauna Loa Observatory nelle Hawaii. Tutti i dati climatici utilizzati in questo studio sono stati ottenuti come descritto nel paragrafo 3.1.

In Aquacrop la coltura è caratterizzata da cinque componenti principali a cui sono associate altrettante risposte dinamiche: fenologia, canopy cover (CC), profondità delle radici, produzione di biomassa e produzione di granella. Tutte le dinamiche sopraelencate sono relazionate tra loro e si sviluppano in dipendenza l'una dell'altra. L'unica eccezione è per la produzione di granella, legata alla biomassa dalla semplice equazione lineare:

$$Y = B \cdot HI \quad (5.17)$$

Anche se il modello Aquacrop utilizza un parametro di Harvest Index (HI), non calcola le suddivisioni della biomassa nei vari organi (foglie, radici, ecc.): la produzione di biomassa è dipendente esclusivamente dell'espansione della canopy e della profondità delle radici (Steduto et al. 2008).

Il ciclo fenologico delle colture è funzione del tempo termico, rappresentato dalla somma termica (Growing Degree Day, GDD) per mezzo della seguente equazione:

$$GDD = \frac{(T_{Max} - T_{min})}{2} - T_{base} \quad (5.18)$$

dove: GDD è il valore in gradi della somma termica, T_{Max} e T_{min} sono rispettivamente la temperatura giornaliera massima e minima e T_{base} è la temperatura basale, ovvero la temperatura sotto la quale lo sviluppo della pianta si arresta. Aquacrop prevede l'impiego anche di una temperatura sopra la quale la crescita si arresta, ma non è mai stata raggiunta nei casi presi in esame in questo studio.

Aquacrop distingue 3 tipi di colture principali: colture da granella o cereali, produzione di radici o tuberi, coltivazioni di vegetali a foglia. Ognuno di questi tipi ha le sue corrispondenti fasi del ciclo di sviluppo e accrescimento. La variazione genetica delle specie e delle coltivazioni può essere implementata nel modello attraverso la variazione delle varie fasi di sviluppo, o anche attraverso il tasso di espansione della canopy, il tasso di sviluppo delle radici, il parametro di produttività dell'acqua ed altri fattori di risposta alle condizioni ambientali.

Il manto vegetale (canopy) in Aquacrop è una caratteristica fondamentale essendo il mezzo attraverso il quale viene determinata la quantità di acqua traspirata e di conseguenza la quantità di biomassa prodotta. L'espansione della canopy è espressa attraverso la variabile CC. In condizione di assenza di stress, l'espansione dall'emergenza allo sviluppo completo della canopy segue una crescita esponenziale durante la prima metà del ciclo e una decrescita esponenziale nella seconda metà. Le equazioni che rappresentano lo sviluppo del CC sono:

$$CC = CC_0 e^{CGC \cdot t} \quad (5.19)$$

nella fase di crescita e:

$$CC = CC_x - (CC_x - CC_0) \cdot e^{-CGC \cdot t} \quad (5.20)$$

In fase di decrescita, CC rappresenta il Canopy Cover al tempo t, CC_0 è il canopy cover all'istante iniziale, CGC è il coefficiente di crescita del canopy espresso in frazione di CC per giorno o per grado-giorno, CC_x è il canopy cover massimo e t è il tempo espresso in giorni o gradi giorni. Dopo il completamento dello sviluppo della vegetazione, la canopy può avere un periodo di lunghezza variabile prima di entrare in fase di senescenza. Durante questo periodo l'effetto di fattori che riducono la capacità fotosintetica (es. stress idrico) è considerato attraverso l'introduzione di coefficienti che diminuiscono il CC_x . Completata questa fase, inizia la fase di senescenza in cui il CC decresce.

In Aquacrop le risposte al deficit idrico sono espresse in: riduzione del tasso di espansione, riduzione del coefficiente di accrescimento CGC ed accelerazione della senescenza. Queste tre risposte sono formalizzate tramite lo stesso algoritmo. La carenza di acqua è quantificata attraverso un coefficiente di stress idrico (K_s) che varia tra 1, condizione di assenza di stress, e 0, condizione di stress completo. Lo stress avviene quando si raggiunge un limite soglia di acqua contenuta nella zona di suolo in cui sono presenti le radici. Tale limite (p , compreso tra 0 e 1), raggiunge il valore soglia per una specifica risposta al deficit idrico, K_s è espresso mediante la seguente equazione:

$$K_s = \frac{1 - (e^{p \cdot f_{shape}} - 1)}{(e^{f_{shape}} - 1)} \quad (5.21)$$

con f_{shape} parametro che stabilisce la forma della funzione K_s , la quale può essere una tra le tre funzione in Fig. 5.3

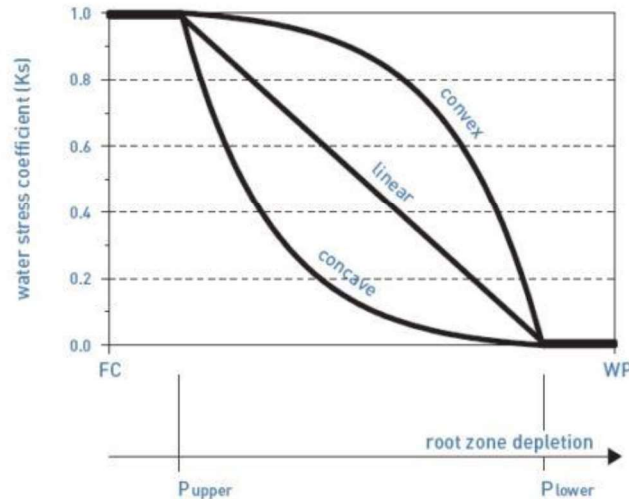


Fig. 5.12: Possibili andamenti di K_s in funzione della diminuzione del contenuto di acqua relativo. La funzione assume andamento lineare quando $fshape = 1$, concava per $fshape < 0$ e convessa per $fshape > 0$ (Steduto et al., 2008).

L'equazione centrale che governa la crescita delle colture nel modello Aquacrop è la 5.2, pertanto tutti i processi secondari che contribuiscono alla produzione di biomassa vengono incorporati nel coefficiente di produttività di acqua (WP) (Steduto et al. 2007). Il parametro WP di Aquacrop è normalizzato per ET_0 e la concentrazione di anidride carbonica in atmosfera. Esso può variare moderatamente in risposta al regime di fertilità e rimane costante in presenza di deficit idrico, eccetto quando è raggiunto un grave stress idrico. La normalizzazione di WP nei confronti delle condizioni climatiche, rende il modello applicabile a diverse località e stagioni.

Il suolo nel modello Aquacrop è configurato come un sistema funzione di profondità variabile permettendo fino a 5 strati di diversa tessitura lungo il profilo.

Per ogni tessitura del terreno il modello associa alcune caratteristiche idrauliche mentre nel caso di combinazioni inserite dall'utente le stima attraverso funzioni di pedotrasferimento. Le caratteristiche idrauliche includono: il coefficiente di drenaggio (τ), conduttività idraulica a saturazione (k_{sat}), contenuto di acqua volumetrico a saturazione (θ_{sat}) e punto di appassimento (θ_{WP}).

In Aquacrop le componenti inerenti la gestione agronomica possono essere suddivise in due categorie: gestione di campo e gestione dell'acqua. Nella prima sono incluse tutte le informazioni inerenti la fertilità mentre nella seconda sono incluse tutte le variabili riguardanti l'irrigazione.

5.2. Simple Algorithm For Yield (SAFY)

La versione originale del modello SAFY è basata sulla teoria formulata da Monteith (1977), che lega la produzione di fitomassa totale alla frazione di radiazione solare fotosinteticamente attiva assorbita (APAR) attraverso un coefficiente d'efficienza d'uso della luce. La prima differenza sostanziale con Aquacrop è quindi che nel caso di SAFY la variabile climatica di input principale non è l'evapotraspirazione ma la radiazione solare giornaliera.

Il modello SAFY simula principalmente tre variabili dinamiche, dal giorno dell'emergenza a quello del termine della senescenza, utilizzando un intervallo di tempo giornaliero. Tali variabili dinamiche sono: biomassa epigea secca (indicata come DAM, dall'inglese Dry Above-ground Mass), il Green Leaf Area Index (al quale in questo lavoro ci si è riferito semplicemente come LAI, ma indicato nel modello come GLAI) e la produzione di granella (GY, Grain Yield).

La produzione giornaliera di DAM (di seguito indicata come ΔDAM) è legata alla radiazione solare globale (R_g) per mezzo di tre parametri:

- l'efficienza climatica ϵ_c , che è il rapporto tra la radiazione fotosinteticamente attiva e la radiazione globale;
- Frazione della luce assorbita dalla canopy ϵ_l ;
- efficienza di conversione della luce in biomassa $ELUE$ ($g \cdot MJ^{-1}$), frazione dell'energia foto sintetica prodotta come DAM dall'APAR;

Oltre che da questi tre parametri, la produzione di biomassa DAM è regolata da una funzione di stress termico $F_T(T_a)$, la quale garantisce che la crescita della vegetazione ΔDAM sia maggiore quando la temperatura atmosferica (T_a) è più vicina alla temperatura ottimale (T_{opt}), e che si arresti quando la T_a risulti essere esterna all'intervallo $[T_{min}, T_{Max}]$ in cui la vegetazione può svilupparsi. La funzione $F_T(T_a)$ è espressa con l'equazione (5.8):

$$\begin{cases} F_T(T_a) = 1 - \left(\frac{T_{opt} - T_a}{T_{opt} - T_{min}} \right)^2; T_{min} < T_a < T_{opt} \\ F_T(T_a) = 1 - \left(\frac{T_{opt} - T_a}{T_{Max} - T_{opt}} \right)^2; T_{opt} < T_a < T_{Max} \\ F_T(T_a) = 0; T_a < T_{min}; T_a > T_{Max} \end{cases} \quad (5.22)$$

L'equazione finale della produzione di biomassa è quindi:

$$\Delta DAM = Rg(t) \cdot \varepsilon_c \cdot \varepsilon_l \cdot ELUE \cdot F_T(T_a) = APAR \cdot ELUE \cdot F_T(T_a) \quad (5.23)$$

La frazione di luce assorbita dalla vegetazione ε_l è una funzione del GLAI tramite l'equazione (5.10) basata sulla legge di Beer:

$$\varepsilon_l = 1 - e^{(-k \cdot GLAI)} \quad (5.24)$$

dove k è un coefficiente di estinzione della luce all'interno della canopy.

L'andamento temporale del GLAI è suddiviso in 2 fasi. Nella prima fase il GLAI cresce ($\Delta GLAI+$). Questa fase è compresa tra l'emergenza e l'inizio della senescenza ed espressa per mezzo della seguente equazione:

$$\Delta GLAI+ = \Delta DAM \cdot P_L(\sum T_a) \cdot SLA \quad (5.25)$$

con P_L funzione di allocazione della biomassa alle foglie, una funzione empirica della temperatura dell'aria espressa dall'equazione (5.12):

$$P_L(\sum T_a) = 1 - P_{La} \cdot e^{P_{Lb} \sum T_a} \quad (5.26)$$

con P_{La} e P_{Lb} parametri empirici, prossimi allo zero, di ordine di grandezza rispettivamente di 10^{-2} e 10^{-3} , adattati dall'equazione espressa da Maas (1993). Più il parametro P_{La} si avvicina allo zero più P_L decresce esponenzialmente con il tempo termico, da un valore di 1 all'emergenza della pianta fino a un valore di 0 alla fine della fase di produzione delle foglie.

La seconda fase dell'andamento del GLAI è compresa tra l'inizio e la fine della senescenza. In questa fase il GLAI decresce ($\Delta GLAI-$) come espresso dall'equazione (5.13):

$$\Delta GLAI- = GLAI \cdot \frac{(\sum T_a - S_{TT})}{P_{fen_SenB}} \quad (5.27)$$

dove P_{fen_SenB} è il tasso dei parametri di senescenza.

La terza variabile temporale simulata da SAFY è la produzione di granella (GY). Questa è proporzionale alla biomassa per mezzo di un parametro di allocazione della biomassa alla granella P_y :

$$\Delta GY = DAM \cdot P_y \quad (5.28)$$

La versione modificata di SAFY utilizzata in questo studio introduce l'influenza del contenuto d'acqua nel suolo nella determinazione delle tre variabili temporale sopra espresse. In particolare nella produzione di biomassa (ΔDAM) è stato aggiunto un fattore di stress idrico K_s , un fattore adimensionale di riduzione della traspirazione dipendente dalla disponibilità di acqua nel suolo. Tale fattore è compreso tra 0 e 1, dove per $K_s=1$ non si è in presenza di stress idrico. Esso è calcolato come bilancio tra

la quantità di acqua totale disponibile nel suolo nella zona delle radici (TAW) e la quantità di acqua facilmente disponibile (RAW). Se la quantità di acqua necessaria alla vegetazione (determinata dall'evapotraspirazione effettiva) è inferiore o pari alla quantità di acqua facilmente disponibile (RAW), non si è in presenza di stress idrico, pertanto $K_s=1$. Nel caso in cui la richiesta di acqua (indicata in mm come root zone depletion, D_r) sia superiore al valore di RAW allora si è in presenza di stress idrico ed il valore di K_s è espresso dall'equazione 5.15:

$$K_s = \frac{TAW - D_r}{TAW - RAW} = \frac{TAW - D_r}{(1 - p) \cdot TAW} \quad (5.29)$$

dove D_r è la quantità d'acqua estratta dalle radici (Root Depletion) espressa in mm, e p è la frazione di TAW che una coltura può estrarre dalla zona delle radici senza soffrire stress idrico. In presenza di stress idrico K_s segue l'andamento lineare mostrato in Fig. 5.3. Il bilancio idrico introdotto in SAFY segue le linee guida presentate nel FAO irrigation and drainage paper n. 56 (Allen et al., 1998).

L'equazione della produzione di biomassa (ΔDAM) usata nella versione modificata di SAFY è:

$$\Delta DAM = APAR \cdot ELUE \cdot F_T(T_a) \cdot K_s \quad (5.30)$$

con K_s calcolato come indicato in precedenza.

6. Analisi di sensibilità

6.1. Introduzione

I modelli sono un importante strumento per comprendere le interazioni tra l'ambiente, la sua gestione da parte dell'uomo e la crescita delle colture. Essendo una rappresentazione matematica di processi naturali, le equazioni e le parametrizzazioni che li compongono, inevitabilmente comportano semplificazioni della realtà. Ciò implica incertezza ed inaccuratezza negli output (Saltelli et al., 2000). L'analisi di sensibilità (SA, dall'inglese Sensitivity Analysis) ha lo scopo di quantificare l'influenza di ogni fattore (parametro o variabile di input) sulla variabilità dell'output del modello. Tale analisi è uno strumento chiave nella comprensione del comportamento del modello al cambiamento dei fattori che lo descrivono (Cariboni et al., 2007). In particolare, è utile per tre motivazioni principali (Confalonieri, 2010, Refsgaard et al., 2007):

1. individuazione dei parametri che hanno un basso impatto sull'output, in questo modo tali parametri, essendo scarsamente influenti, possono essere fissati come costanti per semplificare il modello;
2. individuazione dei parametri con alto impatto sull'output, per una semplificazione della calibrazione o per fornire linee guida nella gestione di strategie o politiche agricole o, nel caso specifico di questa tesi, limitare il numero di parametri da ricalibrare durante l'applicazione dei metodi di assimilazione;
3. individuazione di uno squilibrio del modello, mettendo in evidenza i casi in cui alcuni parametri hanno una rilevanza significativamente più alta degli altri.

Come descritto nella sezione 2.2, è possibile distinguere due differenti tecniche di SA: Local Sensitivity Analysis e Global Sensitivity Analysis (Saltelli et al., 2008). I metodi di analisi di sensibilità globale (Global Sensitivity Analysis, GSA) valutano gli effetti che la variazione simultanea di numerosi fattori di input ha sull'output del modello in ampi intervalli di variazione (Saltelli et al., 2008). Un'accurata panoramica sui metodi di SA è stata fatta da Cariboni et al. (2007), in cui è mostrata l'inadeguatezza dei metodi di sensibilità locali per modelli di crescita delle colture, a causa della complessità di questi ultimi e della necessità di conoscere le variazioni di output dovute alle interazioni tra parametri. Pertanto per analizzare un modello di crescita delle colture è indispensabile affidarsi ad un metodo di analisi di sensibilità globale.

I metodi GSA possono essere suddivisi in tre gruppi: i) metodi basati sulla varianza, come SOBOL (Sobol, 1993), Fourier Amplitude Sensitivity Test (FAST) (Cuckier et al., 1978), Extended FAST (EFAST) (Saltelli et al. 1999); ii) metodi screening, come ad esempio il metodo di Morris (Morris, 1991); iii) metodi basati sulla regression, come Latin Hyper Cube Sampling o il Random and Quasi-Random LpTau (Confalonieri et al., 2010a).

In Confalonieri et al. (2010a) è stato eseguito un paragone tra numerose tecniche di SA, applicandole al modello di crescita del riso WARM, classificando i parametri in ordine crescente di rilevanza e mostrando come queste classifiche cambino al variare degli scenari in cui il modello è applicato. Sono stati inoltre valutate l'accuratezza e l'efficienza dei vari metodi analizzati, concludendo che non è sempre necessario applicare il metodo più accurato per individuare i parametri più rilevanti. Infatti viene mostrato che il metodo risultato più rapido (avente quindi un costo computazionale più basso) tra quelli analizzati, il metodo di Morris, produce risultati comparabili con quelli ottenuti da metodi con costi computazionali più elevati.

Se il modello è descritto da un ridotto numero di parametri, o è possibile escludere a priori alcuni parametri dall'analisi, è conveniente usare un metodo basato sulla varianza. Tra i metodi di questo tipo più usati ci sono senza dubbio Sobol ed EFAST. L'efficienza di questi metodi è simile, in termini di costi computazionali, la differenza tra i due è che EFAST evidenzia meglio che SOBOL l'influenza tra le interazioni dei parametri sulla varianza del modello (Vazquez-Cruz et al. 2014). In letteratura entrambi i metodi sono molto usati per analizzare la sensibilità di un limitato numero di parametri. Chen et al., (2014) ad esempio, ha preferito utilizzare gli indici di Sobol mentre in altri studi (e.g. Ma et. al, 2011; Li et al., 2014) è stato preferito il metodo EFAST.

Confalonieri (2010) ha proposto una combinazione dei metodi Morris e Sobol, sfruttando i vantaggi computazionali dei Metodi di Morris per individuare i parametri non influenti ed escluderli dall'applicazione del metodo Sobol, usato per individuare con maggior precisione i parametri più influenti. Confalonieri (2010) ha utilizzato tale procedura su due diversi modelli, CropSyst e WOFOST, inoltre ha applicato il test di Grubbs per l'individuazione dei valori anomali (Grubbs, 1969) per verificare se un parametro era significativamente più influente degli altri, assumendo l'omogeneità degli indici di sensibilità come indicatore di bilanciamento del modello.

Vanyutrecht et al. (2014) ha usato la stessa strategia, combinando Morris con EFAST. Questi autori hanno mostrato che le condizioni ambientali condizionano molto il grado di influenza dei parametri di input sugli output, stabilendo una classifica d'influenza per ogni scenario climatico, oltre a mostrare la validità della combinazione di questi due metodi GSA applicati al modello Aquacrop in svariati scenari. Vanyutrecht et al. (2014) hanno pertanto fornito un'importante analisi del modello Aquacrop per scenari climatici molto diversi, mostrando una classifica di importanza dei parametri che facilita l'uso di questo modello, pur sottolineando l'importanza di ripetere tale studio al cambiare delle condizioni al contorno per ottenere una più precisa classificazione dei parametri.

Sulla base degli studi incontrati in letteratura, è stato scelto di analizzare i modelli Aquacrop e SAFY combinando due diversi metodi di SA, Morris ed EFAST, come proposto da Vanuytrecht et al. (2014). Gli obiettivi della SA svolta sono principalmente due:

1. individuare i parametri dei due modelli maggiormente influenti sulla produzione di granella di una coltura di frumento tenero, al fine di effettuare la ricalibrazione dei soli parametri principali in fase di assimilazione;
2. effettuare un confronto dei modelli quantificandone la complessità e la plasticità, in modo da fornire maggiori informazioni da utilizzare in fase di scelta del modello.

Di seguito sono riportati i metodi ed i risultati inerenti alla SA effettuata per questo studio sui modelli SAFY e Aquacrop in due differenti località della Cina, Xiaotanshan e Yangling, utilizzando i database climatici mostrati nel paragrafo 3.1. Pur essendo una SA specifica per i siti presi in esame, la cui risultante classificazione di importanza dei parametri è da limitare a questi casi particolari, è possibile fare un confronto qualitativo tra i due modelli per mezzo degli indici di complessità e plasticità.

6.2. Metodi di Global Sensitivity Analysis (GSA) utilizzati

In questo studio è stata applicata la strategia sopra descritta (Confalonieri, 2010; Vanuytrecht et al., 2014) che prevede l'impiego combinato di due metodi di GSA, un metodo di screening, che, per mezzo della sua rapidità di calcolo, permette di individuare facilmente i fattori del modello non influenti, ed un metodo basato sulla scomposizione della varianza dell'output del modello. In questa seconda analisi vengono esclusi i parametri precedentemente riconosciuti come non influenti, in modo

da ridurre l'elevato costo computazionale di questo tipo di metodo ed ottenere una più accurata classificazione dei parametri maggiormente influenti. I metodi di GSA utilizzati in questo studio sono il metodo di Morris (Morris, 1991), di screening, ed il metodo basato sulla varianza Extended Amplitude Sensitivity Test (EFAST; Saltelli et al., 1999).

I modelli analizzati sono Aquacrop e SAFY, precedentemente descritti nel capitolo 5. Essi sono stati confrontati per mezzo di due indici, uno di complessità (Confalonieri et al., 2009) e l'altro di plasticità (Confalonieri et al., 2012), calcolati per mezzo della SA, i quali quantificano queste due caratteristiche dei modelli permettendone un confronto qualitativo.

6.2.1. Metodo di Morris

Il metodo di Morris (Morris, 1991) è un metodo di GSA appartenente al gruppo degli “screening methods”. Esso si basa sul principio di quantificare gli effetti elementari (d_i) che la variazione degli input, o fattori (variabili e parametri), produce sull'output del modello. Più specificatamente, l'obiettivo del metodo è determinare se gli effetti della variazione dei fattori sono: trascurabili, lineari e additivi, o non lineari e conseguentemente coinvolti nelle interazioni con gli altri fattori (Campolongo et al., 2007).

Definiamo come disegno sperimentale del metodo, l'insieme delle combinazioni di fattori che sono utilizzate per la SA, ciascuna individuata in maniera randomizzata one-at-time e definita come punto del disegno sperimentale (Monod et al., 2006). Il disegno sperimentale di Morris ha le seguenti caratteristiche (Pujol et al., 2009): i punti sono campionati da una regione (Ω) strutturata come una griglia regolare di p dimensioni (dove p è il numero di livelli della griglia). Il disegno sperimentale è strutturato in gruppi di punti, chiamati traiettorie. Tali traiettorie sono casuali ma seguono uno schema specifico, infatti ogni punto della traiettoria differisce dal precedente solo per un fattore (one-factor-at-time, OAT). La regione di sperimentazione Ω è uno spazio a k dimensioni, rappresentato con una griglia di p livelli, dove k è il numero di fattori da esaminare e p sono i livelli della griglia. In questo studio è stato utilizzato un numero di punti pari al numero di livelli ($k=p$). Il numero di traiettorie (r) è stato impostato a 20, seguendo le indicazioni date in letteratura (Campolongo et al. 2007), per le quali è sufficiente impostare $r=10$ per ottenere un numero di simulazioni casualmente distribuito in tutto l'iperspazio Ω .

Ogni traiettoria è composta da k punti, ogni punto è un fattore di input del modello (insieme di parametri e variabili di input).

La strategia proposta da Campolongo et al. (2007), e usata in questo lavoro, propone un miglioramento della strategia classica (Morris, 1991) di campionamento. Tale miglioramento punta ad ottenere una scansione più appropriata del dominio degli input senza incrementare il numero di esecuzioni del modello. La strategia è quella di generare un alto numero di traiettorie, tra le 500 e le 1000, in modo che l'intero iperspazio sia coperto dalle traiettorie. Ciò assicura che il valore di ogni singolo fattore assuma valori distribuiti in tutto il range di variazione e che le combinazioni di tali fattori siano distribuite in tutto l'iperspazio. Tra queste Q traiettorie generate solo r traiettorie verranno effettivamente scelte per l'esecuzione del modello. Il numero di traiettorie generate in questo lavoro è Q=500. Tra queste sono state scelte le r traiettorie (con r=20) con lo spread maggiore. Lo spread è definito come la distanza d_{ml} tra una coppia di traiettorie m e l, secondo l'equazione (6.1):

$$d_{ml} = \begin{cases} 0; m = l \\ \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \sqrt{\sum_{z=1}^{k+1} [X_i^m(z) - X_j^l(z)]^2}; m \neq l \end{cases} \quad (6.1)$$

dove: k è il numero di fattori di input, m e l sono 2 traiettorie, $X_i^m(z)$ indica la z-esima coordinata dell'i-esimo punto della traiettoria m e analogamente $X_j^l(z)$ indica la z-esima coordinate del j-esimo punto della traiettoria l. Sono scelte infine le r traiettorie che meglio massimizzano la distanza d_{ml} .

Come visto precedentemente, ogni fattore di input del modello X_i (con $i=1, \dots, k$), varia attraverso i p livelli selezionati nell'iperspazio. I valori X_i sono scelti in un intervallo compreso tra 0 e 1, con passo $1/(p+1)$ come mostrato in (6.2):

$$I = \left\{ 0, \frac{1}{p+1}, \frac{2}{p+1}, \dots, \frac{p}{p+1}, 1 \right\} \quad (6.2)$$

I fattori hanno bisogno di essere tutti definiti nell'intervallo (0,1), pertanto è necessario normalizzare il range di variazione di ciascun fattore per portarli nell'intervallo tra zero ed uno per mezzo di una semplice proporzione. Una volta create le traiettorie i parametri sono riconvertiti nel loro intervallo di variazione iniziale.

L'effetto elementare (d_i) dell'i-esimo fattore (X_i) dovuto alla perturbazione Δ è la differenza relativa tra l'output del modello ottenuto con e senza perturbazione (6.3):

$$d_i(X_1, X_2, \dots, X_n) = \frac{y(X_1, \dots, X_{i-1}, X_i + \Delta, X_{i+1}, \dots, X_k) - y(X_1, X_2, \dots, X_k)}{\Delta} \quad (6.3)$$

dove Δ è la perturbazione, e assume il valore di ogni elemento dell'intervallo (6.2), p è il numero di livelli, X il singolo parametro, i l'indice del parametro soggetto alla perturbazione, y è l'output del modello. Morris (1991) suggerisce come indice di misura della sensibilità, la media (μ) e la deviazione standard (σ) della distribuzione degli effetti elementari ($d_i(X_1, X_2, \dots, X_n)$).

La media (μ) è misurata come:

$$\mu_i = \frac{1}{r} \sum_{k=1}^r (d_i^k), i = 1 : k \quad (6.4)$$

e rappresenta il grado di importanza del fattore i -esimo. Campolongo et al. (2007) suggerisce di considerare la media dei valori assoluti (6.5), per eludere il problema che nella media (6.4), gli effetti elementari con valore opposto si eliminano l'un l'altro:

$$\mu_i^* = \frac{1}{r} \sum_{k=1}^r |d_i^k|, i = 1 : k \quad (6.5)$$

Studi empirici (Campolongo et al., 2007) dimostrano che μ^* rappresenta un indice di sensibilità totale migliore rispetto alla semplice media.

La deviazione standard (σ) è:

$$\sigma_i = \sqrt{\frac{1}{r} \sum_{k=1}^r (d_i^k - \mu_i)^2}, i = 1 : k \quad (6.6)$$

e rappresenta la misura della non linearità rispetto all' i -esimo fattore o l'importanza delle interazioni coinvolte con il fattore i -esimo, o entrambe le cose. Il metodo di Morris non consente infatti di distinguere tra la non linearità e le interazioni dei fattori, rimarcando comunque, che l'analisi dei dati può fornire informazioni circa questo fenomeno (Pujol, 2009). Usando μ^* e σ , è possibile distinguere tre gruppi di fattori: i) fattori trascurabili, che hanno μ^* inferiore a un valore soglia, ii) fattori con effetti lineari privi di interazioni, con μ^* superiori al valore soglia e σ inferiore ad un valore soglia, iii) fattori con effetti non lineari e/o interazioni, i quali hanno μ^* e σ entrambi superiori ai valori soglia. Tali valori soglia sono decisi arbitrariamente dall'utente, a seconda delle esigenze e degli obiettivi che si vogliono perseguire.

In questo studio, il metodo di Morris è stato applicato per individuare i fattori trascurabili con lo scopo di semplificare l'applicazione del metodo di EFAST. Per questa ragione è stato sufficiente considerare esclusivamente l'indice di sensibilità totale μ^* , in quanto la sensibilità dovuta all'interazione dei parametri è stata valutata in seguito per mezzo del metodo EFAST.

6.2.2. Extended Fourier Amplitude Sensitivity Test (EFAST)

Fourier Amplitude Sensitivity Test (FAST) è un metodo di SA appartenente al gruppo dei metodi cosiddetti variance-based. Esso si basa sulla scomposizione della varianza di Sobol (Sobol, 1993) e propone di calcolare gli indici di sensibilità usando la varianza totale dell'output Y e le varianze condizionali conoscendo i parametri. Questi metodi richiedono il calcolo dei momenti del primo e del secondo ordine:

$$E(Y^r) = \int_{K^n} f^r(x_1, \dots, x_n) \cdot p(x_1, \dots, x_n) dx \quad (6.7)$$

dove r è l'ordine del momento (1 o 2), f(...) è l'insieme degli algoritmi del modello considerato, K^n è lo spazio a n dimensioni dove i parametri $x=(x_1, \dots, x_n)$ sono definiti e p è la funzione di densità di probabilità (pdf) della loro legge di distribuzione. L'idea del metodo FAST è di trasformare l'integrale multidimensionale, che è troppo complesso perché possa essere calcolato, in un integrale a una dimensione, attraverso il cambio di variabile $x_i=G_i(\sin(\omega_i s))$, dove s è uno scalare, $\omega_i s$ è un insieme di frequenze intere associate a differenti parametri, G_i è una funzione chiamata funzione di trasferimento. Esistono numerose funzioni di trasferimento sviluppate negli anni (Cukier et al., 1978; Koda et al., 1979; Saltelli et al., 1999). Quella usata in questo lavoro è stata proposta da Lu et al. (2001):

$$G_i(u) = F_i^{-1}\left(\frac{1}{\pi} \arcsin(u) + C\right) \quad (6.8)$$

dove F_i^{-1} è la funzione inversa della distribuzione dei fattori cumulativi x_1 . La curva creata dalle variazioni di s, la forma da cui dipende la funzione di trasferimento G_i , esplora lo spazio dei parametri K^n in tutte le direzioni. Questa curva è chiamata "search function". Utilizzando un valore intero per ω_i la funzione f deve essere una funzione periodica di s con periodo 2π . In questo modo è possibile sviluppare la funzione f in serie di Fourier:

$$f(s) = \sum_{j=-\infty}^{+\infty} (A_j \cos(js) + B_j \sin(js)) \quad (6.9)$$

A_j e B_j sono i coefficienti di Fourier:

$$\begin{aligned} A_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(js) ds \\ B_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(js) ds \end{aligned} \quad (6.10)$$

nel dominio (dei soli numeri interi):

$$j \in \{-\infty, \dots, -1, 0, 1, \dots, +\infty\} \quad (6.1)$$

Usando il teorema di Parseval (cfr. Saltelli et al., 1999), lo spettro (Λ) di sviluppo nelle serie di Fourier f è:

$$\sum \Lambda_j = \frac{1}{2\pi} \int_{-x}^x f^2(s) ds \quad (6.2)$$

ne segue che:

$$2 \sum_{j=1}^{+\infty} \Lambda_j = \sum \Lambda_j - \Lambda_0 = E(Y^2) - E^2(Y) = D \quad (6.13)$$

La varianza totale D di $Y=f(s)$ può essere espressa come somma dello spettro delle serie di Fourier di f . La scomposizione delle serie di Fourier definisce f come somma di funzioni seno e coseno, con periodo multipli di 2π , moltiplicato per i coefficienti A_j e B_j . Questa scomposizione è unica perché A_j e B_j non dipendono da s . Lo spettro Λ_j è il quadrato della norma delle funzioni moltiplicative $\sin(js)$ e $\cos(js)$, e caratterizza l'importanza delle oscillazioni di f alla frequenza $j/2\pi$. Il principio del metodo FAST (Cukier et al., 1978) è di far oscillare gli input x_i a differenti frequenze ed osservare l'importanza dei componenti della funzione f a queste frequenze. L'importanza dei componenti di f , che sono associate con le frequenze ω , è quantificata dalla somma degli elementi dello spettro corrispondenti a queste frequenze (ed alle sue armoniche). La porzione di varianza di Y dipendente dal fattore x_i è:

$$D_i = 2 \sum_{p=1}^{\infty} \Lambda_{p\omega_i} \quad (6.4)$$

L'indice di sensibilità principale diventa:

$$S_i = \frac{D_i}{D} = \frac{2 \sum_{p=1}^{\infty} \Lambda_{p\omega_i}}{D} \quad (6.5)$$

Il metodo FAST stabilisce una corrispondenza unica tra i coefficienti di Fourier (calcolati per le frequenze ω_i associate con il fattore x_i) e la sensibilità del primo ordine dell'output relativa al fattore x_i . Usando il metodo FAST è possibile calcolare solo l'effetto principale della variazione degli input sugli output, perché esso trascura gli effetti del second'ordine, da cui è possibile quantificare la sensibilità delle interazioni tra le variazioni dei parametri. Calcolare nel dettaglio i termini di interazione è complesso perché deve considerare tutte le variazioni lineari delle frequenze ω_i . Saltelli et al. (1999) mostra un metodo per quantificare abbastanza rapidamente l'influenza di ogni interazione, calcolando l'indice di sensibilità totale (ST_i), che è la somma dell'effetto principale di x_i più l'interazione dei termini a tutti gli ordini includendo x_i . Tale metodo consiste nell'assegnare un alto valore di frequenza al

fattore x_i e dei bassi valori alle frequenze degli altri fattori. In Saltelli et al. (1999) si afferma che l'interazione tra i fattori x_i e il fattore $x_{\sim i}$ ($x_{\sim i}$ rappresenta il fattore generico che si sta confrontando, eccetto x_i) è concentrata attorno al picco di frequenza e pertanto si propone di calcolare la varianza parziale $D_{\sim i}$, la quale include gli effetti di tutti i parametri eccetto x_i , come mostrato dall'equazione (6.16):

$$D_{\sim i} = \sum_{j=1}^{\frac{\omega_i}{2}} \Lambda_j \quad (6.6)$$

L'indice di sensibilità totale può dunque essere calcolato usando l'equazione (6.17):

$$ST_i = 1 - \frac{D_{\sim i}}{D} = 1 - \frac{\sum_{j=1}^{\frac{\omega_i}{2}} \Lambda_j}{2 \sum_{j=1}^{+\infty} \Lambda_j} \quad (6.7)$$

Riassumendo quanto detto finora e semplificando le equazioni mostrate, l'indice di sensibilità del singolo parametro è espresso dalla (6.18):

$$S_i = \frac{D_i}{D} \quad (6.8)$$

mentre l'indice di sensibilità totale è espresso dalla (6.19):

$$ST_i = 1 - \frac{D_{\sim i}}{D} \quad (6.9)$$

con: $D_{\sim i}$ somma della varianza di tutti i fattori, eccetto il valore i -esimo (ovvero quello in esame).

L'innovazione del metodo EFAST rispetto al classico FAST è nel modo di esplorare lo spazio di n dimensioni in cui vengono scelti i parametri. Questo metodo introduce l'uso di una curva di ricerca definita dall'equazione (6.20):

$$x_i = \frac{1}{2} + \frac{\arcsin[\sin(\omega_i s)]}{\pi} \quad (6.20)$$

Saltelli et al. (1999) suggeriscono come valore ottimo di ω_i un valore compreso nell'intervallo $[16 N_r, (16 N_r + 48 N_r)]$, dove N_r è il numero di curve di ricerca da usare per ricoprire l'iperspazio.

6.2.3. Intervalli di variazione dei parametri

Gli intervalli in cui i parametri vengono fatti variare durante la SA sono scelti ampi affinché possano comprendere tutte le variabilità degli scenari sperimentali osservati. Per Aquacrop, gli intervalli dei parametri inerenti la coltura (Tab. 6.1) sono stati definiti incrementando e diminuendo del 25% i valori utilizzati da Jin et al. (2014) per uno studio di calibrazione del modello condotto sul frumento tenero nel sito sperimentale di Xiaotanshan.

Tab. 6.8: Parametri del modello Aquacrop: nome identificativo del parametro, intervallo in cui il parametro è fatto variare durante la SA e descrizione del parametro.

Name	Value (min- max)	Description
Phenological development		
dos	260 - 280	day of sowing (Julian Day)
eme	30 - 90	Growing Degree Days (GDD) from sowing to emergence
ccs	2.5 - 7.5	Soil surface covered by an individual seedling at 90 % emergence (cm ²)
den	90000- 300000	Number of plants per hectare
sen	1360 - 3000	GDD from sowing to start senescence
flo	249.5 - 748.5	GDD from sowing to flowering
flolen	80 - 120	Length of the flowering stage (GDD)
cdlf	0.5 - 1.5	Crop determinancy linked with flowering
hilen_cd	16 - 48	Building up of Harvest Index starting at flowering (days)
exc	25 - 75	Excess of potential fruits (%)
mat	1400 - 3000	Length of the crop cycle (GDD)
Canopy development		
cgc	0.016 - 0.050	Canopy growth coefficient: increase in canopy cover (fraction soil cover per day)
ccx	0.80 - 0.99	Maximum canopy cover in fraction soil cover
cdc	0.05 - 0.15	Canopy decline coefficient: decrease in canopy cover (in fraction per day)
cgc4ggd	0.01 - 0.02	CGC for GGD: increase in canopy cover (in fraction soil cover per GDD)

cdc4ggd	0.006 - 0.01	CDC for GGD: decrease in canopy cover (in fraction per GDD)
Root development		
root	256 - 768	GDD from sowing to maximum rooting depth
rtx	0.6 - 1.8	Maximum effective rooting depth (m)
rtexup	0.01 - 0.03	Maximum root water extraction ($\text{m}^3 \text{ water}/\text{m}^3 \text{ soil day}$) in top quarter of root zone
rtxlw	0.0025 - 0.0075	Maximum root water extraction ($\text{m}^3 \text{ water}/\text{m}^3 \text{ soil.day}$) in bottom quarter of root zone
rtmin	0.15 - 0.45	Minimum effective rooting depth (m)
Transpiration		
kc	0.07 - 0.2	Crop coefficient when canopy is complete but prior to senescence
kedcl	0.075 - 0.225	Decline of crop coefficient (%/day) as a result of senescence, nitrogen deficiency, etc.
evarde	30 - 90	Effect of canopy cover in reducing soil evaporation in late season stage
Production		
wp	9-27	Water productivity normalized for ET_0 and CO_2 (g/m^2)
wp_yfp	50 - 150	Water Productivity normalized for ET_0 and CO_2 during yield formation (as % of wp)
cpcO2	25 - 75	Crop performance under elevated atmospheric CO_2 concentration (%)
hi	25 - 75	Reference Harvest Index (HI_0) (%)
hipsveg	1-20	Coefficient describing positive impact on HI of restricted vegetative growth during yield formation
hingsto	4-12	Coefficient describing negative impact on HI of stomatal closure during yield formation
hinc	7.5 - 22.5	Allowable maximum increase (%) of specified HI
Water stress		
pexup	0.2 - 0.4	Soil water depletion factor for canopy expansion: upper threshold, fraction of total available water (TAW)

pexlw	0.4 - 0.7	Soil water depletion factor for canopy expansion: lower threshold (fraction TAW)
psto	0.20 - 0.75	Soil water depletion fraction for stomatal control: upper threshold (fraction TAW)
psen	0.8- 1.0	Soil water depletion factor for canopy senescence: upper threshold (fraction TAW)
ppol	0.20 - 0.99	Soil water depletion factor for pollination (p - pol): upper threshold (fraction TAW)
ecsss	1-3	Electrical Conductivity of soil saturation extract at which crop starts to be affected by soil salinity (dS/m)
ecss	1-10	Electrical Conductivity of soil saturation extract at which crop can no longer grow (dS/m)
o hipsfl	5-15	Possible increase (%) of HI due to water stress before flowering
anaer	1-8	Anaerobic point below saturation limiting aeration (vol %)
Temperature stress		
To_crop	-10	Base temperature below which crop development stops (°C)
Tmax_cro p	25 - 35	Upper temperature above which crop development stops (°C)
polmn	5-12	Minimum air temperature below which pollination starts to fail (°C)
polmx	35 - 45	Maximum air temperature above which pollination starts to fail (°C)
stbio	5-16	Minimum GDD required for full biomass production
Soil		
cn	33 - 100	CN (Curve Number)
rew	5-13	Readily evaporable water from top layer (mm)
th	0.825 - 2.457	Thickness
fc	28 - 42	Soil Water Content at Field Capacity
pwp	7-15	Soil Water content at Wilting Point
Ksat	120 - 360	Costant of Saturation

fk	2-6	Evaporation decline factor for stage II
sat	0.15 - 0.58	Saturation
Kex	1 - 1.5	Soil evaporation coefficient for fully wet and non-shaded soil surface
SWC	0.5 - 1.5	Soil Water Content specified at particular depths
NSD	1-3	Number of soil depths considered
SD	0.3 - 1	Soil Depth
WC	20 - 50	Water Content
Ssf	20 - 90	Soil fertility/salinity stress coefficient (%)
Ssf	20 - 90	Soil fertility/salinity stress coefficient (%)

Per SAFY, i parametri inerenti alle colture sono stati variati del 25% intorno ai valori presentati da Duchemin et al (2007) e presentati in Tab. 6.2.

Tab. 6.9: Parametri del modello SAFY: nome identificativo del parametro, intervallo in cui il parametro è fatto variare durante la SA e descrizione del parametro.

Name	Value (min- max)	Description
Phenological development		
Pfen_MrgD	275 - 366	Day of the year of emergence
Pgro_Ms0	2 - 7	Emergence Dry Mass Value (g m ⁻²)
Pfen_PrtA	0.01 - 0.5	Make vary the origin slope of the partitioning function
Pfen_PrtB	0.0001 - 0.02	Make vary the day of max LAI (partition=0)
Pfen_SenA	900 - 2000	Temperature threshold to start senescence (°C)
Pfen_SenB	1000 - 20000	Rate of senescence (°C)
Pfen_flw	400 - 1000	Flowering GDD
Delta_flw	0.55 - 0.85	Proportional Factor flw/Sen
SMT_sen	1600 - 2500	GDD to complete senescence
Pfen_stel	100 - 400	Jointing GDD
Delta_stel	0.15 - 0.3	Proportional Factor stel/Sen
h_max	0.6 - 1	maximum crop height (m)
Canopy development		
Kc_tab_ini	0.15 - 0.3	Crop coefficient at the initial stage

Kc_tab_mid	1 - 1.2	Crop coefficient at mid stage
Kc_tab_end	0.1 - 0.2	Crop coefficient at final stage
Pgro_Sla	0.015 - 0.29	Specific Leaf Area ($\text{m}^2 \text{g}^{-1}$)
		Climatic efficiency: ratio of incoming
Pgro_R2P	0.4 - 0.6	photosynthetically active radiation (PAR) to global radiation
Pgro_Kex	0.4 - 0.8	Light-interception efficiency coefficient: fraction of PAR absorbed by the canopy (APAR)
Production		
Pgro_Lue	1.4 - 2.6	Effective light-use efficiency: ratio of energy produced as DAM from APAR (gMJ^{-1})
Pgro_P2G	0.004 - 0.007	Partition coefficient to grain
Root development		
Rd_tab_ini	0.005 - 0.2	Root depth at emergence (m)
Rd_tab_end	1 - 1.8	Root depth at end of growth (m)
Dr	0 - TAW	initial root zone depletion (mm)
Water stress		
p_tab_ini	0.4 - 0.6	Fraction of readily available water at the initial stage
p_tab_mid	0.4 - 0.6	Fraction of readily available water at the mid stage
p_tab_end	0.7 - 0.9	Fraction of readily available water at the final stage
DoyIrr_1	30 - 60	Day of first irrigation
DoyIrr_2	60 - 90	Day of second irrigation
DoyIrr_3	90 - 120	Day of third irrigation
DoyIrr_4	120 - 150	Day of fourth irrigation
Irr_mm_TOT	0 - 400	Total irrigated water
Temperature stress		
Ptfn_Tmin	-5 - 5	Minimum temperature for plant development (°C)
Ptfn_Tmax	35 - 45	Maximum temperature for plant development (°C)

Ptfn_Topt	15 - 25	Optimum temperature for plant development (°C)
Ptfn_TpSn	1.5 - 2.5	Make vary the length of plateau around Ptfn_Topt
Soil		
FC	0.1 - 0.46	Field Capacity
Da	0.9 - 1.6	Soil bulk density (g cm ⁻³)
De	0 - TEW	initial cumulative depth of evaporation (mm)
Ze	0.02 - 0.15	Depth of the soil layer for the evaporation (m)

I parametri inerenti al suolo per entrambi i modelli sono scelti in modo da rappresentare tutte le tessiture di suolo presenti nei 2 siti sperimentali di Xiaotanshan e Yangling.

6.2.4. Indici per il confronto tra i modelli

I modelli Aquacrop e SAFY sono stati confrontati rispetto a due caratteristiche evidenziate per mezzo della SA: la complessità e la plasticità.

La complessità è la parsimonia del modello nel rappresentare il sistema biofisico. Fornisce informazioni circa la quantità dei parametri e la loro rilevanza (Confalonieri et al., 2009). In questo studio si è scelto di quantificare la complessità per mezzo del “Parameter Ratio” (R_p) presentato da Confalonieri et al. (2009). Tale indice è la proporzione tra i fattori rilevanti (S) sul numero totale dei fattori (T):

$$R_p = \frac{S}{T} \quad (6.1)$$

Tutti i fattori, per cui la somma degli indici di EFAST, S_i e ST_i è risultata essere maggiore di 0.15, sono stati considerati come rilevanti.

La plasticità è la tendenza del modello di cambiare il suo comportamento quando è impiegato in condizioni differenti. In accordo con Confalonieri et al. (2012), i cambiamenti del comportamento del modello possono essere quantificati utilizzando un indicatore di plasticità, rappresentato dalla (6.22):

$$L = TDCC \cdot e^{\sigma_{SAM}^{-1}} \quad (6.2)$$

dove TDCC è il “top-down concordance coefficient” presentato da Iman e Conover (1987), σ_{SAM} è la deviazione standard dell’indicatore agrometeorologico SAM (6.23):

$$SAM = \frac{Rain - ET_0}{Rain + ET_0} \quad (6.3)$$

con Rain pioggia totale del periodo interessato dal ciclo della coltura (settembre-giugno) espressa in mm e ET_0 evapotraspirazione totale espressa in mm. L'indice di plasticità proposto da Confalonieri et al. (2012) varia tra 0 e 1.51, dove lo 0 indica la plasticità più alta.

6.3. Risultati della SA

6.3.1. Applicazione del metodo di Morris

I risultati del metodo di Morris (sia per Aquacrop che per SAFY) mostrano che il valore di μ^* , indice di sensibilità totale, rappresentato dalla media del valore assoluto degli effetti elementari (eq. 6.5) sulla produzione di granella di una coltura di frumento tenero, è fortemente influenzato dalle condizioni climatiche. Infatti, i valori di μ^* sono risultati molto diversi non solo tra un sito sperimentale e l'altro, ma sono variati significativamente anche tra i vari anni analizzati per lo stesso sito, come mostrato dalla loro deviazione standard (barra di errore) in Fig. 6.1A e 6.1B.

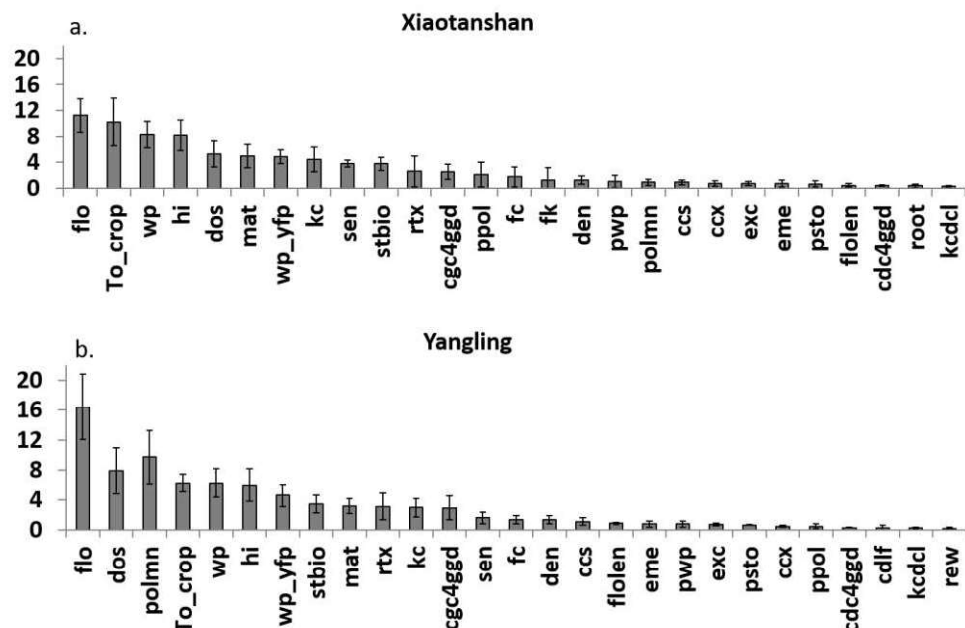


Fig. 6.13A: Valore medio e deviazione standard (barra di errore) dell'indice di influenza di Morris (μ^*) per i 3 diversi cicli di crescita del frumento duro, calcolati applicando la SA ad Aquacrop con i dati climatici di Xiaotanshan (a) e Yangling (b).

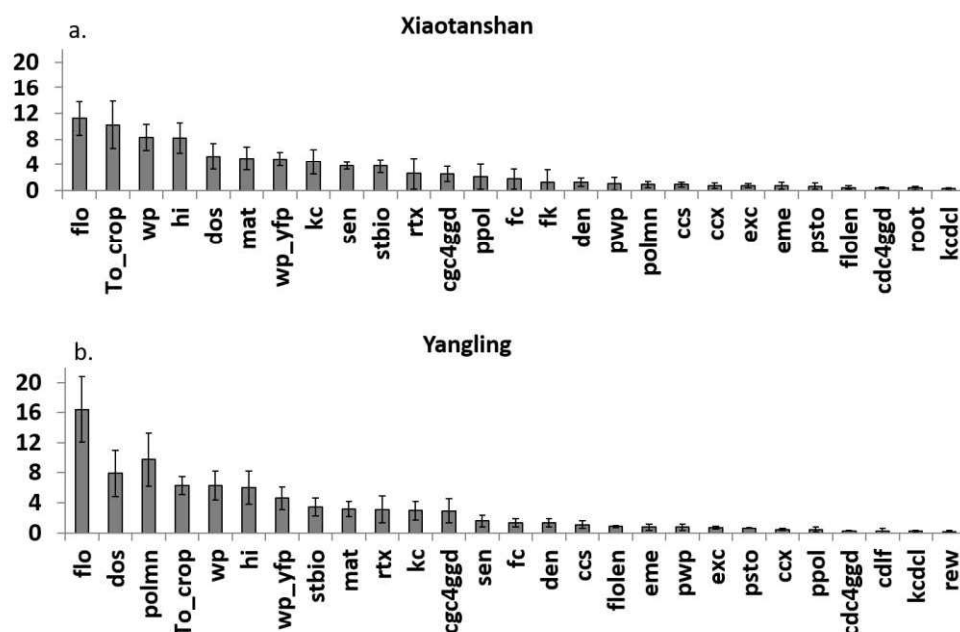


Fig. 6.14B: Valore medio e deviazione standard (barra di errore) dell'indice di influenza di Morris (μ^*) per i 3 diversi cicli di crescita del frumento duro, calcolati applicando la SA a SAFY con i dati climatici di Xiaotanshan (a) e Yangling (b).

In generale sono stati osservati valori maggiori di μ^* applicando la SA ad Aquacrop rispetto a quelli misurati effettuandola su SAFY, mentre confrontando i due siti, μ^* risulta sempre maggiore per Yangling che per Xiaotanshan. Ad esempio il valore massimo di μ^* trovato applicando il metodo Morris ad Aquacrop con i dati di Yangling è stato di $24.9 \text{ t} \cdot \text{ha}^{-1}$, mentre per SAFY, nello stesso sito, il valore maggiore di μ^* è di $17.1 \text{ t} \cdot \text{ha}^{-1}$.

La SA condotta sul modello Aquacrop mostra che i parametri che descrivono la coltura influenzano maggiormente la produzione rispetto ai parametri del suolo. L'analisi su entrambi i siti evidenzia che quasi tutti i parametri inerenti alla fenologia ed allo sviluppo della canopy non sono trascurabili. Tra i parametri inerenti la produzione solo tre risultano non trascurabili e con un μ^* molto alto, infine solo tre parametri inerenti al suolo sono risultati rilevanti, tutti i restanti parametri sono stati considerati trascurabili ed esclusi dalla successiva SA svolta con il metodo EFAST. I parametri di Aquacrop risultati più influenti dalla SA svolta con il metodo di Morris sono gli stessi per i due siti sperimentali, ad eccezione del fattore di declino dell'evapotraspirazione (f_k) e della profondità delle radici dalla semina alla profondità massima raggiunta (root), i quali sono risultati non trascurabili solo per il sito di Xiaotanshan. Per questi due parametri, l'indice di Morris è stato fortemente influenzato dallo scenario, pertanto è opportuno includerli tra i parametri da analizzare con EFAST, in modo da verificare con maggiore accuratezza il loro grado d'influenza.

Il parametro che ha l'indice μ^* più alto è concorde per entrambi i siti ed è il parametro che determina l'inizio della fioritura in GDD (flo). Questo parametro ha determinato un valore di μ^* mediamente intorno alle 13 t·ha⁻¹ per tutti i siti e gli anni analizzati, ad eccezione del set di dati di Yangling 2013-2014, per il quale il valore di μ^* del parametro flo raggiunge un valore di 21.34 t·ha⁻¹. La temperatura basale (To_crop) si è rivelata essere molto influente, più in Xiaotanshan che in Yangling, probabilmente perché nel primo sito le temperature raggiungono valori sotto zero per periodi più lunghi che a Yangling, pertanto le colture sono esposte per periodi più lunghi a stress termico. La temperatura minima sotto la quale l'impollinazione viene inibita (polmn) è un parametro che, pur essendo rientrato tra i parametri non trascurabili per entrambi i siti, è risultato essere molto importante in Yangling, con un valore medio di $\mu^*=9.7$ t·ha⁻¹, e molto meno per Xiaotanshan, con $\mu^*=0.9$ t·ha⁻¹. Questa discrepanza può essere spiegata con la differenza climatica che si ha tra i due siti sperimentali. Una situazione simile è avvenuta anche per il parametro che determina l'inizio della senescenza (sen), ma in questo caso la differenza tra un sito e l'altro è piuttosto ridotta, con $\mu^*=3.8$ t·ha⁻¹ per Xiaotanshan, e $\mu^*=1.6$ t·ha⁻¹ per Yangling.

Le differenze climatiche tra i due siti, pur determinando qualche differenza tra l'influenza dei parametri, non hanno impedito di poter distinguere un gruppo di parametri risultati sempre inferiori ad un valore soglia fissato. Nel presente studio tale valore è stato stabilito, in accordo con quanto fatto da Vanuitrecht et al. (2014), a $\mu^*=0.25$ t·ha⁻¹, sotto il quale tutti fattori sono considerati trascurabili.

La SA ha messo in evidenza che tutti i fattori del suolo sono trascurabili, ad eccezione della capacità di campo (fc), del punto di appassimento (pwp) e del fattore di declino dell'evapotraspirazione (fk). Un altro numeroso gruppo di parametri risultato trascurabile è quello legato allo stress idrico. Infatti, dei nove parametri legati allo stress idrico, solo due sono risultati non trascurabili, la soglia superiore della frazione di esaurimento di acqua contenuta nel suolo per il controllo stomatico (ppol) e la soglia superiore del fattore di esaurimento di acqua nel suolo per l'impollinazione (ppol). Tutti i parametri collegati con le temperature massime di stress termico sono risultati trascurabili, poiché tali temperature sono state raramente raggiunte negli scenari considerati. Tra i parametri inerenti la produzione, quattro sono risultati trascurabili: il coefficiente che descrive l'impatto positivo del Harvest Index (HI) sulla crescita vegetativa ristretta durante la formazione di granella (hipsveg), il coefficiente che descrive l'impatto negativo del HI sulla chiusura stomatica durante la formazione

della granella (hingsto), l'aumento percentuale massimo ammissibile di HI specificato ed il parametro che descrive le prestazioni della coltura sotto elevata concentrazione di CO₂ (cpco2). Tra i parametri inerenti la traspirazione, solo l'effetto sulla CC della riduzione di evapotraspirazione del suolo nella parte finale del ciclo (evada) è trascurabile. Tutti i parametri legati allo sviluppo delle radici sono risultati trascurabili o scarsamente influenti.

La SA per SAFY ha messo in evidenza un insieme di parametri per cui il valore di μ^* è risultato superiore al valore soglia di influenza, fissato a 0.25 t·ha⁻¹. Questo insieme è concorde per entrambi i siti sperimentali, ad eccezione di pochi parametri, risultando però il valore medio di μ^* per Yangling maggiore che per Xiaotanshan. Ad esempio il parametro che ha determinato il valore più alto μ^* è risultato essere il quello che fa variare la pendenza della funzione di allocazione della biomassa alle foglie (Pfen_PrtA), con un valore di $\mu^* = 9.01$ t·ha⁻¹ per il sito di Yangling e di $\mu^* = 2.59$ t·ha⁻¹ per Xiaotanshan. L'insieme dei parametri influenti trovato per Xiaotanshan differisce da quello trovato per Yangling per soli quattro parametri. Il parametro che determina la senescenza in GDD (SMT_sen) è molto influente per i dati di Yangling, con $\mu^* = 3.75$ t·ha⁻¹, e trascurabile per i dati di Xiaotanshan, con $\mu^* = 0.16$ t·ha⁻¹. La densità apparente del suolo (Da), similmente, è risultata influente a Yangling e trascurabile a Xiaotanshan, con $\mu^* = 1.45$ t·ha⁻¹ nel primo caso e $\mu^* = 0.02$ t·ha⁻¹ nel secondo. Infine i parametri Rd_tab_ini e Delta_flw (vedere Tab. 6.2 per i dettagli sui parametri) sono risultati moderatamente influenti solo per lo scenario di Yangling. Lo sviluppo fenologico si è confermato essere molto importante anche per questo modello, sia per numero di parametri che per grandezza dell'indice μ^* . Infatti, tutti i parametri appartenenti al gruppo dello sviluppo fenologico sono risultati molto influenti, eccetto h_max e Pfen_stel (vedi Tab. 6.2), che sono invece trascurabili. Anche i parametri in relazione con lo stress termico e con la produzione sono risultati influenti, eccetto la temperatura massima di sviluppo della pianta (Ptn_Tmax) che come per Aquacrop è risultata trascurabile perché quasi mai raggiunta durante i cicli colturali considerati. Tra i parametri descriventi lo sviluppo della canopy, tre sono risultati influenti: l'area fogliare specifica (Pgro_Sla), l'efficienza climatica (PgroR2P) e il coefficiente di luce intercettata (PgroKex), mentre i tre parametri legati al coefficiente colturale (Kc) sono risultati trascurabili. Tutti i fattori legati allo stress idrico sono trascurabili, ad eccezione dell'irrigazione totale. Come per Aquacrop, anche per SAFY, i parametri inerenti al suolo non sono influenti per gli scenari climatici usati in questo studio.

Infatti per SAFY solo la capacità di campo è risultata influente, mentre i restanti parametri del suolo sono trascurabili.

6.3.2. Applicazione del metodo EFAST

In Fig. 6.2 sono mostrati gli indici di sensibilità di EFAST per Aquacrop per gli scenari climatici dei due siti sperimentali, separati per ogni ciclo colturale esaminato.

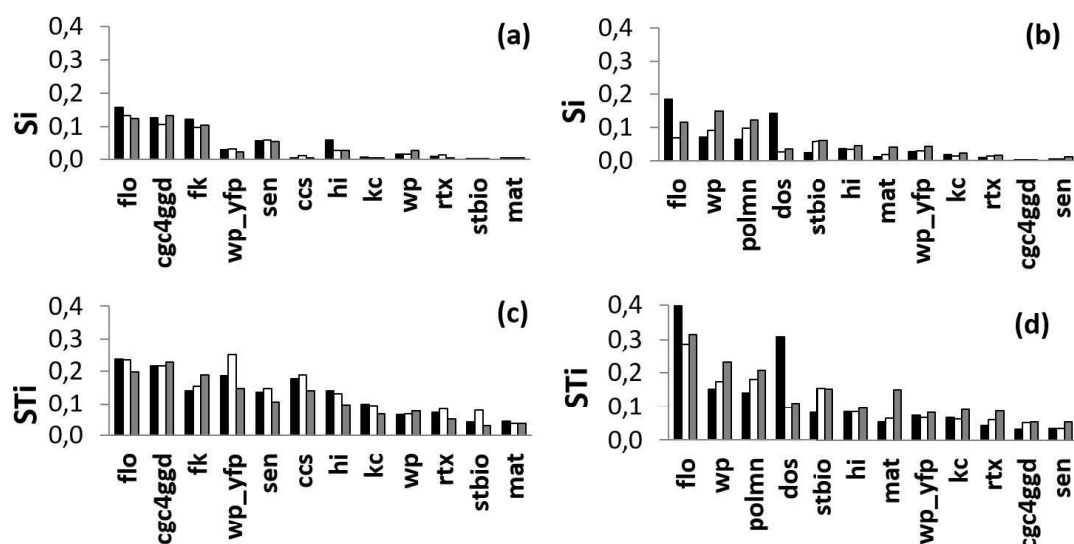


Fig. 6.15: Risultati della SA svolta con EFAST per Aquacrop. (a) Indice principale (S_i) e (c) Indice di sensibilità Totale (ST_i) per Xiaotanshan (barra nera 2008-2009, barra bianca 2009-2010, barra grigia 2010-2011).; (b) indice principale (S_i) e (d) indice di sensibilità totale (ST_i) per Yangling (barra nera 2012-2013, barra bianca 2013-2014, barra grigia 2014-2015).

Gli indici rappresentano gli effetti della variazione del singolo parametro (S_i) e delle interazioni tra i parametri (ST_i), come mostrato nel paragrafo 6.2.2. In Fig. 6.2 sono mostrati esclusivamente i parametri significativi, sono stati dunque omessi tutti i parametri la cui somma dei due indici era prossima allo 0 (minore di 0.01). La SA per Aquacrop ha confermato che i parametri riguardanti la coltura sono più influenti di quelli del suolo. L'insieme di parametri influenti è piuttosto simile per i due siti testati, eccetto per 3 parametri: la data di maturità in GDD (mat) la data di semina (dos) e la temperatura minima per la pollinazione (polmn), risultati rilevanti solo per Yangling, mentre il coefficiente di copertura del suolo al termine del ciclo (ccs), il coefficiente di incremento di crescita di canopy per GDD e il coefficiente di declino di evapotraspirazione (fk), sono risultati influenti solo per Xiaotanshan. Al contrario i parametri flo, sen, hi, wp_yfp, wp, rtx, kc e stbio (Tab. 6.1) sono risultati influenti per entrambi i test. Il parametro più influente per entrambi i siti è la data dell'inizio della fioritura, con $S_i = 0.14$, $ST_i = 0.22$ per Xiaotanshan e $S_i = 0.12$, $ST_i = 0.46$, per Yangling. Per Yangling l'indice S_i è stato influenzato dalle condizioni climatiche stagionali, in particolare per la fioritura (flo) e la data della semina (dos), per cui

l'indice S_i ottenuto per il dataset Yangling 2012-2013 risulta più alto che per tutti gli altri anni analizzati. Non ci sono differenze significative di ST_i tra i vari anni analizzati, ad eccezione dei parametri flo e dos, per cui i valori maggiori vengono raggiunti ancora una volta per il dataset Yangling 2012-2013.

In Fig. 6.3 sono mostrati gli indici di sensibilità per SAFY, per ogni ciclo colturale e per ogni scenario analizzato.

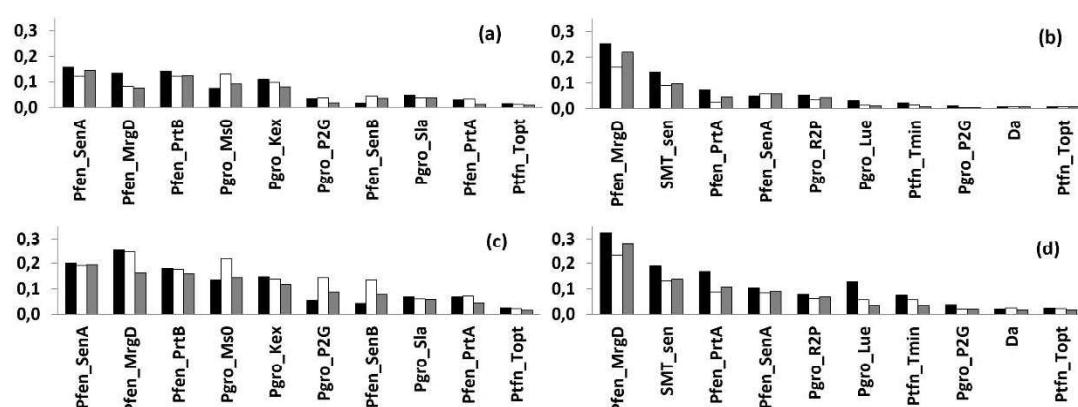


Fig. 6.16: Risultati per la SA effettuata con il metodo EFAST sul modello SAFY. (a) Indice principale (S_i) e (c) Indice di sensibilità Totale (ST_i) per Xiaotanshan (barra nera 2008-2009, barra bianca 2009-2010, barra grigia 2010-2011).; (b) indice principale (S_i) e (d) indice di sensibilità totale (ST_i) per Yangling (barra nera 2012-2013, barra bianca 2013-2014, barra grigia 2014-2015). Le abbreviazioni dei parametri sono spiegate in Tab. 2.

Come per Fig. 6.2, anche in questo caso i parametri con indice di sensibilità prossimo allo 0 sono stati omessi. Per il modello SAFY, il clima gioca un ruolo molto più rilevante nella sensibilità dei parametri, comparato con il modello Aquacrop. Infatti, in questo caso, l'insieme dei parametri influenti è piuttosto differente tra i 2 scenari analizzati. Indici con valori maggiori sono stati ottenuti per il sito di Yangling. In generale, i parametri inerenti lo sviluppo fenologico sono i più influenti, in particolare il giorno dell'emergenza (PfenMrgD) e la temperatura soglia di inizio senescenza (Pfen_SenA). Gli indici S_i e ST_i di PfenMrgD risultano generalmente più alti nel per il dataset di Yangling, con valori di 0.10 and 0.22, rispettivamente, per Xiaotanshan, e 0.21 e 0.28, per Yangling. I risultati per Xiaotanshan 2008-2009 sono molto simili ai risultati ottenuti per il sito di Yangling, probabilmente perché le temperature di questo anno sono state più alte delle medie stagionali tipiche per quella zona, somigliando di più alle temperature medie stagionali dell'area di Yangling. L'altro parametro per cui la somma degli indici è superiore alla soglia di 0.15 per tutti gli scenari è Pfen_SenA, ma a differenza del caso precedente, i valori per Xiaotanshan sono più alti che per Yangling. Si ha infatti, $S_i = 0.14$ e $ST_i = 0.20$ per Xiaotanshan e $S_i = 0.05$ e $ST_i = 0.1$ per Yangling. Per il sito di Xiaotanshan gli altri 2 parametri

influenti sono Pfen_PartB (che fa variare il giorno del LAI massimo) e Pgro_MS0 (valore della biomassa all'emergenza). Questi due parametri risultano essere più influenti quando le temperature sono basse. Gli altri parametri influenti per il sito di Yangling sono due parametri legati all'inizio ed alla fine della senescenza: la temperatura soglia all'emergenza (Pfen_PartA) e i GDD necessari a completare la senescenza (SMT_sen).

In Fig. 6.4 sono riassunti i risultati della SA svolta con EFAST per entrambi i modelli. Ogni colonna rappresenta la somma degli indici S_i (parte grigia) e ST_i (parte bianca). La linea orizzontale impostata a 0.15 rappresenta la soglia arbitraria di influenza.

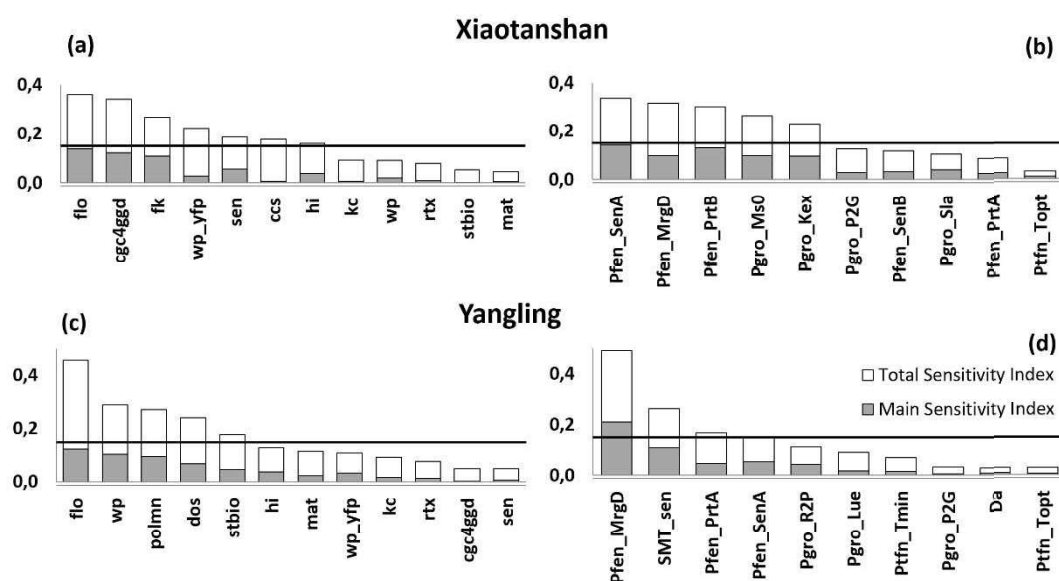


Fig. 6.17: Media degli indici di sensibilità di EFAST principale (S_i , in grigio) e totale (ST_i , in bianco): (a) SA per Aquacrop con dataset di Xiaotanshan; (b) SA per SAFY con dataset di Xiaotanshan; (c) SA per Aquacrop con dataset di Yangling; (d) SA per SAFY con dataset di Yangling.

6.3.3. Plasticità e complessità di Aquacrop e SAFY

La plasticità e la complessità di Aquacrop e SAFY sono state valutate calcolando l'indice L (6.21) proposto da Confalonieri et al. (2012). I risultati sono riassunti in Tab. 6.3, in cui è mostrato che la plasticità dei due modelli varia tra un sito e l'altro.

Tab. 6.10: Deviazione standard dell'indicatore agrometeorologico sintetico calcolato per i differenti anni e siti e la media per tutti i dataset climatici utilizzati. In corrispondenza i valori del "Top-Down concordance coefficient (TDCC) e plasticità L calcolato per i due modelli.

Site	σ_{SAM}	TDCC		L	
		SAFY	Aquacrop	SAFY	Aquacrop
Xiaotanshan	0.065	0.968	0.951	0.380	0.373

Yangling	0.105	0.368	0.614	0.150	0.251
mean	0.127	0.405	0.716	0.169	0.299

La varianza del indice SAM è più alta per Yangling che per Xiaotanshan, ciò indica una maggiore variabilità climatica per gli anni esaminati inerenti al sito di Yangling. Ciò sembra aver influenzato i risultati della SA, infatti, i valori di TDCC, i quali esprimono la correlazione tra le classifiche di influenza dei parametri, sono alti (quindi poco plastici) per Xiaotanshan, e piuttosto bassi per Yangling. La plasticità è quindi più bassa in Xiaotanshan che in Yangling, e le differenze tra i valori di L di Aquacrop e SAFY non sono consistenti tra un sito e l'altro, ma molto evidenti tra un modello e l'altro, almeno per quanto riguarda il sito di Yangling. Mediamente, il valore di plasticità (L) è risultato più alto per SAFY che per Aquacrop.

La complessità di questi due modelli di crescita delle colture è stata misurata usando il rapporto R_p (6.20) presentato da Confalonieri et al. (2009). I parametri sono stati considerati rilevanti se la somma degli indici di sensibilità di EFAST è maggiore del valore di soglia, arbitrariamente fissato a 0.15. Questo valore è stato fissato esclusivamente per quantificare e comparare la complessità di Aquacrop e SAFY. Risulta che per Aquacrop $R_p = 0.11$, indicando che i parametri rilevanti sono mediamente l'11% dei parametri totali, mentre per SAFY è stato ottenuto un valore di $R_p = 0.08$.

6.4. Discussione sulla GSA dei modelli Aquacrop e SAFY

È noto dalla letteratura che la sensibilità di un modello dipende dalle condizioni al contorno applicate. In questo caso specifico, le condizioni al contorno consistono nel dataset climatico e nell'intervallo di variazione dei parametri. Dalla letteratura (Shin et al., 2013) è noto inoltre che gli intervalli di variazione dei parametri influenzano la loro classifica di sensibilità. Sia i parametri di Aquacrop che SAFY sono stati inizialmente calibrati seguendo le indicazioni trovate in letteratura (Jin et al., 2014, Duchemin et al., 2007). Gli intervalli dei parametri inerenti la coltura scelti in questo lavoro sono molto ampi, maggiori rispetto a quelli usati generalmente in letteratura, come ad esempio in Vanuytrecht et al (2014), poiché si intendeva includere tutti i valori plausibili, sia per la coltura analizzata sia per i tipi di suolo presenti negli scenari presi in considerazione. Gli intervalli dei parametri del suolo sono stati scelti

uguali per tutti i test, in modo da ridurre la variabilità delle condizioni al contorno. Scopo principale di questa SA era ottenere i parametri più rilevanti per il sito di Yangling, sito di interesse per lo studio dei metodi di assimilazione, principale obiettivo di questa tesi, pertanto è su quel dataset che si è focalizzata maggiormente l'attenzione. Lo studio è stato ampliato anche al sito di Xiaotanshan perché per quel sito si avevano più informazioni di campo e per dare una maggiore validità ai risultati ottenuti, pur tenendo presente che sono strettamente legati ai dataset impiegati in questo lavoro. In virtù di quanto detto, le condizioni climatiche sono dunque le uniche variabili che distinguono uno scenario dall'altro, in modo da poter più semplicemente distinguere se i parametri sono sempre influenti o sempre trascurabili, o se il loro grado di influenza varia in funzione del dataset climatico. I risultati ottenuti confermano quanto trovato in letteratura (Vanuytrecht et al., 2014): applicando il metodo di Morris è possibile identificare facilmente i parametri trascurabili, essendo Morris un metodo rapido da applicare. Con riferimento al modello Aquacrop, l'insieme dei parametri trascurabili identificati in questo studio coincide quasi completamente con quelli identificati da Vanuytrecht et al. (2014), eccetto per il parametro indicante il coefficiente di declino della canopy (cdc), il quale è risultato tra i più influenti in letteratura e trascurabile in questo studio. Ciò è dovuto al fatto che cdc dipende dal tipo di coltura. Poiché in questo studio ci si è limitati ad analizzare il comportamento dei modelli per il frumento tenero, esso varia in un intervallo più piccolo, per questo non influenza significativamente la produzione simulata dal modello, mentre in Vanuytrecht et al. (2014) esso varia in un intervallo più ampio, adatto a rappresentare diversi tipi di coltura.

L'insieme dei parametri trascurabili di Aquacrop elencato nel paragrafo 6.3.1 ha generalmente un impatto limitato sulla variazione degli output, per questo è possibile escluderli dai metodi di calibrazione o di assimilazione o da ulteriori studi di sensibilità che si intende applicare a scenari diversi da quelli utilizzati in questo studio. In accordo con la letteratura (Confalonieri, 2010; Vanuytrecht et al., 2014), i risultati di questo studio confermano che l'applicazione del metodo di Morris consente di identificare in modo affidabile i parametri trascurabili di un modello. Questo è anche confermato dall'analisi di sensibilità fatta per SAFY. Come mostrato nel paragrafo 6.3.1, per questo modello è stato individuato un gruppo di parametri trascurabili univoco per tutti i test svolti. Bisogna tuttavia essere prudenti nell'utilizzare i risultati ottenuti in questo studio per il modello SAFY, in quanto mancando uno studio di SA

in letteratura non è possibile effettuare un confronto con scenari diversi da quelli analizzati, quindi non si può dire con certezza che la trascurabilità di questi parametri sia valida per qualsiasi scenario.

Il metodo di Morris ha dato prova di essere efficiente nell'identificazione dei parametri trascurabili, ma non egualmente per la classificazione della sensibilità dei parametri. Guardando la classifica della sensibilità dei parametri rispetto all'indice di Morris μ^* (Fig. 6.1), è possibile notare che le condizioni climatiche hanno un effetto più incisivo per alcuni parametri che per altri. Ad esempio, analizzando i risultati ottenuti per Aquacrop (Fig. 6.1 (a) e Fig. 6.1 (c)), è possibile osservare che il parametro flo (giorno di inizio della fioritura in GDD) presenta per tutti i test un valore di μ^* approssimativamente uguale a $13 \text{ t} \cdot \text{ha}^{-1}$, eccetto per il test effettuato con il dataset di Xiaotanshan 2010-2011, in cui risulta essere $8.5 \text{ t} \cdot \text{ha}^{-1}$ e Yangling 2013-2014, in cui risulta essere $21 \text{ t} \cdot \text{ha}^{-1}$. Ciò indica che, ad esempio, i dati climatici influenzano il parametro flo maggiormente di un parametro come sen (che indica l'inizio della senescenza in DGG), per cui il valore di μ^* è costantemente circa $7 \text{ t} \cdot \text{ha}^{-1}$ per tutti i test effettuati. L'affidabilità della classifica della sensibilità dei parametri è confermata dal metodo EFAST. Particolarmente per Aquacrop è possibile stabilire un gruppo di parametri influenti per tutti gli scenari considerati: flo, sen, hi, wp_yfp, wp, rtx, kc e stbio (Tab. 6.1). Questi parametri sono riconosciuti come influenti anche in uno studio precedente (Vanuytrecht et al., 2014), anche se con differente valore degli indici di sensibilità.

Per SAFY invece, gli indici di sensibilità sono concordi solo all'interno dello stesso sito sperimentale, ma variano significativamente tra un sito e l'altro. Non è possibile identificare un unico gruppo di parametri influenti valido per entrambi gli scenari. Gli unici a rientrare nel gruppo dei parametri influenti per entrambi gli scenari sono PfenMrgD e PfenSenA (Tab. 6.2). EFAST ha confermato di essere un metodo robusto, ma ha evidenziato che i risultati della SA sono fortemente legati alle condizioni climatiche alle caratteristiche dello scenario, rendendo impossibile classificare in modo universale la sensibilità dei parametri.

Per entrambi i modelli utilizzati risulta che i parametri descrittivi la fenologia della pianta giocano sempre un ruolo decisivo nella stima della produzione di granella, essendo significativamente più influenti degli altri gruppi di parametri. In particolare l'inizio del ciclo colturale, rappresentato dalla data di semina per Aquacrop (dos) e dalla data dell'emergenza per SAFY (Pgro_MrgD), sono risultati essere sempre molto

influenti. La coincidenza delle fasi fenologiche con periodi dell'anno con determinate precipitazioni e temperature sembra essere la causa che principalmente influisce sulle variazioni dell'output.

Come atteso, la SA effettuata per Aquacrop è stata più complessa che quella fatta per SAFY. Aquacrop descrive più accuratamente di SAFY i processi di crescita della coltura, usando un numero maggiore di equazioni e di fattori di input. Pertanto la SA effettuata per il primo modello ha un costo computazionale maggiore che quella effettuata per il secondo. Usando un computer con un processore Intel(R) Core(TM) i7-4770 CPU a 3.40 GHz e 16 GB of RAM, il tempo di calcolo per l'esecuzione di EFAST su Aquacrop è stato nell'ordine delle ore, mentre per SAFY nell'ordine dei minuti.

La SA è uno step necessario per applicare qualsiasi metodo di assimilazione, quindi il costo computazionale di questa operazione potrebbe essere un fattore che influenza la scelta del modello da usare.

La SA eseguita in questo studio ha permesso di calcolare la plasticità e la complessità dei due modelli di crescita delle colture. La complessità, definita come il rapporto tra parametri influenti e parametri totali, è risultata maggiore per Aquacrop che per SAFY. La plasticità, quantifica quanto le condizioni climatiche cambino l'influenza sull'output di gruppi di parametri. Confalonieri et al. (2012), paragonando tre modelli, hanno riscontrato che il modello con la maggiore complessità (in quel caso WOFOST) era quello con la maggiore plasticità. Essi hanno considerato che un elevato valore di plasticità è un aspetto positivo per un modello di crescita delle colture, in quanto riflette la sua capacità di cambiare comportamento in risposta alle diverse condizioni ambientali, riflettendo apparentemente la plasticità fenotipica tipica delle colture (Confalonieri et al., 2012). Nel contesto dell'assimilazione dati, o di applicazioni su scala regionale del modello, una alta plasticità da un lato potrebbe complicare l'individuazione dei parametri per cui sia più necessario ridurre l'incertezza, d'altro canto però un modello con una plasticità bassa necessita di una calibrazione più accurata.

In questo studio la plasticità varia per entrambi i modelli e per entrambi i siti sperimentali, come conseguenza della diversa variabilità dei modelli climatici. Per ottenere una valutazione accurata della plasticità sarebbe necessario usare un campione più grande di siti e di cicli colturali. Mediamente risulta che SAFY ha una plasticità

più alta che Aquacrop, sebbene sia meno complesso, mostrando che non sempre i modelli con complessità maggiore hanno anche maggiore plasticità.

La motivazione principale che ha mosso la SA svolta in questo studio è dovuta all'esigenza dettata dai metodi di assimilazione di ridurre i parametri da ricalibrare, e di utilizzare pertanto quelli che maggiormente influenzano l'output del modello, che in questo caso specifico è la produzione di granella. Per tale ragione gli sforzi sono stati principalmente concentrati nello studio del comportamento dei modelli Aquacrop e SAFY per le condizioni climatiche ed ambientali utilizzate in seguito per l'assimilazione, ovvero quelle del sito sperimentale di Yangling per i cicli colturali compresi tra settembre 2012 e giugno 2015. La scelta di introdurre anche il sito sperimentale di Xiaotanshan è stata dettata dall'esigenza di effettuare un confronto tra i due modelli scelti, cercando di compararne la plasticità e la complessità, seppur in modo qualitativo a causa del ridotto numero di scenari utilizzati.

La lista dei parametri influenti è stata individuata combinando due diversi metodi di SA: il metodo di Morris, per una preliminare individuazione dei parametri trascurabili, e il metodo EFAST, computato escludendo i parametri trascurabili individuati in precedenza. Da tale studio si evince che i parametri che descrivono la fenologia sono sempre molto influenti nello stimare la produzione di granella, in particolare quelli che caratterizzano l'inizio del ciclo (giorno della semina per Aquacrop e giorno dell'emergenza per SAFY). Pertanto la conoscenza del giorno di semina ed emergenza permette di incrementare inequivocabilmente l'accuratezza della stima di produzione.

Nel confronto tra i due modelli è risultato che SAFY è leggermente meno complesso di Aquacrop. Ciò comporta una maggiore semplicità nella calibrazione e nell'uso di tale modello in combinazione con altri algoritmi, quali possono essere appunto i metodi di assimilazione. SAFY, inoltre, ha una plasticità maggiore di Aquacrop. La plasticità può essere considerata un'indicazione di come, in condizioni diverse, i processi relativi ai parametri cambiano la loro importanza relativa nell'influenzare l'output. Avere un'idea della plasticità di un modello è utile quando questo è usato per analizzare il sistema che rappresenta, ad esempio per capire quali sono le caratteristiche delle colture (i.e. i parametri del modello inerenti la coltura) che influenzano maggiormente la produttività, per uno scenario con specifiche caratteristiche climatiche, pedologiche o di gestione. Nel caso di modelli agro

ambientali questo assume una grande importanza, a causa della plasticità fenotipica pronunciata tipica delle entità modellate (Confalonieri et al., 2012).

7. Metodi di assimilazione dati in modelli di accrescimento delle colture

Un'introduzione sui metodi di assimilazione, in particolare sui metodi di assimilazione dati nei modelli di accrescimento delle colture è stata fornita nel paragrafo 2.3. Di seguito sono indicati nel dettaglio i metodi di assimilazione sviluppati per questa tesi e applicati al set di dati raccolto nella zona rurale di Yangling, descritto nel capitolo 3.

7.1. Metodo di assimilazione *EnKF* usato per *SAFY*

Per integrare i dati ottenuti da telerilevamento riguardanti il LAI del frumento tenero con il modello *SAFY*, è stato utilizzato un metodo di assimilazione di tipo *updating* basato sul Ensemble Kalman Filter. Il principio su cui si basa questo metodo è di “aggiornare” l'andamento del LAI simulato dal modello *SAFY* all'istante corrispondente alla data dell'acquisizione da telerilevamento, con il dato di LAI stimato tramite il metodo Prosail-ANN (paragrafo 4.2). In realtà il valore di LAI assimilato non è esattamente il valore di LAI misurato, ma quello “filtrato” dall'errore sulla misura e dall'errore sul modello.

Prima di applicare il metodo è stato necessario identificare per mezzo dell'analisi di sensibilità i parametri che maggiormente influenzano la produzione di granella simulata dal modello (capitolo 6). In Tab. 7.1 è mostrata la lista dei parametri principali così individuati ed il relativo intervallo in cui sono fatti variare durante l'assimilazione. Una prima calibrazione di tutti i parametri è stata fatta utilizzando i valori indicati in letteratura per il frumento tenero (Duchemin et al., 2007). Successivamente, per i parametri ritenuti più influenti, è stato definito un intervallo di variazione in un intorno del valore di riferimento, impostando tale valore come valore medio dell'intervallo ed impostando il valore minimo e massimo arbitrariamente, tenendo in considerazione le informazioni a disposizione sul sito in esame, i valori scelti in letteratura (Duchemin et al., 2007) e il parere di esperti.

Tabella 7.1: Lista dei parametri più influenti del modello *SAFY* ricalibrati durante l'assimilazione. Tale lista è risultata dalla SA descritta nel capitolo 6.

Parameter	Min	Max	Description
-----------	-----	-----	-------------

Pgro_R2P	0.7	0.9	Climatic efficiency: ratio of incoming PAR to global radiation
Pgro_Lue	1.75	1.95	Effective light-use efficiency: ratio of energy produced as DAM from APAR (g·MJ ⁻¹)
Pgro_P2G	0.0098	0.012	Partition coefficient to grain
Pfen_MrgD	288	300	Day of the year of emergence
Pfen_PrtA	0.16	0.2	Make vary the origin and slope of the partitioning function
Pfen_SenA	1260	1540	Temperature threshold to start senescence (°C)
Ptfn_Topt	20	22	Optimum temperature for plant development (°C)
Da	1.35	1.55	Soil bulk density (g cm ⁻³)
SMT_sen	2070	2530	GDD to complete senescence

Il metodo di assimilazione Ensemble Kalman Filter sviluppato per SAFY è basato sulla teoria sviluppata da Burgers et al. (1998) e segue l'applicazione ai modelli di crescita delle colture descritta da Makowski et al. (2006). È possibile suddividere l'algoritmo di assimilazione nelle seguenti fasi:

- 1) Generazione di un insieme (anche detto ensemble) e di n vettori V (in questo caso specifico si è impiegato n=100). Ogni vettore contiene il valore di ogni parametro P_i incluso in Tab 7.1, in questo caso i = 1,2,...,9. Ogni singolo elemento di ciascun vettore V^j (con j=1,...,n) è composto dal valore nominale di P_i, ovvero il valore medio dell'intervallo di variazione mostrato in Tab. 7.1 più un errore e_i^j, estratto casualmente da una distribuzione normale troncata N(0,1) con limiti inferiore e superiore pari al minimo ed al massimo dell'intervallo corrispondente (7.1):

$$e = [V^1, V^2, \dots, V^n] = \begin{bmatrix} P_1 + e_1^1 & P_1 + e_1^2 & \dots & P_1 + e_1^n \\ P_2 + e_2^1 & P_2 + e_2^2 & \dots & P_2 + e_2^n \\ \dots & \dots & \dots & \dots \\ P_9 + e_9^1 & P_9 + e_9^2 & \dots & P_9 + e_9^n \end{bmatrix} = \begin{bmatrix} P_1^1 & P_1^2 & \dots & P_1^n \\ P_2^1 & P_2^2 & \dots & P_2^n \\ \dots & \dots & \dots & \dots \\ P_9^1 & P_9^2 & \dots & P_9^n \end{bmatrix} \quad (7.5)$$

- 2) Simulazione con il modello SAFY con il fine di ottenere per ogni vettore dell'ensemble il valore del LAI simulato dal modello alla data della prima osservazione di LAI (cioè acquisizione satellitare), utilizzando per ogni simulazione del modello il set di parametri che compone ogni vettore V^j

dell'ensemble. A tale valore di LAI viene aggiunto un errore ε^j . L'errore ε^j è generato casualmente da una distribuzione normale $N(0, Q)$, in cui la deviazione standard Q quantifica l'incertezza del modello ed è stata arbitrariamente impostata nel presente studio al 20% del valore nominale dei parametri.

Si definisce in questo modo la matrice φ_{t_1} :

$$\varphi_{t_1} = \begin{bmatrix} L^1 + \varepsilon^1 & L^2 + \varepsilon^2 & \dots & L^n + \varepsilon^n \\ P_1^1 & P_1^2 & \dots & P_1^n \\ \dots & \dots & \dots & \dots \\ P_9^1 & P_9^2 & \dots & P_9^n \end{bmatrix} \quad (7.6)$$

per la quale ogni colonna è un vettore V^j dell'ensemble e rappresenta una configurazione casuale del modello per il set di parametri P^j (composto dai nove parametri di Tab. 7.1) e il corrispondente valore di LAI (compreso di errore del modello) al tempo t_1 corrispondente alla data della prima acquisizione dell'immagine per la stima del LAI da RS.

3) Generazione di un ensemble di misure tale che

$$eM_{t_1} = [M_{t_1}^1 \quad M_{t_1}^2 \quad \dots \quad M_{t_1}^n] \quad (7.7)$$

per il quale ogni elemento $M_{t_1}^j$ è composto dal LAI osservato al tempo t_1 ,

i.e. ricavato da dati ottenuti tramite RS, più un errore $\tau_{t_1}^j$:

$$M_{t_1}^j = M_{t_1} + \tau_{t_1}^j \quad (7.8)$$

L'errore $\tau_{t_1}^j$ è estratto casualmente dalla distribuzione gaussiana

$N[0, \text{var}(\tau_{t_1}^j)]$, dove $\text{var}(\tau_{t_1}^j)$ è una varianza che rappresenta l'errore di misura,

stimato dalla comparazione tra LAI ottenuto tramite l'algoritmo Prosail-ANN e quello misurato in situ durante le campagne di misura.

In questo studio la varianza dell'errore di misura è stato scelto come espresso dalla (7.5):

$$\sqrt{\text{var}(\tau_t)} = 0.29 \cdot M_t \quad (7.9)$$

4) Calcolo della matrice di varianza-covarianza ($\Sigma_{t_1}^j$) di φ_{t_1} per gli n ensemble:

$$\Sigma_{t_1}^j = \begin{bmatrix} \text{var}(L_t^j) & \text{cov}(P_1^j L_t^j) & \text{cov}(P_2^j L_t^j) & \dots & \text{cov}(P_n^j L_t^j) \\ \text{cov}(L_t^j P_1^j) & \text{var}(P_1^j) & \text{cov}(P_2^j P_1^j) & \dots & \text{cov}(P_n^j P_1^j) \\ \text{cov}(L_t^j P_2^j) & \text{cov}(P_1^j P_2^j) & \text{var}(P_2^j) & \dots & \text{cov}(P_n^j P_2^j) \\ \dots & \dots & \dots & \dots & \dots \\ \text{cov}(L_t^j P_9^j) & \text{cov}(P_1^j P_9^j) & \text{cov}(P_2^j P_9^j) & \dots & \text{var}(P_9^j) \end{bmatrix} \quad (7.10)$$

5) Calcolo del guadagno di Kalman per mezzo della matrice di varianza-covarianza:

$$K_{t_1} = \frac{\sum_{t_1} R^T}{R \sum R^T + \text{var}(\tau_{t_1})} \quad (7.11)$$

Dove $R=(1,0,0,\dots,0)$, con numero di elementi pari a n.

6) Aggiornamento di φ_{t_1} come mostrato dalla (7.8):

$$\varphi_{t_1,K}^j = \varphi_{t_1}^j + K_{t_1} (M_{t_1}^j - R\varphi_{t_1}^j) \quad (7.8)$$

Dove j è ancora una volta la j-esima colonna di φ_{t_1} (o anche l'indice del vettore dell'ensemble)

7) Sostituzione del LAI e del valore dei 9 parametri fondamentali (i.e. degli elementi di φ_{t_1}) con quelli calcolati allo step 6 (gli elementi di $\varphi_{t_1,K}$)

8) Ripetizione dell'algoritmo dal punto 3, per ogni data di acquisizione delle immagini da RS. Tutti gli step vengono ripetuti per ogni istante $t=[t_1, t_2, \dots, t_f]$ dove t_f è l'ultima data di acquisizione. Quando gli ultimi dati sono stati assimilati, il modello SAFY viene fatto eseguire fino alla fine del ciclo colturale, ottenendo finalmente come output, l'andamento del LAI, della biomassa e la produzione di granella.

In Figura 7.1, a titolo di esempio, è riportato l'andamento del LAI, della biomassa della produzione di granella simulata da SAFY a seguito dell'assimilazione per updating per mezzo del metodo basato sul EnKF, dove sono evidenti le discontinuità nell'andamento delle variabili biofisiche in corrispondenza delle assimilazioni.

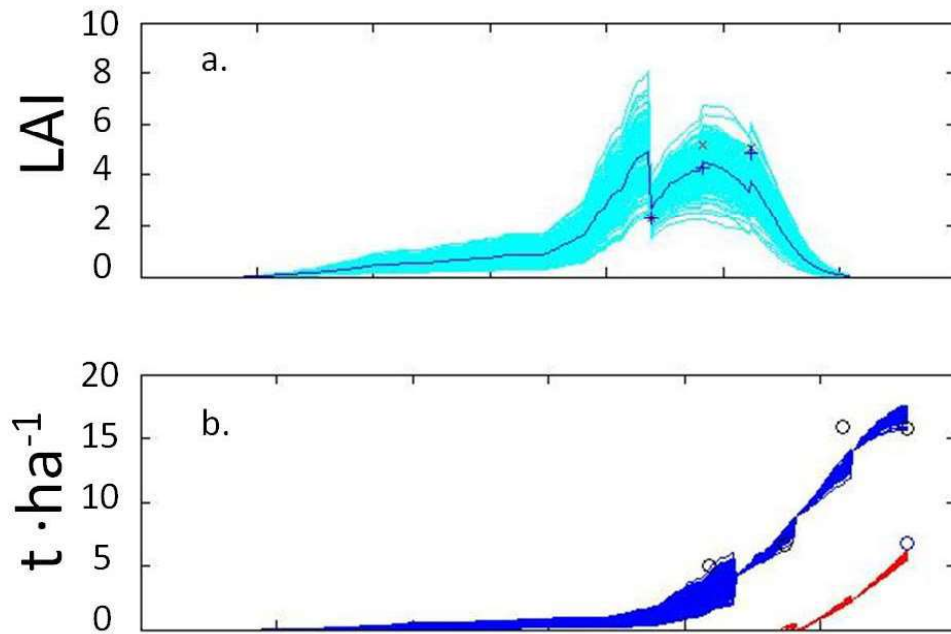


Fig. 7.18: Andamento delle variabili biofisiche simulate dal modello SAFY a seguito dell'applicazione del metodo di assimilazione basato sul EnKF fatta su Yangling per il dataset 2013-2014. a. andamento del LAI per ogni combinazione dei parametri dell'ensemble (linee in azzurro), valore medio tra tutti gli andamenti di LAI dell'ensemble (linea blu), valore puntuale di LAI misurato in campo (+ in blu), valore puntuale di LAI stimato da RS e utilizzato per l'assimilazione (x in rosso). b. andamento della biomassa misurata in $t \cdot ha^{-1}$ per ogni combinazione dei parametri dell'ensemble (linee in blu), andamento della produzione di granella misurato in $t \cdot ha^{-1}$ per ogni combinazione dei parametri dell'ensemble (linee in rosso), valori puntuali di biomassa misurati in campo (cerchi in nero), valore finale di produzione di granella (cerchio in blu).

7.2. Metodo di assimilazione PSO usato per Aquacrop

Il modello Aquacrop, nella versione utilizzata, aveva un codice non accessibile a modifiche di programmazione, che consentissero d'introdurre la possibilità di arresto e di riavvio del modello a ciascuna data di assimilazione delle variabili osservate. Pertanto non è stato possibile apportare le modifiche necessarie per applicare un metodo di assimilazione per updating come l'EnKF visto in precedenza. Per questa ragione, con il modello Aquacrop è stato applicato un metodo di assimilazione per calibrazione (Dorigo et al., 2007). L'idea di base è di utilizzare un metodo di ottimizzazione che ricalibri i parametri al fine di minimizzare l'errore tra il CC stimato da RS e quello simulato dal modello. Inizialmente era stato scelto di impiegare il metodo di ottimizzazione del Simplexso (Dantzig, 1947) per minimizzare il Mean Square Error (MSE) tra il CC simulato (CC_s) e il CC stimato da RS (CC_r):

$$MSE = \sum_{i=1}^n \frac{(CC_s(i) - CC_r(i))^2}{n} \quad (7.9)$$

Tale metodo di ottimizzazione aveva un costo computazionale troppo elevato. Pertanto è stato scelto di utilizzare il metodo Particle Swarm Optimization (PSO), il quale è basato su un principio relativamente semplice, richiede pochi parametri di input, può essere facilmente applicato con altri modelli ed ha mostrato di avere un'elevata efficienza di calcolo (Eberhart et Kennedy, 1995). Il concetto di base nel metodo PSO è che un gruppo di *m particles* ($m = 15$ in questo studio), si muove senza uno schema preciso all'interno di uno spazio di ricerca (in inglese *search space*) a d dimensioni. Il numero di *particles* scelto in questo studio è pari a 15, un valore più basso di quello suggerito in letteratura per garantire un funzionamento ottimale del metodo (Li et al., 2015). Tale scelta è stata fatta perché il metodo PSO applicato ad Aquacrop con un numero di *particles* (cioè di particelle) superiore a 15 richiedeva delle considerevoli risorse computazionali non disponibili per questo studio. Ogni *particle* può modificare la sua posizione e velocità in base sia al miglior punto di generazione corrente (p_{id}) sia in base al punto migliore tra tutti i *particles* generate nello sciame (p_{gd}). Il metodo di assimilazione PSO è descritto nel dettaglio da Li et al. (2015), la sua applicazione per Aquacrop è riassunta nei seguenti passaggi:

- 1) La posizione iniziale (o valore iniziale) e la velocità di ogni particella è stabilita come suggerito da letteratura (Li et al., 2015). I parametri sono inizialmente calibrati seguendo le indicazioni trovate in letteratura (Jin et al., 2014) mentre i parametri da ricalibrare sono selezionati dall'analisi di sensibilità (capitolo 6) e sono elencati in tabella 7. I dati di input utilizzati sono quelli precedentemente descritti nel capitolo 3.
- 2) Il modello Aquacrop è fatto girare tramite il file eseguibile distribuito dalla FAO e controllato tramite uno script in Matlab.
- 3) Una funzione costo J è stata definita per valutare la discrepanza tra il CC simulato (CC_s) e quello stimato tramite RS (CC_r) come mostrato nell'equazione (7.10). La minimizzazione della funzione J è determinata se l'algoritmo raggiunge i valori dei parametri di input ottimali.

$$J = \sqrt{\sum_{i=1}^N \left(\frac{CC_{s_i} - CC_{r_i}}{CC_{r_i}} \right)^2} \quad (7.10)$$

4) Viene ricercato il miglior punto del'i-esimo particle (p_{id}) e di tutti i particles dello sciame (p_{gd}). I p_{id} e p_{gd} sono ricercati ad ogni iterazione.

5) La posizione e la velocità dei particles è aggiornata come mostrato dall'equazione (7.11):

$$\begin{aligned} v_{id}^{k+1} &= v_{id}^k + c_1 \xi (p_{id}^k - x_{id}^k) + c_2 \eta (p_{gd}^k - x_{id}^k) \\ x_{id}^{k+1} &= x_{id}^k + v_{id}^{k+1} \end{aligned} \quad (7.11)$$

6) Se l'obiettivo dell'iterazione non è raggiunto entro il numero massimo di iterazioni prefissato, la posizione è sostituita dalla successiva (i parametri quindi assumono valori differenti dipendenti dalla nuova posizione e dalla velocità di movimento nello spazio di d-dimensioni in cui sono definiti) e l'algoritmo riprende dallo step 2. In questo studio, per limiti di potenza computazionale, il limite massimo è stato fissato a 30, indicato in letteratura (Li et al., 2015) come il numero minimo di iterazioni per l'applicazione del metodo PSO. Scegliere un numero limite di iterazioni maggiore non avrebbe consentito di eseguire lo script in un tempo accettabile.

7) Se l'obiettivo dell'iterazione è raggiunto, il CC, la biomassa e la produzione sono finalmente simulati.

Tab. 7.11: Lista dei parametri più influenti del modello Aquacrop variati durante l'assimilazione per calibrazione applicando il metodo PSO. L'intervallo di variazione è stato definito sullo stesso criterio di quanto fatto per i parametri variati nell'assimilazione per updating nel modello SAFY tramite il metodo EnKF.

Parameter	Min	Max	Mean	Description
flo	1500	1700	1600	Growing Degree Days (GDD) from sowing to flowering
wp	16	20	18	Water productivity normalized for ETo and CO ₂ (g m ⁻²)
polmn	6	10	8	Minimum air temperature below which pollination starts to fail (°C)
dos	274	284	279	Day of sowing (Day of year)
stbio	5	16	11	Minimum GDD required for full biomass production
hi	35	45	40	Reference Harvest Index (HI ₀) (%)
mat	2200	2600	2400	Length of the crop cycle (GDD)
kc	0.08	1.3	1.1	Crop coefficient when canopy is fully covering the ground but prior to senescence
rtx	1	1.4	1.2	Maximum effective rooting depth (m)

7.3. Applicazione puntuale e spazializzata dei metodi di assimilazione

I metodi di assimilazione descritti in precedenza sono stati applicati sia su scala di campo che su scala regionale.

Per l'applicazione su scala di campo, i metodi sono stati semplicemente applicati puntualmente, i.e. sono stati utilizzati i dati climatici raccolti su scala regionale (come indicato nel paragrafo 3.1) per stimare la produzione in ogni punto per i quali sono state effettuate le misure in campo. I dati di LAI assimilati con il metodo EnKF per SAFY e di CC con il PSO per Aquacrop, corrispondenti ai punti osservati in campo, sono stati estratti dalle mappe elaborate per mezzo dell'algoritmo Prosail-ANN. In questo modo è stato possibile validare la produzione stimata dai modelli a seguito dell'assimilazione, con quella misurata in campo durante le campagne di misura svolte tra il 2013 e il 2015 (paragrafo 3.2).

L'applicazione spazializzata invece è stata svolta esclusivamente per i cicli colturali 2013 e 2014. Questa scelta è stata fatta perché solo per questi due anni erano disponibili i dati ufficiali statistici di produzione di granella per la zona sperimentale interessata.

Sia per l'applicazione puntuale che spazializzata, le variabili biofisiche sono state assimilate usando i valori ricavati da 3 immagini acquisite tra marzo e maggio tramite sistemi di telerilevamento.

Per l'applicazione spazializzata, è stato necessario dapprima identificare le aree coltivate con frumento tenero. È stato quindi effettuato un processo di classificazione dei pixel compresi nell'immagine per stabilire se si trattava di suolo coltivato con frumento tenero. Per effettuare questa classificazione sono state utilizzate 5 immagini acquisite tra febbraio e giugno. Le immagini sono state convertite in mappe di LAI per mezzo dell'algoritmo Prosail-ANN (capitolo 4). Inizialmente tutte le aree agricole sono state identificate per mezzo di un file vettoriale che identificava i confini di ogni campo, in questo modo è stato possibile escludere immediatamente tutte le aree non agricole (centri abitati, strade, costruzioni, aree di vegetazione naturale). Questo file vettoriale è stato fornito dal National Engineering Centre for Information Technology in Agriculture, il quale da molto tempo gestisce l'area rurale di Yangling come area sperimentale. I restanti pixel sono stati sottoposti al seguente processo di classificazione, il quale stabiliva se un pixel era appartenente alla classe "superficie coltivata con frumento tenero" se soddisfaceva tutte le seguenti condizioni:

- LAI crescente tra la 1^a e la 3^a data di acquisizione;
- LAI della 3^a data di acquisizione superiore a 3
- LAI della 4^a data di acquisizione superiore a 3.4
- LAI della 5^a acquisizione più piccolo di 0.5 (data successiva alla mietitura)

Questi vincoli sono scelti in funzione del ciclo colturale tipico del frumento tenero e delle misure di LAI effettuate in situ. Tale procedura è stata applicata per una area ristretta rispetto a quella utilizzata per gli esperimenti in campo perché i dati statistici riferiti alla produzione di frumento tenero erano disponibili solo per l'area amministrativa di Yangling, mentre l'intera area sperimentale toccava aree amministrative più ampie per le quali non è stato possibile risalire a dati statistici ufficiali. Di tutti i punti osservati durante le campagne di misura, solo 8 erano inclusi nell'area amministrativa di Yangling e sono stati tutti classificati dalla procedura sopra descritta come punti con presenza di frumento tenero. Non tutti i pixel inclusi nei confini dei campi sono stati riconosciuti come coltivazioni di frumento, ma ogni campo per cui almeno l'80% dei punti è stato classificato come frumento è stato interamente considerato come tali. Per ridurre ulteriormente il costo computazionale degli algoritmi di assimilazione, per ogni immagine è stata effettuata un'ulteriore classificazione dei pixel basata sull'evoluzione del LAI nel tempo. I valori di LAI delle immagini sono stati arrotondati alla prima cifra decimale ed affinché fossero sempre multipli di 0.2. Il numero totale di classi è il prodotto tra il numero di intervalli tra il valore di LAI minimo e massimo per ogni immagine. In pratica, per ogni immagine viene definito un intervallo di valori di LAI:

$$\begin{aligned}
 I^1 &= [LAI_{\min}^1, LAI_{\min}^1 + 0.2, LAI_{\min}^1 + 2 * 0.2, \dots, LAI_{\min}^1 + x * 0.2, \dots, LAI_{\max}^1 - 2 * 0.2, LAI_{\max}^1 - 0.2, LAI_{\max}^1] \\
 I^2 &= [LAI_{\min}^2, LAI_{\min}^2 + 0.2, LAI_{\min}^2 + 2 * 0.2, \dots, LAI_{\min}^2 + x * 0.2, \dots, LAI_{\max}^2 - 2 * 0.2, LAI_{\max}^2 - 0.2, LAI_{\max}^2] \\
 I^3 &= [LAI_{\min}^3, LAI_{\min}^3 + 0.2, LAI_{\min}^3 + 2 * 0.2, \dots, LAI_{\min}^3 + x * 0.2, \dots, LAI_{\max}^3 - 2 * 0.2, LAI_{\max}^3 - 0.2, LAI_{\max}^3]
 \end{aligned}
 \tag{7.12}$$

dove ogni intervallo I^j è composto da n^j elementi pari a:

$$n^j = \frac{(LAI_{\max}^j - LAI_{\min}^j)}{0.2} + 1
 \tag{7.13}$$

con $LAI_{\max}$ e $LAI_{\min}$ arrotondati per difetto al multiplo di 0.2 più vicino. In questo modo ogni classe verrà definita dalla combinazione di tutti gli elementi dei 3 intervalli appena identificati:

$$\begin{cases}
C_1 = [LAI_{\min}^1, LAI_{\min}^2, LAI_{\min}^3]; \\
C_2 = [LAI_{\min}^1, LAI_{\min}^2, LAI_{\min}^3 + 0.2]; \\
C_3 = [LAI_{\min}^1, LAI_{\min}^2, LAI_{\min}^3 + 2 \cdot 0.2]; \\
\vdots \\
C_y = [LAI_{\min}^1, LAI_{\min}^2, LAI_{\max}^3]; \\
C_{y+1} = [LAI_{\min}^1, LAI_{\min}^2 + 0.2, LAI_{\min}^3]; \\
C_{y+2} = [LAI_{\min}^1, LAI_{\min}^2 + 0.2, LAI_{\min}^3 + 0.2]; \\
\vdots \\
C_N = [LAI_{\max}^1, LAI_{\max}^2, LAI_{\max}^3];
\end{cases} \quad (7.14)$$

Con N numero totale di classi. Una volta definite le N classi, e arrotondato per difetto ogni pixel delle 3 mappe di LAI al multiplo di 0.2 più vicino, si è proceduto con la classificazione, cioè, ogni punto è stato identificato per la sua serie di 3 valori di LAI (uno per ogni immagine) e inserito nella corrispettiva classe di appartenenza, per la quale tale serie risultava essere uguale. In questo modo è stato possibile ridurre il numero di esecuzioni dell'algoritmo da circa 2 milioni (ovvero il numero di pixel totale che compone ogni immagine acquisita da RS) a circa 1800 (il numero N di classi totale), riducendo enormemente i costi computazionali.

Una procedura simile è stata applicata per il CC, ma in questo caso l'intervallo di variazione tra una classe e la successiva non è costante, a causa della non linearità della relazione tra LAI e CC (eq. 3.5). Inoltre è stato necessario ampliare gli intervalli tra una classe e l'altra, in modo da ridurre il numero di classi totali, perché la lentezza computazionale del metodo PSO applicato ad Aquacrop ha imposto un numero limitato di iterazioni. Le 1800 classi identificate per il LAI sono state ridotte a circa 100 per il CC.

8. Risultati: applicazione dei metodi di assimilazione

8.1. Applicazione puntuale dei metodi di assimilazione

Prima di procedere all'applicazione su scala regionale, i due metodi di assimilazione, EnKF per SAFY e PSO per Aquacrop, sono stati applicati su scala di campo, validando i risultati con LAI e CC stimati dalle immagini satellitari per mezzo dell'algoritmo PROSAIL-ANN. Tale validazione è stata compiuta utilizzando i punti osservati al suolo durante le campagne di misura fatte tra il 2013 e il 2015 (paragrafo 3.2). La produzione di granella misurata dai campioni raccolti in situ è stata comparata con la produzione stimata dai due metodi di assimilazione.

In Allegato A sono mostrati gli andamenti di LAI, Biomassa e Produzione di Granella simulati dal metodo EnKF per SAFY per ogni punto di controllo a terra, riportando per ogni grafico i valori delle grandezze biofisiche misurati.

In Fig. 8.1 è mostrato il confronto tra la produzione stimata e misurata, per entrambi i metodi e i modelli usati. In Fig. 8.1a la produzione misurata in campo è confrontata con la produzione stimata applicando il metodo di assimilazione basato sul EnKF al modello SAFY.

Per questo metodo di assimilazione è stata effettuata una calibrazione dei parametri partendo dalle indicazioni fornite in letteratura (Duchemin et al., 2007), tali valori sono stati modificati manualmente affinché l'andamento del LAI misurato a terra per tre punti campione estratti casualmente coincidesse il più possibile con quello simulato dal modello. I tre punti campione sono stati scelti casualmente dal database di misure raccolte durante la campagna del 2014, poiché per questa sono state fatte il numero maggiori di osservazioni, sia per numero di punti che per date.

Fig. 8.1b e Fig.8.1c mostrano un confronto tra la produzione misurata e stimata utilizzando il metodo di assimilazione basato sul PSO applicato al modello Aquacrop. Per questo modello si è proceduto inizialmente come per il precedente, calibrando i parametri inizialmente partendo da informazioni fornite dalla letteratura (Jin et al., 2014), per poi modificarli in modo che l'andamento del CC simulato coincidesse con l'andamento del CC misurato in situ per gli stessi punti campione utilizzati in precedenza. I risultati dell'applicazione di questo metodo con questa procedura di calibrazione dei parametri sono mostrati in Fig. 8.1b, nella quale si può notare che la produzione varia in un intervallo compreso tra 4 e 6 t·ha⁻¹, con un RRMSE di 0.27. Applicando quindi un'unica calibrazione ai parametri, il modello Aquacrop a seguito

dell'assimilazione con metodo PSO sembra rispondere poco alle variazioni del dataset climatico. È stato scelto pertanto di effettuare una calibrazione specifica per ogni dataset climatico, utilizzando per ogni calibrazione punti campione presi dalle relative campagne di misura, quindi un punto dal database di un anno specifico per la calibrazione dei parametri utilizzati con il dataset climatico dello stesso anno. In Fig. 8.1c sono mostrati i risultati applicando al metodo PSO-Aquacrop una calibrazione specifica per ogni anno. Si evidenzia che in questo caso l'intervallo in cui varia la produzione si è leggermente ampliato, passando a degli estremi compresi tra 3 e 7 t·ha⁻¹, con una lieve riduzione anche del RRMSE.

Confrontando invece il metodo EnKF applicato a SAFY con il metodo PSO applicato ad Aquacrop, si nota che per il primo si ottiene un RRMSE di 0.18 mentre per il secondo (considerando il caso di calibrazione specifica) il valore di RRMSE è di 0.24. Questo significa che per gli scenari analizzati, l'assimilazione fatta usando il metodo EnKF con il modello SAFY stima meglio la produzione del metodo PSO applicato ad Aquacrop. È da notare che il modello Aquacrop in assenza di assimilazione è più performante del modello SAFY. Si può dedurre che il metodo di assimilazione basato sul EnKF consenta di giungere a risultati migliori rispetto al metodo basato sul PSO. Non è però da trascurare l'effetto di interazione metodo-modello, SAFY infatti è un modello semplice creato appositamente per interagire con dati provenienti da misure reali, mentre Aquacrop è un modello più sofisticato che non prevedeva inizialmente nessuna interazione con dati esterni. Altro confronto importante è relativo al tempo di esecuzione dei due metodi. Usando un PC con in Intel Core i7-4770 CPU a 3.4 GHz, il tempo di esecuzione del metodo PSO applicato ad Aquacrop è di circa 40 minuti per ogni punto simulato, mentre per il metodo EnKF applicato a SAFY è di circa 30 secondi per punto.

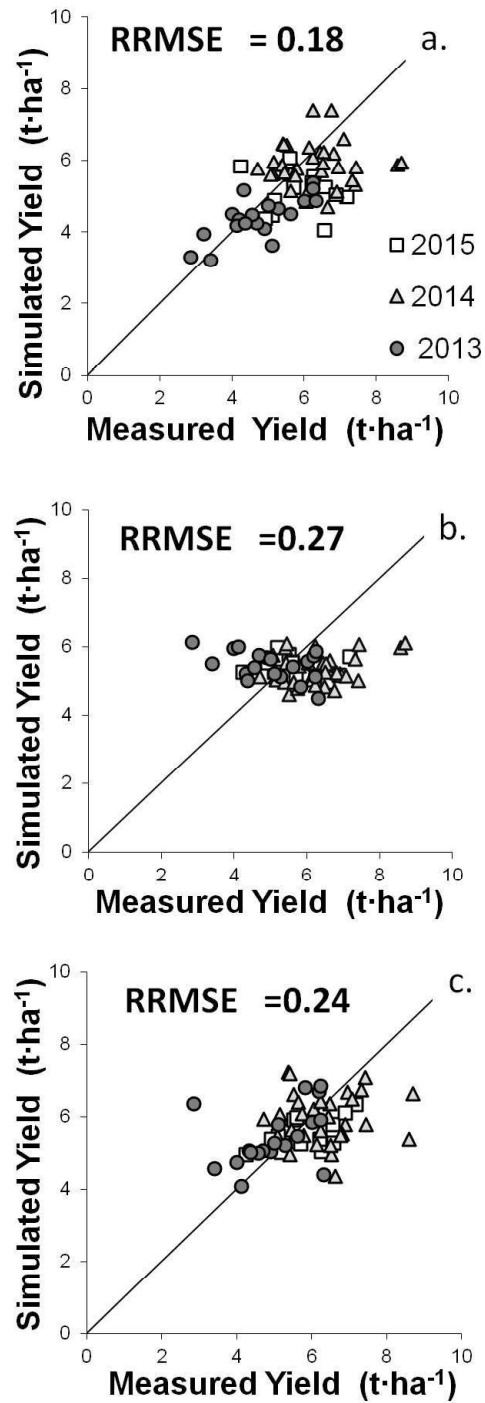


Fig. 8.19: Applicazione puntuale di due differenti tecniche di assimilazione: a. confronto tra produzione di frumento tenero misurato in situ e simulato utilizzando il metodo di assimilazione EnKF applicato al modello SAFY. b. confronto tra produzione di frumento tenero misurato in situ e simulato utilizzando il metodo di assimilazione PSO applicato ad Aquacrop a seguito di un'unica calibrazione. c. confronto tra produzione di frumento tenero misurato in situ e simulato utilizzando il metodo di assimilazione PSO applicato ad Aquacrop a seguito di una calibrazione specifica per ogni dataset climatico utilizzato per le simulazioni. È inoltre mostrata la retta 1:1.

8.2. Applicazione spazializzata dei metodi di assimilazione

L'applicazione spazializzata dei metodi di assimilazione è stata fatta col fine di stimare e mappare la variabilità di produzione di frumento tenero in un contesto a scala regionale. La stima spazializzata della produzione è stata limitata all'area amministrativa di Yangling. Questa scelta è stata imposta dal fatto che solo per questa area è stato possibile ottenere dati statistici ufficiali riguardo la produzione per ettaro, la produzione totale e la superficie totale seminata con frumento tenero. La validazione delle stime di produzione del frumento è stata pertanto possibile, utilizzando le informazioni fornite da due enti ufficiali di statistica: il National Bureau of Statistics of the People's Republic of China (il quale ha fornito i dati per il 2013) e lo Shaanxi Provincial Bureau of Statistics (il quale ha fornito i dati per il 2014). L'applicazione spazializzata è stata limitata alle aree in cui è stato coltivato frumento. Fig. 8.2 mostra le mappe di produzione stimata di frumento tenero per il 2013 (Fig.8.2 a.) e per il 2014 (Fig.8.2 b.) usando il metodo EnKF con il modello SAFY per l'area amministrativa di Yangling.

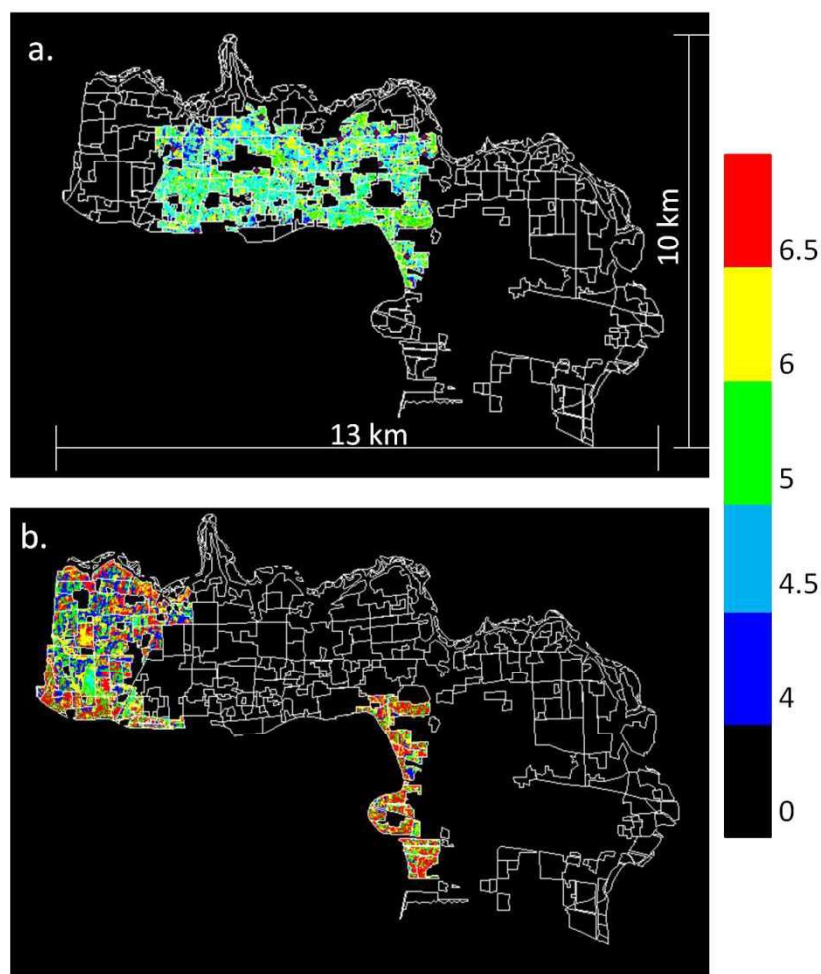


Fig. 8.20: Mappa di produzione di frumento tenero ($\text{t}\cdot\text{ha}^{-1}$) per Yangling stimata usando il metodo di assimilazione EnKF con il modello di crescita delle colture SAFY per il 2013 (a) e il 2014 (b).

I dati statistici ufficiali per il 2013 riportano una produzione media di $5.50 \text{ t}\cdot\text{ha}^{-1}$, una produzione totale di $1.09\cdot 10^4$ ed una superficie totale coltivata a frumento tenero di 1980 ha. Il metodo usato per individuare le superfici di coltivazioni di frumento tenero (paragrafo 7.3) ha fornito una stima di superficie coltivata a frumento di 1983 ha, pertanto questo metodo consente di identificare con buona approssimazione le coltivazioni di frumento tenero. La produzione media per unità di superficie stimata utilizzando il metodo EnKF con SAFY risulta essere di $4.98 \text{ t}\cdot\text{ha}^{-1}$, con una produzione totale di $0.99\cdot 10^4 \text{ t}$, una stima inferiore di circa il 9.4% rispetto ai dati statistici ufficiali.

Per il 2014 i dati statistici forniti dal National Bureau of Statistics of the People's Republic of China ha riportato una produzione per area di $6.13 \text{ t}\cdot\text{ha}^{-1}$, una produzione totale di $0.93\cdot 10^4 \text{ t}$ e una superficie totale di area seminata con frumento tenero pari a 1470 ha. Per questo stesso anno, la stima di superficie di campi di frumento condotta in questo studio è di 1564 ha. Confrontando il dato stimato con quello ufficiale, risulta quindi un errore di stima di circa il 7%. In questo caso la produzione per area è stata

di $5.67 \text{ t}\cdot\text{ha}^{-1}$, con una produzione totale di $0.82\cdot 10^4 \text{ t}$, pertanto per il 2014 risulta una sottostima di circa il 12.2% rispetto ai dati statistici ufficiali.

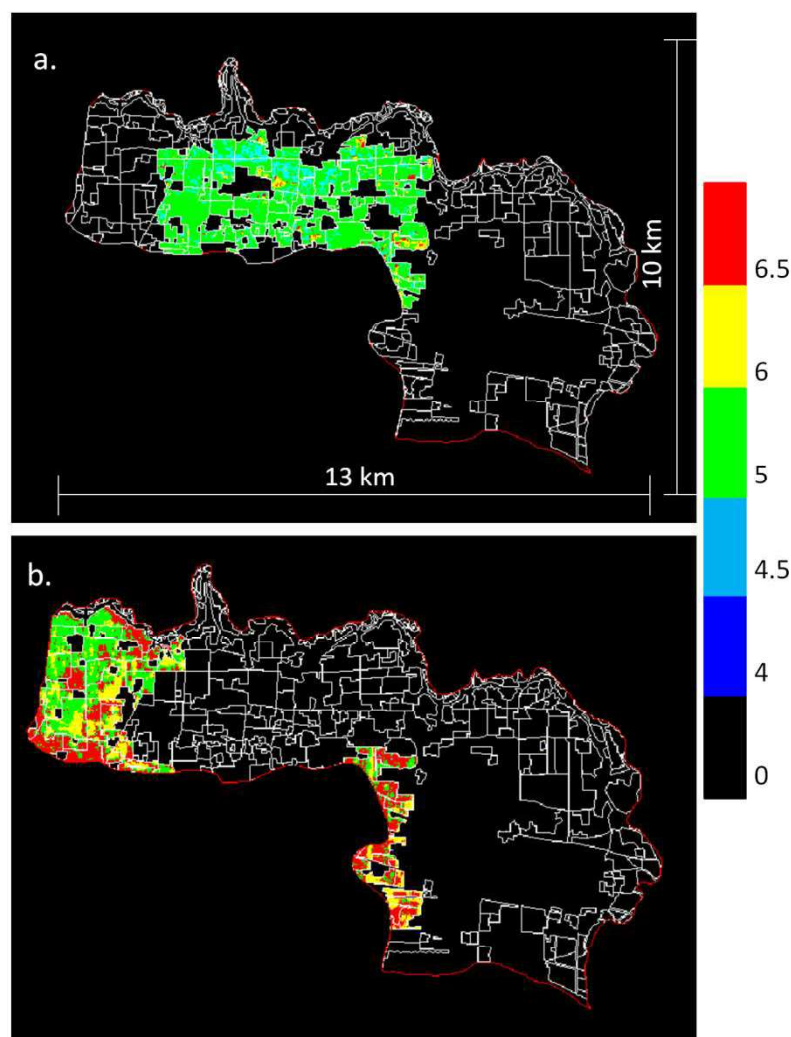


Fig. 8.21: Mappa di produzione di frumento tenero ($\text{t}\cdot\text{ha}^{-1}$) per Yangling, stimata usando il metodo di assimilazione PSO con il modello Aquacrop per il 2013 (a) e il 2014 (b).

La produzione è stata stimata anche utilizzando il metodo di assimilazione PSO applicato al modello Aquacrop, per gli stessi anni di riferimento. Le mappe di produzione per i due anni analizzati sono riportate in Fig. 8.3. Tutte le stime ottenute con Aquacrop mostrate in Fig. 8.3 sono state ottenute applicando una calibrazione dei parametri specifica per ogni dataset climatico. Usando il metodo PSO con Aquacrop, la produzione media per unità di superficie risultante per l'area di interesse nel 2013 è di $5.56 \text{ t}\cdot\text{ha}^{-1}$, molto vicino al valore statistico ufficiale, mentre la produzione totale stimata è di $1.10\cdot 10^4 \text{ t}$, presenta quindi un errore di stima di circa $1.18 \cdot 10^4 \text{ t}$. Per il 2014, la produzione stimata utilizzando lo stesso metodo è di $5.97 \text{ t}\cdot\text{ha}^{-1}$, mentre la produzione totale è di circa $0.93\cdot 10^4 \text{ t}$.

La Fig. 8.4 riassume il confronto tra i risultati dell'assimilazione e i dati statistici ufficiali per l'area amministrativa di Yangling.

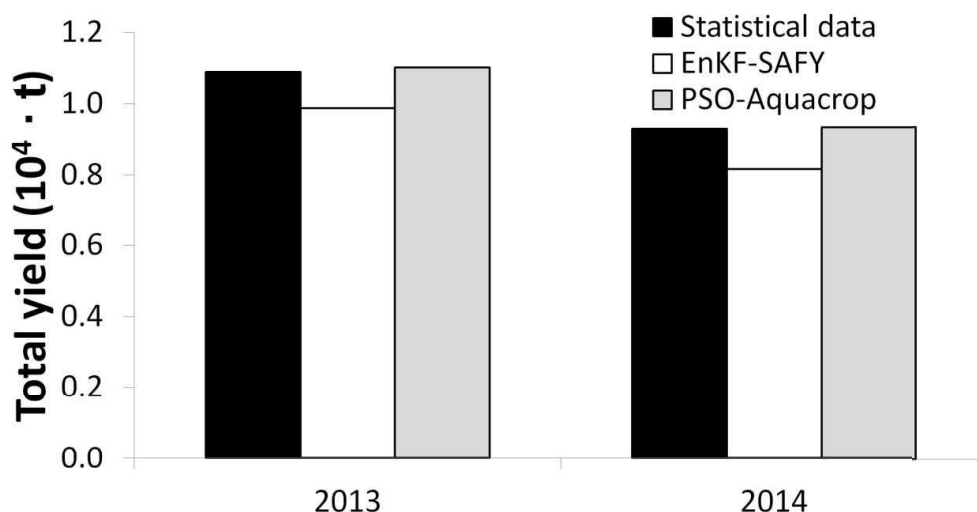


Fig. 8.22: Confronto tra la produzione stimata da: indagini statistiche ufficiali (in nero), metodo di assimilazione EnKF applicato al modello SAFY (in bianco), metodo di assimilazione PSO applicato ad Aquacrop (in grigio).

8.3. *Discussione dei risultati*

L'applicazione dei metodi di assimilazione ha permesso di utilizzare in applicazioni puntuali e spazializzate entrambi i modelli utilizzati ottenendo una stima di produzione di granella con un adeguato livello di accuratezza.

Nell'applicazione puntuale, l'accuratezza del modello SAFY a seguito dell'assimilazione per mezzo del metodo EnKF è risultata maggiore di quella del modello Aquacrop, per il quale è stato invece utilizzato il metodo di assimilazione basato sul PSO. Infatti, il valore di RRMSE tra la produzione di frumento misurata durante le campagne di misura e quella stimata per mezzo di tali procedure è, nel primo caso di circa 0.18 mentre nel secondo è di circa 0.25. Entrambi i metodi sono stati inizialmente applicati utilizzando la stessa calibrazione dei parametri per tutte le simulazioni. Con questa procedura di calibrazione, il metodo EnKF ha dimostrato di riuscire a ricalibrare i parametri principali di SAFY in maniera autonoma, permettendo un'adeguata risposta del modello alle variazioni in input delle condizioni climatiche. Il metodo PSO, invece, ha dimostrato di necessitare di una calibrazione specifica per ogni dataset climatico, poiché la ricalibrazione autonoma dei parametri che maggiormente influenzano la produzione fatta dal metodo di assimilazione non era sufficiente a garantire un'adeguata risposta alle variazioni in input del database climatico.

L'accuratezza nello stimare la produzione del metodo EnKF applicato a SAFY è molto buona per valori bassi di produzione ma diminuisce al crescere di tali valori, fino ad arrivare ad un evidente sottostima della produzione per valori prossimi alle 6 t·ha⁻¹. L'altro metodo di assimilazione usato, invece, sovrastima per valori più bassi di 4 t·ha⁻¹ e sottostima per valori sopra le 6 t·ha⁻¹. Entrambi i metodi mostrano la tendenza a saturare per valori superiori a 6.5 t·ha⁻¹.

Nell'applicazione su scala regionale, entrambi i metodi hanno simulato la produzione totale e la produzione per unità di superficie con buona approssimazione, comparate con i dati statistici ufficiali. I risultati sono abbastanza concordanti con quelli ottenuti con l'applicazione puntuale dei metodi di assimilazione eccetto che per il metodo PSO-Aquacrop per cui l'accuratezza nell'applicazione spazializzata è più alta delle aspettative. Una delle cause è probabilmente da ricercare nella calibrazione dei parametri specifica per ogni dataset climatico utilizzato, aumentando, di fatto, l'accuratezza nella predizione della produzione. A causa dei limiti computazionali di Aquacrop è stato necessario ridurre le classi di LAI in cui suddividere i pixel delle immagini da 1800 (usate per il metodo EnKF-SAFY) a 100. La riduzione così elevata è stata in parte aiutata anche dalla bassa variabilità dei valori che assumeva il CC, specialmente nella fase finale del ciclo colturale. Questa riduzione ha contribuito a ridurre la variabilità dei pixel, tendendo a suddividere molti pixel in poche classi, che avendo subito una calibrazione specifica all'origine, simulano con buona precisione le grandezze biofisiche della coltura (tra cui la produzione di granella). In pratica, pur ottenendo un risultato molto vicino a quello stimato da enti statistici ufficiali, non si ha una reale distribuzione pixel per pixel della produzione, ma piuttosto una distribuzione quasi uniforme di pochi dati di produzioni simulati con un'accuratezza molto alta.

Nonostante l'errore sui dati stimati da RS, trovato per entrambe le variabili di input, ma in particolar modo elevato per il LAI (capitolo 4), le stime della produzione sono state fatte con una buona approssimazione.

Per il metodo di assimilazione EnKF applicato a SAFY, ciò dipende dalle caratteristiche del Filtro di Kalman. Questo metodo, infatti, “filtra” l'errore di misura, introdotto come un effetto casuale, riducendo la sua influenza sull'output. L'effetto dell'errore può essere anche ridotto incrementando il numero di ensemble, ma ciò inevitabilmente aumenterebbe il tempo di calcolo dell'algoritmo. Per questo metodo

quantificare l'errore sulle misure e stabilire un adeguato numero di ensemble è fondamentale (Makowski et al., 2006).

Per il metodo PSO applicato ad Aquacrop, l'errore di input del CC era inferiore che per il LAI utilizzato dal modello SAFY, anche se l'accuratezza dell'output è sembrata essere maggiormente dipendente dalla calibrazione del modello per ogni scenario climatico. Questo metodo è così risultato essere fortemente dipendente dalla calibrazione iniziale dei parametri, la quale risulta piuttosto complessa a causa dell'elevato numero di parametri che descrivono il modello. La dipendenza dalla calibrazione, unita con l'inaccessibilità del codice di Aquacrop nella versione utilizzata e con l'elevato costo computazionale, fanno del metodo di assimilazione PSO applicato al modello Aquacrop una procedura non sfruttabile per applicazioni operative.

Il valore di RRMSE ottenuto con il metodo di assimilazione per updating basato sul EnKF e in questo studio applicato al modello SAFY risulta essere di circa il 18%. In letteratura (Wu et al., 2012; Ines et al., 2013; Huang et al., 2013; Jiang et al., 2014; Dong et al., 2016) il valore di RRMSE trovato per applicazioni di differenti metodi di assimilazione in differenti modelli di crescita delle colture sembra essere tra il 5 e il 16%, pertanto apparentemente migliore di quanto risultato in questo lavoro. Tuttavia è importante considerare che alcuni di questi studi (Wu et al., 2012; Ines et al. 2013; Jiang et al., 2014) hanno utilizzato più di 20 assimilazioni per ogni ciclo colturale, mentre altri studi (Huang et al., 2013; Dong et al. 2016) hanno usato meno immagini degli studi precedenti (circa 10) con un'alta risoluzione e con un errore sulle stime delle variabili biofisiche più basso. Il metodo testato in questo studio, invece, ha utilizzato un numero ridotto di immagini (appena 3) per ogni ciclo colturale con un errore stimato superiore al 20%, raggiungendo livelli di accuratezza nelle stime di produzione comparabili con i livelli raggiunti dagli studi incontrati in letteratura.

9. Conclusioni

L'obiettivo principale di questa tesi è stato di proporre e valutare un approccio innovativo per la stima della produzione di frumento tenero che permettesse un'applicazione spazializzata dei modelli di crescita delle colture, sia su scala di campo che su scala regionale, cercando allo stesso tempo di migliorare l'accuratezza dei modelli utilizzati. Si è pertanto seguito la strategia di integrare i modelli di crescita delle colture con dati acquisiti tramite telerilevamento, utilizzando metodi di assimilazione.

La prima parte di questa tesi è stata dedicata alla stima di LAI e CC, ricavati da tre serie temporali d'immagini (una per ogni ciclo di crescita della vegetazione), utilizzando il metodo PROSAIL-ANN proposto da Bacour et al. (2006). In tale metodo, le simulazioni fatte dal modello di trasferimento radiativo PROSAIL vengono utilizzate per l'addestramento di una rete neurale artificiale (ANN) col fine di mettere in relazione le riflettanze misurate dai sensori a bordo dei satelliti e creare così le mappe di LAI e CC. Il metodo PROSAIL-ANN è stato scelto per ricavare il LAI e il CC sfruttando le stesse simulazioni fatte da PROSAIL, inoltre l'efficacia di tale metodo per sensori multispettrali è stata ampiamente dimostrata in letteratura (e.g. Bacour et al., 2006; Baret et al., 2006; Jacquemoud et al., 2007). La validazione dei valori di CC e LAI stimati è stata fatta per mezzo di misure effettuate su una serie di punti osservati in situ. Da tale validazione ne è conseguito che l'errore stimato per il LAI è maggiore di quello stimato per il CC, in accordo con quanto affermato in letteratura, infatti nel primo caso si ottiene un RRMSE di circa 29% nel secondo di circa il 10%. Anche il valore di RRMSE è in accordo con gli studi sopra citati, infatti in letteratura (Bacour et al., 2006; Baret et al., 2006; Jacquemoud et al., 2007) esso varia tra il 25-39% per il LAI e tra il 10-15% per il CC. Il metodo utilizzato ha permesso di quantificare l'errore sulla stima delle variabili biofisiche, informazione molto utile per l'applicazione del metodo di assimilazione per updating basato su EnKF.

La seconda parte della tesi è stata dedicata all'analisi dei modelli di crescita delle colture. In questo studio sono stati presi in esame due differenti modelli largamente descritti nei capitoli precedenti, Aquacrop e SAFY. I due modelli sono stati analizzati per mezzo di uno studio di analisi di sensibilità (Sensitivity Analysis), il quale ha permesso di individuare i parametri che più influenzano la stima di produzione, alle

condizioni climatiche specifiche di questo studio. Inoltre la SA ha consentito di effettuare un confronto tra i due modelli, seppur qualitativo visto il basso numero di scenari utilizzati. In particolare sono state quantificate e messe a confronto la plasticità e la complessità dei modelli, risultando Aquacrop più complesso di SAFY e quest'ultimo con una plasticità maggiore del primo.

Dal confronto fatto tra i due modelli per mezzo della SA, si evince una migliore predisposizione di SAFY a interagire con altri algoritmi, rendendolo un miglior candidato per interazioni con metodi di assimilazione. I risultati ottenuti dalla SA non sono stati però tanto evidenti da consentire di scartare il modello Aquacrop per applicazioni spazializzate per mezzo di metodi di assimilazione, perché i risultati di plasticità e complessità non hanno indicato che Aquacrop sia da considerare in assoluto un modello complesso e con bassa plasticità, ma semplicemente più complesso e con plasticità minore rispetto a SAFY. Da letteratura (Confalonieri et al., 2012) risulta infatti che a confronto con altri modelli di crescita delle colture Aquacrop non è considerabile né tra i più complessi né tra quelli con plasticità più bassa. Numerosi studi (e.g. Mkhabela et al., 2012; Ghanbbari et al., 2013; Jin et al., 2014) hanno dimostrato elevate prestazioni del modello nello stimare la produzione di frumento tenero a seguito di un'accurata calibrazione dei parametri, pertanto il confronto effettuato tra i due modelli deve considerarsi valido esclusivamente per gli scenari considerati, in quanto sia Aquacrop che SAFY risultano avere una bassa complessità ed un'alta plasticità se comparati con i modelli di crescita delle colture maggiormente incontrati in letteratura (Confalonieri et al., 2012).

La SA effettuata ha messo in risalto la forte dipendenza tra la sensibilità dei singoli parametri (e delle loro interazioni) sull'output del modello e le condizioni al contorno (climatiche e ambientali) in cui l'analisi è svolta. Ciò porta ad affermare che è opportuno effettuare uno studio di sensibilità per ogni scenario che si intende utilizzare, in particolare se l'intento è di applicare un metodo di assimilazione al modello, in quanto al variare delle condizioni al contorno e dei relativi intervalli di variazione dei parametri (che ne descrivono alcune caratteristiche), può variare in modo significativo l'influenza di un parametro sul output. Essendo limitato il numero di parametri da inserire nell'assimilazione, non effettuare una SA per il sito specifico potrebbe portare ad una scelta equivocata dei parametri da ricalibrare con l'assimilazione.

Nella terza parte della tesi sono stati testati 2 differenti metodi di assimilazione di dati ottenuti da RS nei modelli di crescita delle colture, un metodo di assimilazione per updating basato su EnKF, il quale assimila il LAI stimato nel modello SAFY, e un altro basato su un metodo di assimilazione per calibrazione, il PSO, il quale assimila il CC stimato nel modello Aquacrop. La scelta di applicare due differenti metodi di assimilazione per i due modelli scelti è dovuta all'inaccessibilità del codice di Aquacrop, condizione che ha reso impossibile l'applicazione di un metodo di assimilazione per updating. Ciò ha tuttavia consentito di analizzare le differenze tra i due differenti metodi di assimilazione, uno per calibrazione l'altro per updating. Questi metodi sono stati applicati per uguali condizioni, come ad esempio: uguale numero di assimilazioni, stesso scenario, uguali condizioni dei dati RS, pertanto stessi errori nella stima delle variabili biofisiche ecc. Entrambi i metodi sono stati validati sia con dati di produzione raccolti in situ su scala di campo, sia con dati statistici ufficiali su scala regionale (dove per regione si intende in questo caso l'area amministrativa di Yangling).

Sono risultati di particolare interesse i test effettuati su scala di campo, i quali hanno mostrato la possibilità di impiegare immagini con una risoluzione di 30 m su una scala di misura dove generalmente si prediligono risoluzioni maggiori.

I risultati di stima di produzione ottenuti con Aquacrop su scala di campo sono meno performanti se confrontati con quelli ottenuti con altri metodi di calibrazione, incontrati in letteratura (Doraiswamy et al., 2003; Launay et al., 2005; De Wit et al., 2012). Va specificato però che in letteratura i modelli accoppiati ai metodi di assimilazione per calibrazione sono differenti da Aquacrop. La minor precisione di stima di produzione su scala di campo è dovuta principalmente alle caratteristiche di Aquacrop, per il quale è necessario effettuare una calibrazione molto accurata dei parametri, la quale non riesce ad essere soddisfatta in modo adeguato dal metodo di assimilazione PSO. Tale inadeguatezza è dovuta al fatto che il metodo PSO richiede una potenza di calcolo (in termini di velocità del processore) molto elevata, pertanto per consentire la sua applicazione con gli strumenti di calcolo a disposizione, è stato necessario limitare la velocità delle particelle e il numero di iterazioni al valore minimo consigliato in letteratura (Li et al., 2015). Tali limitazioni hanno pregiudicato le performance del metodo PSO, il quale non è riuscito a ricalibrare in modo ottimale i parametri del modello Aquacrop. Per migliorare i risultati, infatti, è stato necessario dover ricalibrare i parametri in modo specifico per ogni dataset climatico utilizzato.

In apparente contraddizione con i risultati a scala di campo, l'applicazione a scala regionale del metodo di assimilazione PSO con Aquacrop è quella che meglio corrisponde ai dati regionali di produzione forniti dagli enti ufficiali di statistica. Ciò è dovuto in parte alla ricalibrazione specifica effettuata per ogni dataset climatico, la quale ha forzatamente migliorato la stima di produzione, in parte invece alla limitazione delle classi di serie temporale di CC in cui sono stati suddivisi i pixel per venire incontro alle esigenze computazionali del metodo. In questo modo la variazione di produzione stimata per ogni simulazione è stata gravemente limitata, comportando un risultato molto accurato ma non altrettanto affidabile. L'elevato costo computazionale, la complessità della calibrazione e la necessità di ricalibrare il modello per ogni dataset climatico rendono il metodo PSO applicato ad Aquacrop poco performante.

Nel complesso il modello Aquacrop, risulta avere prestazioni potenzialmente buone, a patto di riuscire a effettuare la sua calibrazione in modo accurato. Lo studio effettuato in questa tesi ha mostrato, infatti, la stretta dipendenza tra le prestazioni nella stima della produzione e la calibrazione adeguata dei parametri. La scarsa affidabilità dei risultati ottenuti è da attribuire per lo più alle numerose limitazioni imposte dall'accoppiamento di un modello e di un metodo, entrambi dalle elevate esigenze computazionali.

Tra i modelli esistenti, Aquacrop è certamente quello che meglio descrive il comportamento delle colture in funzione del bilancio idrico, sarebbe quindi di particolare interesse poterlo impiegare per studiare zone particolarmente soggette a siccità, o per meglio osservare i legami tra la crescita e lo sviluppo delle colture e la quantità di acqua presente nel suolo. Come detto in precedenza, l'inaccessibilità del codice di Aquacrop limita la sua applicazione ai soli metodi di assimilazione per calibrazione. Le sperimentazioni fatte in questa tesi hanno dimostrato che qualsiasi applicazione del modello Aquacrop in simbiosi con altri metodi e algoritmi risulta macchinosa e complessa fintanto che il codice del modello non sarà modificabile. Recentemente è stata sviluppata una versione open source di Aquacrop, che rende pienamente accessibile il codice sorgente di questo modello (Foster et al., 2017). Sarebbe pertanto interessante valutare le prestazioni di questo modello con un metodo di assimilazione computazionalmente sostenibile, come ad esempio il metodo basato su EnKF sviluppato in questo lavoro.

Il metodo di assimilazione per aggiornamento basato sul EnKF sviluppato e testato in questo studio sul modello Simple Algorithm For Yield (SAFY), ha mostrato buone prestazioni. I risultati ottenuti sono perfettamente in linea con quelli di altri metodi di assimilazione per updating incontrati in letteratura (Wu et al, 2012; Huang et al., 2013; Ines et al., 2013; Jiang et al., 2014; Dong et al., 2016).

L'innovazione di questo metodo non è stata nell'apportare un sostanziale aumento nell'accuratezza nelle stime di produzione rispetto ai metodi esistenti, ma nel dimostrare la possibilità di utilizzare un ridotto numero di immagini acquisite tramite Remote Sensing, con un errore in ingresso di stime di LAI superiore al 25% e con un errore in uscita nella stima di produzione di circa il 18%. Tale caratteristica amplia le possibilità di applicazione delle metodologie che accoppiano dati RS e stime di produzione per mezzo di modelli di crescita delle colture anche a zone del mondo in cui, a causa delle condizioni climatiche o economiche, non è possibile reperire a basso costo, per ogni ciclo di crescita della coltura, una numerosa serie temporale di dati ad alta qualità. I metodi di assimilazione per updating incontrati in letteratura fanno uso di un numero compreso tra le 8 e le 24 assimilazioni (alle quali corrispondono altrettante immagini acquisite tramite sensori multispettrali), per mezzo del metodo EnKF-SAFY sviluppato in questo lavoro è stato possibile ridurre il numero di assimilazioni a 3. La metodologia sviluppata necessita di ulteriori studi per poter essere affinata e generalizzata a qualsiasi tipo di scenario. In particolare è necessario sperimentarla su un maggior numero di scenari con sostanziali differenze climatiche e ambientali, possibilmente ottenendo un maggior numero di osservazioni in campo, sia per numero di punti che per frequenza nelle osservazioni. Di particolare importanza sarebbe ottenere delle misure in campo con un'accuratezza nota, poiché sia i dati sperimentali di campo che i dati delle statistiche ufficiali utilizzati in questo studio sono affetti da un errore difficile da quantificare.

Per uno studio futuro sarebbe interessante sperimentare questa metodologia utilizzando immagini acquisite da satelliti in aree meno affette da perturbazioni atmosferiche e con una risoluzione maggiore, sfruttando ad esempio le immagini del satellite Sentinel 2 recentemente messe a disposizione gratuitamente dal European Space Agency, le quali garantiscono una risoluzione di 10 m al suolo e una frequenza di passaggio a latitudine medie di 2-3 giorni.

Dagli studi incontrati in letteratura si evince che il vantaggio principale dell'uso di metodi di assimilazione per calibrazione consiste nel poter applicare questo metodo senza necessità di modifiche al codice del modello e con un ridotto numero di immagini, generalmente compreso tra 3 e 6 (Doraiswamy et al., 2003; Launay et Guerif, 2005; De Wit et al, 2012) ottenendo errori di stima della produzione compresi tra il 10% e il 25%. L'applicazione del metodo PSO-Aquacrop sviluppato per questa tesi va a confermare questa stima, mettendo però in risalto la complessità computazionale di questa metodologia. Si è, infatti, dimostrato che l'esigenza computazionale di tali metodi ne limita molto le performance, specie se accoppiata con modelli che richiedono anch'essi elevati costi computazionali.

Questo lavoro ha messo in evidenza che i metodi di assimilazione per updating risultano essere computazionalmente più accessibili dei precedenti, essi infatti non richiedono specifiche tecniche dei calcolatori particolarmente elevate né tantomeno elevati tempi di attesa per il calcolo. Tale caratteristica è fondamentale in applicazioni spazializzate ad elevata risoluzione.

In conclusione, si può pertanto affermare che per meglio descrivere il comportamento degli agroecosistemi, in termini di produzione delle colture, è preferibile l'impiego di metodi di assimilazioni di dati RS a modelli di crescita delle colture per mezzo di metodi di assimilazione updating, piuttosto che per mezzo di metodi per calibrazione, in quanto i primi risultano, a parità di immagini assimilate, più semplici da utilizzare, presentano meno limitazioni e permettono stime di produzione con buone approssimazioni. Inoltre tali metodi sono quelli che presumibilmente permettono un margine di miglioramento più ampio, utilizzando, ad esempio, modelli di crescita delle colture con un'accuratezza maggiore, aumentando il numero di assimilazioni o usando variabili di input ottenuti da RS con risoluzione e qualità maggiore, miglioramenti non possibili con i metodi di assimilazione per calibrazione a causa delle limitazioni imposte dagli elevati costi di computazione.

Ringraziamenti

Questo lavoro è stato supportato dalla borsa di studio cofinanziata dall'Università degli studi della Tuscia e dal “Dragon Programme training for young scientist scheme” promosso dalla European Space Agency (numero contratto 4000110311/14/I-BG) all'interno del progetto Dragon 3 “Farmaland Drought” (project ID 10448). Ringrazio i colleghi del National Engineering Research Center for Information Technology in Agriculture per la loro collaborazione, per avermi ospitato nelle loro strutture di Beijing nel 2014 e nel 2016 per un totale di 5 mesi, per aver fornito i dati delle campagne di misura effettuate sui siti sperimentali e i dati satellitari raccolti per mezzo delle strumentazioni cinesi. Ringrazio anche Dirk Raes per aver fornito la versione “command-line” del modello Aquacrop e Samuel Buis e gli altri colleghi del INRA Avignon (France) per aver fornito il codice Matlab del modello SAFY e della Global Sensitivity Analysis. Infine un particolare ringraziamento al mio tutore, il Prof. Casa e al cotutore, il Dr. Pignatti, per il supporto e la pazienza dimostrata in questi tre anni di dottorato.

Bibliografia

1. Allen RG, Pereira LS, Raes D, Smith M, W AB. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. Irrig Drain. 1998;1–15.
2. Atzberger C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sens Environ.* 2004;93(1):53–67.
3. Atzberger C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* 2013;5(2):949–81.
4. Bacour C. Design and analysis of numerical experiments to compare four canopy reflectance models. *Remote Sens Environ.* 2002;79(1):72–83.
5. Bacour C, Baret F, Béal D, Weiss M, Pavageau K. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens Environ.* 2006;105(4):313–25.
6. Bannari A, Morin D, Bonn F, Huete AR. A review of vegetation indices. *Remote Sens Rev [Internet]*. 1995;13(1):95–120.
7. Baret F, Guyot G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens Environ.* 1991;35(2):161–73.
8. Baret F, Buis S. Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems. In: Liang S, editor. *Advances in Land Remote Sensing*. Springer Netherlands; 2008. p. 173–201.
9. Baret F, Hagolle O, Geiger B, Bicheron P, Miras B, Huc M, et al. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens Environ.* 2007;110(3):275–86.
10. Baret F, Pavageau K, Béal D, Weiss M, Berthelot B, Regner P. Algorithm Theoretical Basis Document for MERIS Top of Atmosphere Land Products (TOA _ VEG). Contract ESA AO/1-4233/02/I-LG, INRA Nov. 2006;(March):1–37.
11. Basso B, Cammarano D, De Vita P. Remotely sensed vegetation indices: Theory and applications for crop management. *Ital J Agrometeorol.* 2004;53(1):36–53.
12. Beale MH, Hagan MT, Demuth HB. *Neural Network Toolbox™ User ' s Guide R 2013 b*. 2013;
13. Boegh E, Houborg R, Bienkowski J, Braban CF, Dalgaard T, Van Dijk N, et al. Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop- and

- grasslands in five European landscapes. *Biogeosciences*. 2013;10(10):6279–307.
14. Bristow KL, Campbell GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric For Meteorol* [Internet]. 1984;31(2):159–66.
 15. Burgers G, van Leeuwen PJ, Evensen G. Analysis Scheme in the Ensemble Kalman Filter. *Mon Weather Rev*. 1998;126:1719–24.
 16. Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. *Environ Model Softw*. 2007 ;22(10):1509–18.
 17. Cariboni J, Gatelli D, Liska R, Saltelli a. The role of sensitivity analysis in ecological modelling. *Ecol Modell*. 2007;203(1–2):167–82.
 18. Chen, J. M.; Black, T. A. Defining leaf area index for non-flat leaves. *Plant, Cell Environ*. 1992, 15, 421–429. 19. Chen Y, Cournède PH. Data assimilation to reduce uncertainty of crop model prediction with Convolution Particle Filtering. *Ecol Modell*. 2014;290:165–77.
 19. Coleman DC, Hendrix PF. Agroecosystem processes. In: Pomeroy LR, Albers JJ, editors. *Concepts of ecosystem ecology*. New York: Ecological Studies; 1988. p. 149–70.
 20. Combal B, Baret F, Weiss M, Trubuil A, Macé D, Pragnère A, et al. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sens Environ*. 2003;84(1):1–15.
 21. Confalonieri R, Bellocchi G, Bregaglio S, Donatelli M, Acutis M. Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecol Modell*. 2010a;221(16):1897–906.
 22. Confalonieri R, Bregaglio S, Acutis M. Quantifying plasticity in simulation models. *Ecol Modell*. 2012;225:159–66.
 23. Confalonieri R, Bregaglio S, Acutis M. Short communication A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. *Ecol Modell*. 2010b;221(6):960–4.
 24. Confalonieri R. Monte Carlo based sensitivity analysis of two crop simulators and considerations on model balance. *Eur J Agron*. 2010;33(2):89–93.
 25. Confalonieri R, Acutis M, Bellocchi G, Donatelli M. Multi-metric evaluation of the models WARM , CropSyst , and WOFOST for rice. 2009;220:1395–410.
 26. Cukier RI, Levine HB, Shuler KE. Nonlinear sensitivity analysis of multiparameter model systems. *J Comput Phys*. 1978;26(1):1–42.

27. Dantzig GB. Maximization of a linear function of variables subject to linear inequalities. In: Koopmans TC, editor. *Activity Analysis of Production and Allocation*. New York-London: Wiley & Chapman-Hall; 1947. p. 339–47.
28. De Wit AJW, Boogaard HL, Van Diepen CA. Spatial resolution of precipitation and radiation: the effect on regional crop yield forecasts. *Agric For Meteorol*. 2005;135(1):156–68.
29. De Wit AJW, Duveiller G, Defourny P. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agric For Meteorol*. 2012;164:39–52.
30. Debes K, Koenig A, Gross H-M. Transfer Functions in Artificial Neural Networks A Simulation-Based Tutorial. *Brains, minds and media. Journal of New Media in Neural Cognitive Science and Education*. 2005.
31. Delécolle R, Maas SJ, Guérif M, Baret F. Remote sensing and crop production models: present trends. *ISPRS J Photogramm Remote Sens*. 1992;47(2–3):145–61.
32. Dente L., Satalino G., Mattia F., Rinaldi M. Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. *Remote Sensing of Environment* 2008; 112, 1395-1407.
33. Dong T, Liu J, Qian B, Zhao T, Jing Q, Geng X, et al. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *Int J Appl Earth Obs Geoinf*. 2016;49:63–74.
34. Doorenbos J, Kassam AH. Yield response to water. *Irrig Drain Pap n 33 FAO*. 1979;193.
35. Doraiswamy PC, Moulin S, Cook PW, Stern A. Crop Yield Assessment from Remote Sensing. *Photogramm Eng Remote Sensing*. 2003;69(6):665–74.
36. Dorigo W a., Zurita-Milla R, de Wit a. JW, Brazile J, Singh R, Schaepman ME. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int J Appl Earth Obs Geoinf*. 2007;9(2):165–93.
37. Duchemin B, Maisongrande P, Boulet G, Benhadj I. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environ Model Softw*. 2008;23(7):876–92.
38. Eberhart R, Kennedy J. A New Optimizer Using Particle Swarm Theory. *Micro Mach Hum Sci 1995 MHS '95, Proc Sixth Int Symp*. 1995;39–43.
39. Elliott ET, Cole CV. A Perspective on Agroecosystem Science. *Ecology*. 1989;70(6):1597–602.

40. Fang H, Liang S, Kuusk A. Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. *Remote Sens Environ.* 2003;85(3):257–70.
41. Foster T, Brozović N, Butler AP, Neale CMU, Raes D, Steduto P, Fereres E, Hsiao TC. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agric Water Manag.* 2017; 181:8-22
42. Gardner BL. Causes of rural economic development. *Agric Econ.* 2005;32(1):21–41.
43. Ghanbbari A, Tavassoli A. Simulation of Wheat Yield Using AquaCrop Model in Shirvan Region. *IjagcsCom.* 2013;342–52.
44. Glenn EP, Huete AR, Nagler PL, Nelson SG. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors.* 2008;8(4):2136–60.
45. Grubbs FE. Procedures for Detecting Outlying Observations in Samples. *Technometrics.* 1969;11(1):1–21.
46. Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: Mira J, Sandoval F, editors. *From Natural to Artificial Neural Computation: International Workshop on Artificial Neural Networks Malaga-Torremolinos, Spain, June 7--9, 1995 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1995. p. 195–201.
47. Huang Y, Zhu Y, Li W, Cao W, Tian Y. Assimilating Remotely Sensed Information with the WheatGrow Model Based on the Ensemble Square Root Filter for Improving Regional Wheat Yield Forecasts. *Plant Prod Sci.* 2013;16(4):352–64.
48. Huang Z, Turner BJ, Dury SJ, Wallis IR, Foley WJ. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens Environ.* 2004;93(1–2):18–29.
49. Iman RL, Conover WJ, Conover WJ. American Society for Quality A Measure of Top-Down Correlation Society for Quality Stable URL: <http://www.jstor.org/stable/1269344>. 1987;29(3):351–7.
50. Ines AVM, Das NN, Hansen JW, Njoku EG. Remote Sensing of Environment Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens Environ.* 2013;138:149–64. . 57. Jacquemoud S, Verdebout J, Schmuck G, Andreoli G, Hosgood B. Investigation of leaf biochemistry by statistics. *Remote Sens Environ.* 1995;54(3):189–97.
51. Jacquemoud S, Baret F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens Environ.* 1990;34:75–91.

52. Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada PJ, Asner GP, et al. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens Environ.* 2009;113:S56–66.
53. Jiang Z, Chen Z, Chen J, Liu J, Ren J, Li Z, et al. Application of Crop Model Data Assimilation With a Particle Filter for Estimating Regional Winter Wheat Yields. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2014; 7(11):4422–31.
54. Jin X, Feng H, Zhu X, Li Z, Song S, Song X, et al. Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PLoS One.* 2014 Jan;9(1):e86938.
55. Kalpana R, Natarajan S, Mythili S. Remote sensing for crop monitoring - a review. *Agric Rev.* 2003;24(1):31–9.
56. Kang Y, Özdoğan M, Zipper S, Román M, Walker J, Hong S, et al. How Universal Is the Relationship between Remotely Sensed Vegetation Indices and Crop Leaf Area Index? A Global Assessment. *Remote Sens.* 2016;8(7):597.
57. Koda M, Mcrae GJ, Seinfeld JH. Automatic sensitivity analysis of kinetic mechanisms. *Int J Chem Kinet.* 1979 Apr;11(4):427–44.
58. Kötz B, Schaepman M, Morsdorf F, Bowyer P, Itten K, Allgöwer B. Radiative transfer modeling within a heterogeneous canopy for estimation of forest fire fuel properties. *Remote Sens Environ.* 2004;92(3):332–44.
59. Launay M, Guerif M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric Ecosyst Environ.* 2005;111(1–4):321–39.
60. Li Y, Zhou Q, Zhou J, Zhang G, Chen C, Wang J. Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecol Modell.* 2014;291:15–27.
61. Li Z, Jin X, Zhao C, Wang J, Xu X, Yang G, et al. Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *Eur J Agron.* 2015;71:53–62.
62. Liang S. Estimation of Land Surface Biophysical Variables. In: Kong GA, editor. *Quantitative Remote Sensing of Land Surfaces.* New York: Wiley Series in Remote Sensing. JohnWiley & Sons; 2003. p. 246–309.
63. Lobell DB. The use of satellite data for crop yield gap analysis. *F Crop Res [Internet].* 2013;143:56–64. Available from:
64. Lu Y, Mohanty S. Sensitivity analysis of a complex, proposed geologic waste disposal system using the Fourier Amplitude Sensitivity Test method. *Reliab Eng Syst Saf.* 2001;72(3):275–91.

65. Ma G, Huang J, Wu W, Fan J, Zou J, Wu S. Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. *Math Comput Model.* 2011;58(3–4):634–43.
66. Maas SJ. Parameterized model of gramineous crop growth. I. Leaf area and dry mass simulation. *Agron J.* 1993;85:348–353;
67. Makowski D, Guérif M, Jones JW, Graham W. Data assimilation with crop models. In: Wallach D, Makowski D, Jones JW, editors. *Working With Dynamic Crop Models*. 1st ed. ELSEVIER; 2006. p. 151–70.
68. Mayer DG, Butler DG. Statistical validation. *Ecol Modell.* 1993;68(1):21–32.
69. Mesplé F, Troussellier M, Casellas C, Legendre P. Evaluation of simple statistical criteria to qualify a simulation. *Ecol Modell.* 1996;88(1–3):9–18.
70. Mkhabela MS, Bullock PR. Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. *Agric Water Manag.* 2012;110:16–24.
71. Monod H, Naud C, Makowski D. Uncertainty and sensitivity analysis for crop models. In: Wallach D, Makowski D, Jones JW, editors. *Working With Dynamic Crop Models*. 1st ed. 2006. p. 55–99.
72. Monteith JL, Moss CJ. Climate and the Efficiency of Crop Production in Britain [and Discussion]. Vol. 281, *Philosophical Transactions of the Royal Society B: Biological Sciences*. 1977. p. 277–94.
73. Morris MD. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* . 1991;33(2):161–74.
74. Mulla DJ. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst Eng.* 2013;114(4):358–71.
75. Nielsen DC, Miceli-Garcia JJ, Lyon DJ. Canopy Cover and Leaf Area Index Relationships for Wheat, Triticale, and Corn. *Agron J.* 2012;104(6):1569.
76. Passioura JB. Simulation models: science, snake oil, education or engineering. *Agron J.* 1996;88:690–4.
77. Pianosi F, Wagener T. A Simple and Effective Approach to Global Sensitivity Analysis Based on Conditional Output Distributions. 2014;
78. Pujol G. Simplex-based screening designs for estimating metamodels. *Reliab Eng Syst Saf.* 2009;94(7):1156–60.
79. Qiu J, Gao W, Lesht BM. Inverting optical reflectance to estimate surface properties of vegetation canopies. *Int J Remote Sens.* 1998;19(4).

80. Raes D, Steduto P, Hsiao TC, Fereres E. Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description. *Agron J*. 2009;101(3):438–47.
81. Raes D, Steduto P, Hsiao TC, Fereres E. Et0 Calculator, Reference Manual. 2012.
82. Rauff KO, Bello R. A Review of Crop Growth Simulation Models as Tools for Agricultural Meteorology. *Agric Sci*. 2015;(September):1098–105.
83. Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA. Uncertainty in the environmental modelling process - A framework and guidance. *Environ Model Softw*. 2007;22(11):1543–56.
84. Rinaldi M., Ruggieri S., Garofalo P., Vonella A.V., Satalino G., Soldo P. Leaf Area Index Retrieval Using High Resolution Remote Sensing Data. *Italian Journal of Agronomy* 2010, 5, 2:155-166.94. Roerink GJ, Menenti M, Verhoef W. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *Int J Remote Sens*. 2010;21(9):1911–7.
85. Rouse JW, Hass RH, Schell JA, Deering DW. Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resour Technol Satell Symp*. 1973;1:309–17.
86. Rykiel EJ. Testing ecological models: The meaning of validation. *Ecol Modell*. 1996;90(3):229–44.
87. Saltelli a., Tarantola S, Campolongo F. Sensitivity analysis as an ingredient of modeling. *Stat Sci*. 2000;15(4):377–95.
88. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis: The Primer*. 2008. 66-69 p.
89. Saltelli A, Tarantola S, Chan K. Global Sensitivity Analysis of Model Output A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*. 1999;41(1):39–56.
90. Saltelli A, Tarantola S, Campolongo F, Ratto M. *Sensitivity Analysis in Practice*. 2004. 219 p.
91. Sathyanarayana S. A Gentle Introduction to Backpropagation. http://numericinsight.com/uploads/A_Gentle_Introduction_to_Backpropagation.pdf. 2014.
92. Shin M, Guillaume JHA, Croke BFW, Jakeman AJ. Addressing ten questions about conceptual rainfall – runoff models with global sensitivity analyses in R. *J Hydrol*. 2013;503:135–52.
93. Sobol IM. Asymmetric convergence of approximations of the Monte Carlo method. *Comput Math Phys* . 1993;33(10):1391–6.

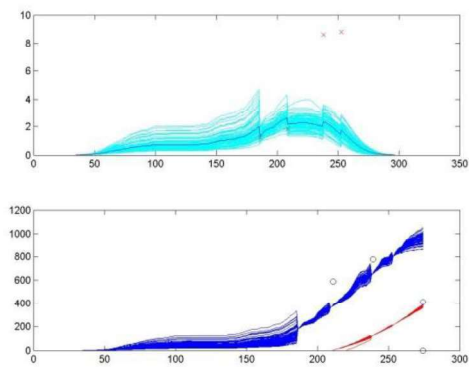
94. Steduto P, Raes D, Hsiao TC, Fereres E, Heng L, Izzi G, et al. AquaCrop : a new model for crop prediction under water deficit conditions. *Options Méditerranéennes*. 2008;Series A(80):285–92.
95. Steduto P, Hsiao TC, Raes D, Fereres E. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agron J*. 2009;101(3):426.
96. Steduto P, Hsiao TC, Fereres E. On The Conservative Behavior Of Biomass Water Productivity. *Irrig Sci*. 2007;25(3):189–207.
97. Tilman D. The Ecological Consequences Of Changes In Biodiversity : Perspectives. *Ecology*. 1999;80(5):1455–74.
98. van Diepen, C. A.; Wolf, J.; van Keulen, H.; Rappoldt, C. WOFOST: a simulation model of crop production. *Soil Use Manag*. **1989**, 5, 16–24.
99. Vanuytrecht E, Raes D, Willems P. Global sensitivity analysis of yield output from the water productivity model. *Environ Model Softw*. 2014;51:323–32.
100. Vazquez-Cruz MA, Guzman-Cruz R, Lopez-Cruz IL, Cornejo-Perez O, Torres-Pacheco I, Guevara-Gonzalez RG. Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. *Comput Electron Agric* . 2014 Jan;100:1–12.
101. Veloso, A. G. M. Modélisation spatialisée de la production, des flux et des bilans de carbone et d'eau des cultures de blé à l'aide de données de télédétection : application au sud-ouest de la France, Université Toulouse III Paul Sabatier (UT3 Paul Sabatier), 2014.
102. Verger A, Baret F, Camacho F. Optimal modalities for radiative transfer- neural network estimation of canopy biophysical characteristics : Evaluation over an agricultural area with CHRIS / PROBA observations. 2011;115:415–26.
103. Verhoef W. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens Environ*. 1984;16(2):125–41.
104. Version E. ENVI User ' s Guide. 2004;1–1150.
105. Viña A, Gitelson AA, Nguy-Robertson AL, Peng Y. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens Environ* . 2011;115(12):3468–78.
106. Wallach D. The Two Forms of Crop Models. In: Wallach D, Makowski D, Jones JW, editors. *Working With Dynamic Crop Models*. Elsevier; 2006(a). p. 3–10.

107. Wallach D. Evaluating Crop Models. In: Wallach D, Makowski D, Jones JW, editors. Working With Dynamic Crop Models. First Edit. ELSEVIER; 2006(b). p. 11–53.
108. Wu S, Huang J, Liu X, Fan J, Ma G, Zou J. Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield. IFIP Adv Inf Commun Technol. 2012;370 AICT(PART 3):410–8.
109. Zhao K, Xu J, Zhao Z, Song L, Xiao K. Cross Comparison of HJ-1A/B CCD and Landsat TM/ETM+ Multispectral Measurements for NDVI, SAVI and EVI Vegetation Index. Remote Sens Technol Appl. 2013;28(4):674–80.
110. U.S Department of the Interior | U.S Geological Survey; 2016 http://landsat.usgs.gov/about_ldcm.php.

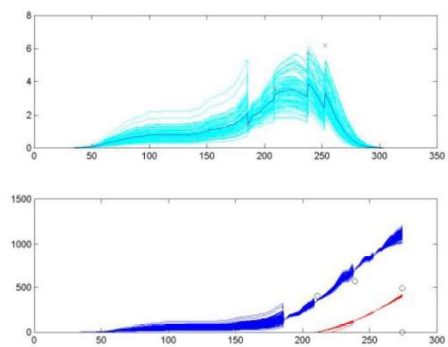
Allegato A

Di seguito sono mostrati gli andamenti delle variabili biofisiche simulate dal modello SAFY a seguito dell'applicazione del metodo di assimilazione basato sul EnKF fatta su Yangling per il dataset dei tre cicli colturali osservati: 2012-2013, 2013-2014, 2014-2015. In ogni grafico in alto è mostrato l'andamento del LAI per ogni combinazione dei parametri dell'ensemble (linee in azzurro), valore medio tra tutti gli andamenti di LAI dell'ensemble (linea blu), valore puntuale di LAI stimato da RS e utilizzato per l'assimilazione (x in rosso), mentre nel grafico in basso è mostrato l'andamento della biomassa misurata in g/m^2 per ogni combinazione dei parametri dell'ensemble (linee in blu), andamento della produzione di granella misurato in g/m^2 per ogni combinazione dei parametri dell'ensemble (linee in rosso), valori puntuali di biomassa misurati in campo (cerchi in nero), valore finale di produzione di granella (cerchio in blu). Per ogni grafico in didascalia è indicata la denominazione del punto.

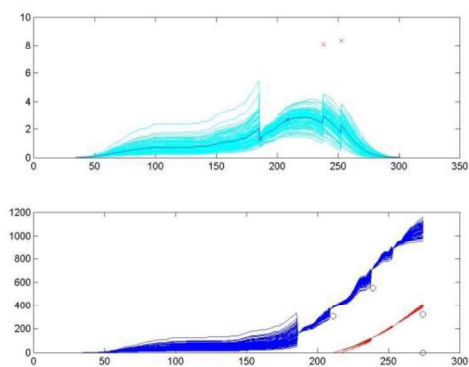
Ciclo Culturale 2012-2013



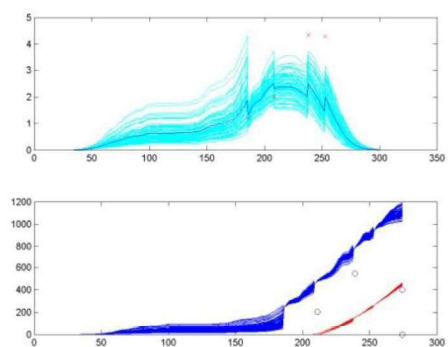
Img 1: FF08



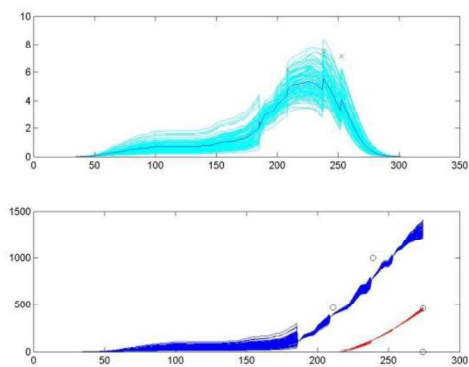
Img 2: FF09



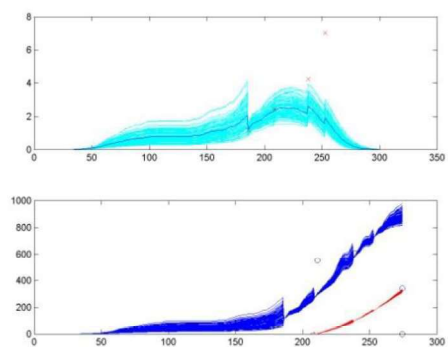
Img 3: FF15



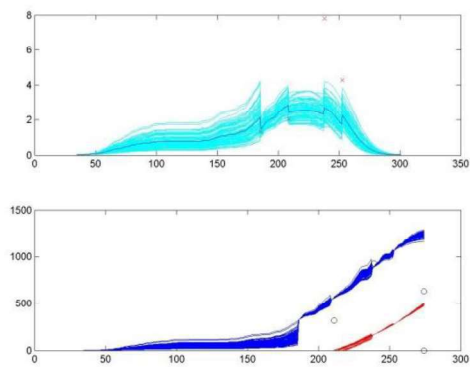
Img 4: FF16



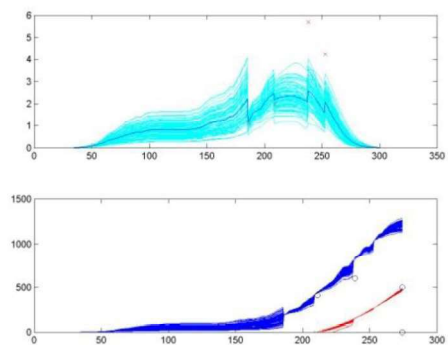
Img 5: WG27



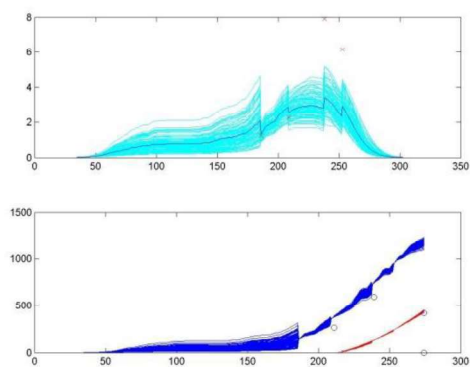
Img 6: WG28



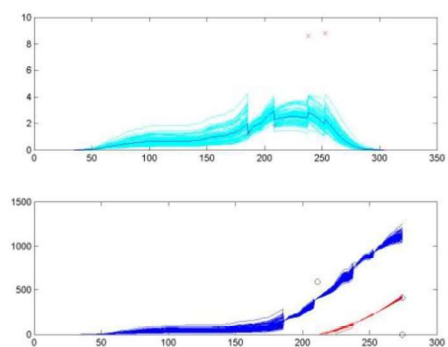
Img 7: WG29



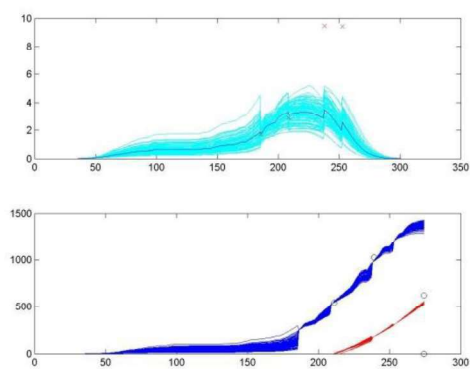
Img 8: WG30



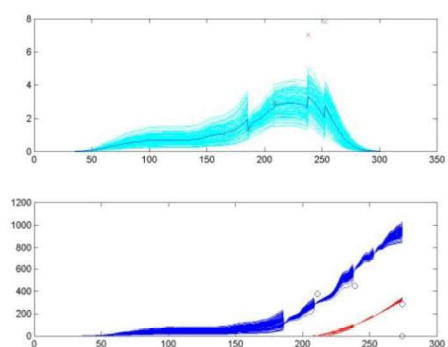
Img 9: WG31



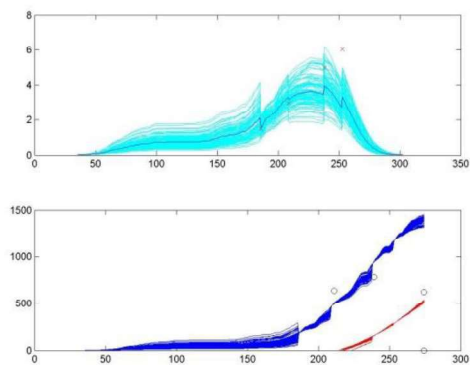
Img 10: WG32



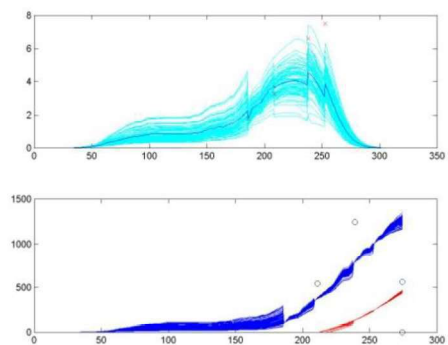
Img 11: WG33



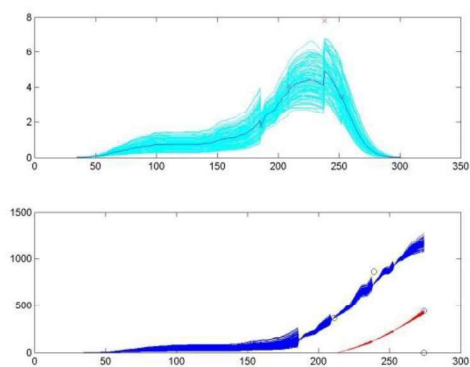
Img 12: WG 35



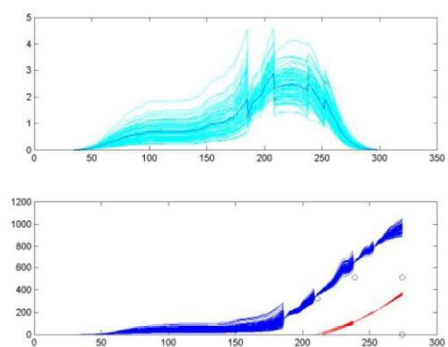
Img 13: WG37



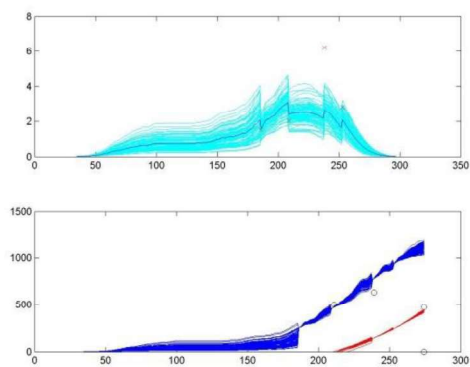
Img 14: WG39



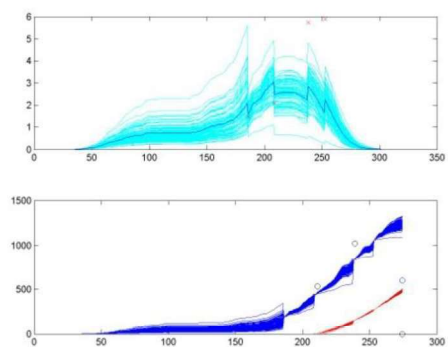
Img 15: WG40



Img 16: WG43

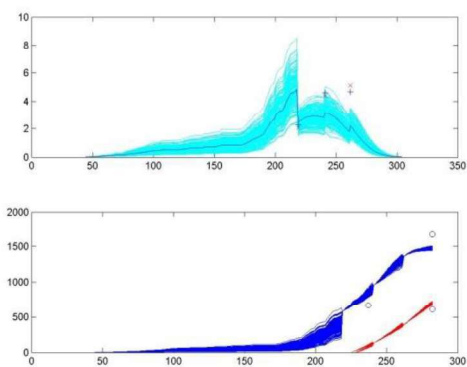


Img 17: YL17

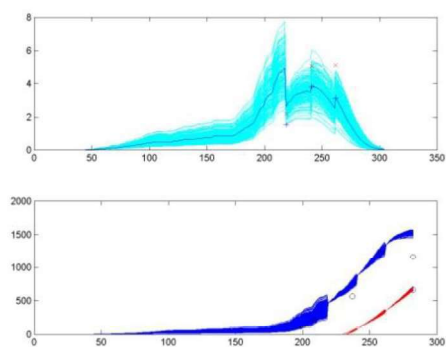


Img 18: YL26

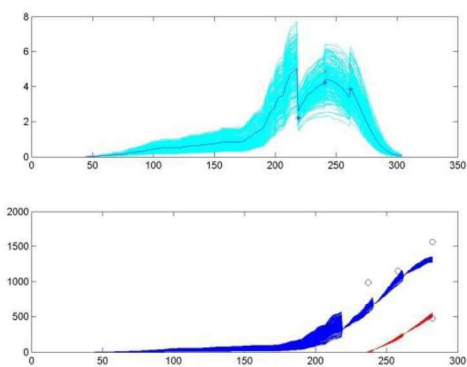
Ciclo Culturale 2013-2014



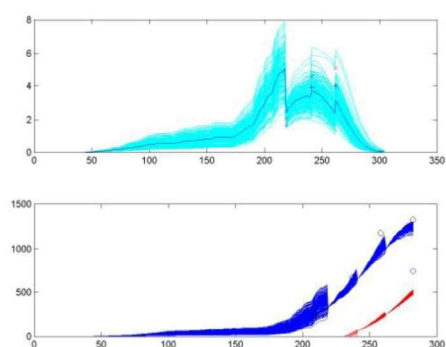
Img 19: RG01



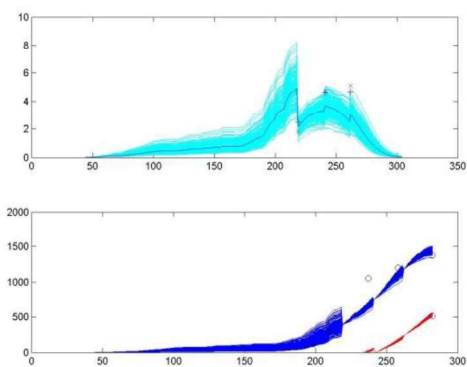
Img 20: RG02



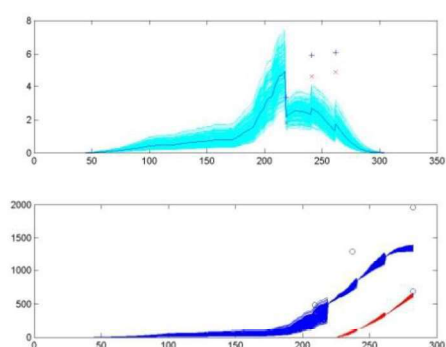
Img 21: RG04



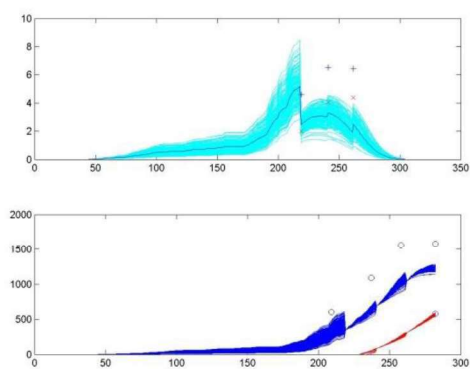
Img 22: RG05



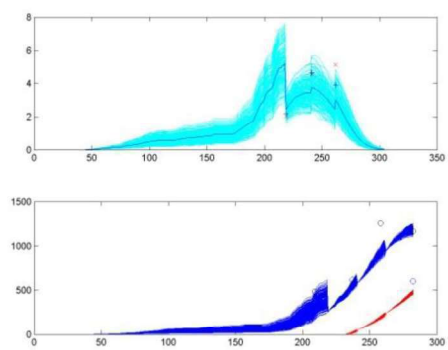
Img 23: RG06



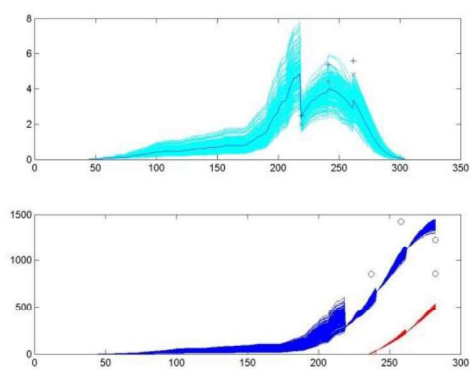
Img 24: RG07



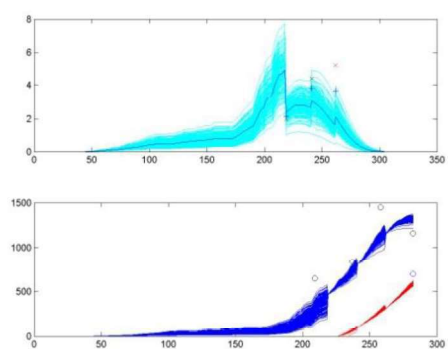
Img 25: RG08



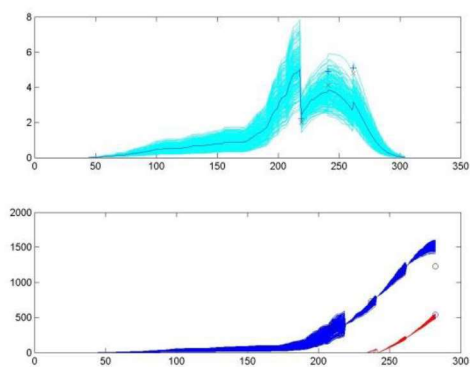
Img 26: RG09



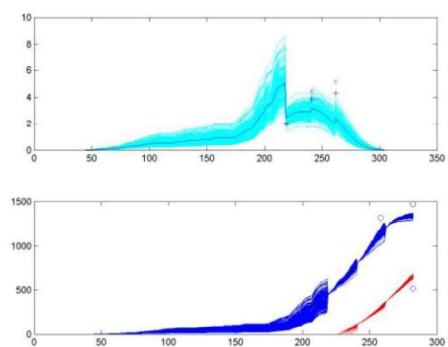
Img 27: RG10



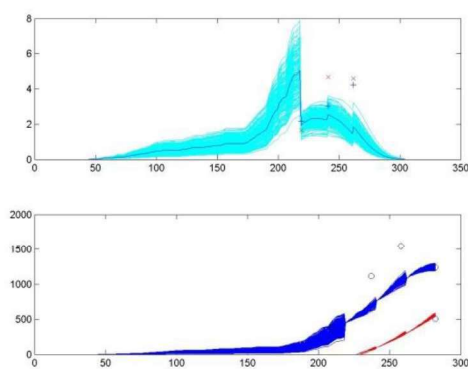
Img 28: RG13



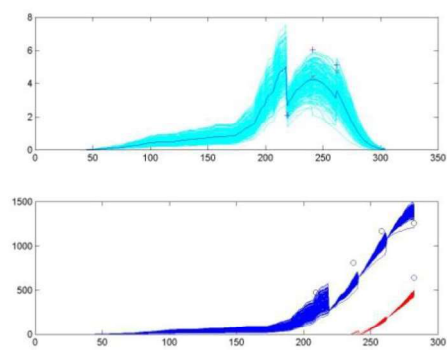
Img 29: RG14



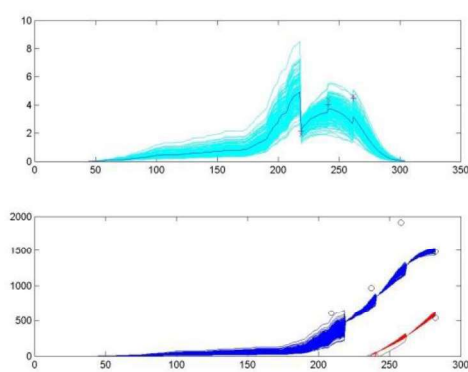
Img 30: RG15



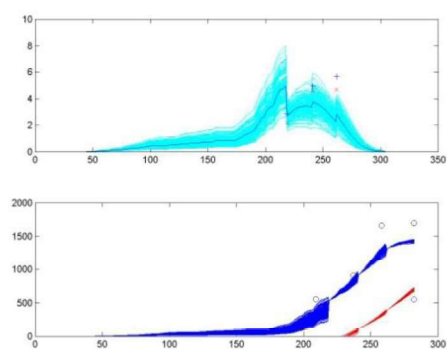
Img 31: RG16



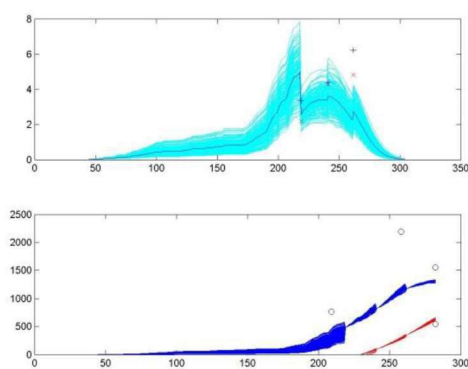
Img 32: RG17



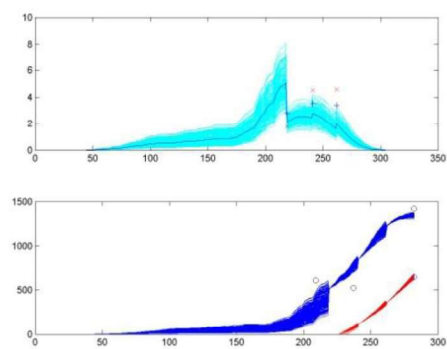
Img 33: RG18



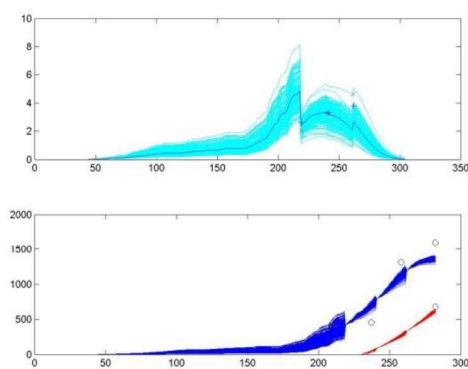
Img 34: RG19



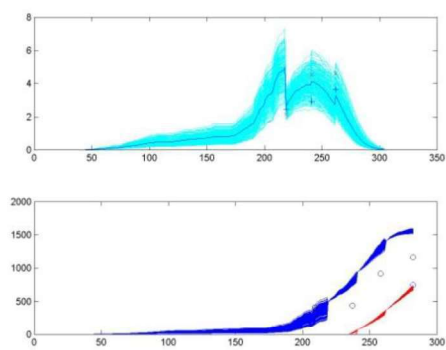
Img 35: RG20



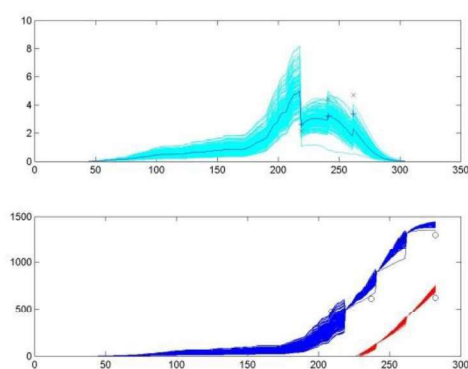
Img 36: RG21



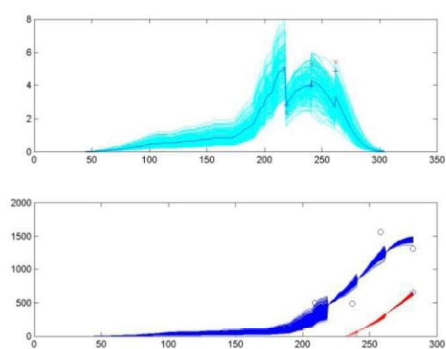
Img 37: RG22



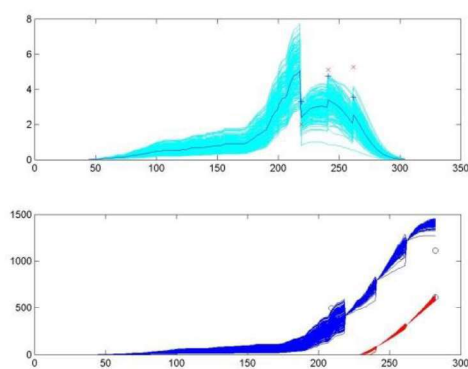
Img 38: RG23



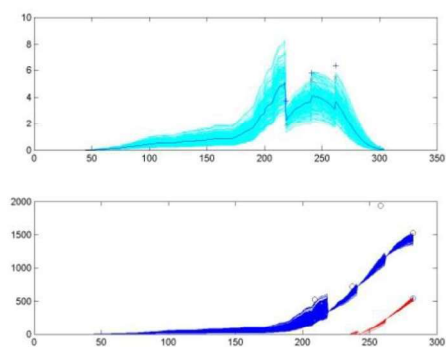
Img 39: RG24



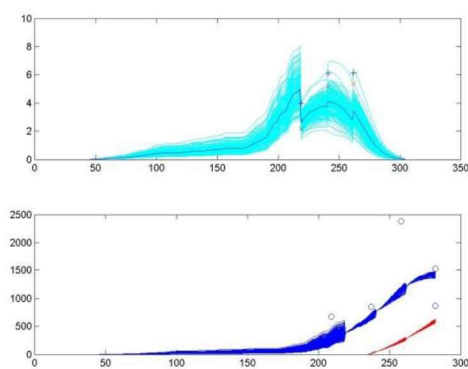
Img 40: WG01



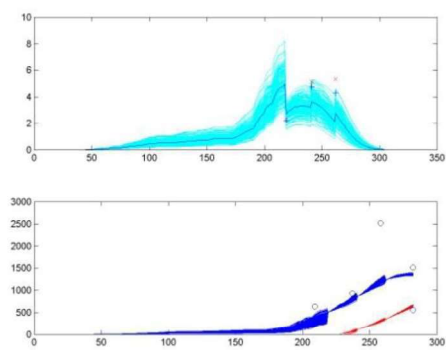
Img 41: WG02



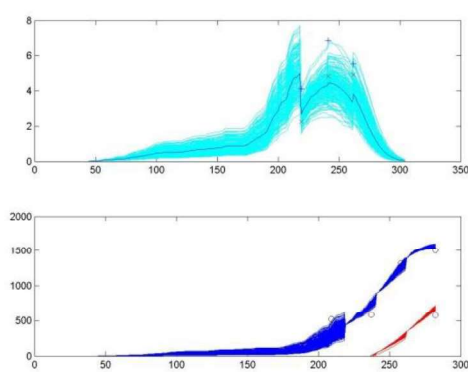
Img 42: WG03



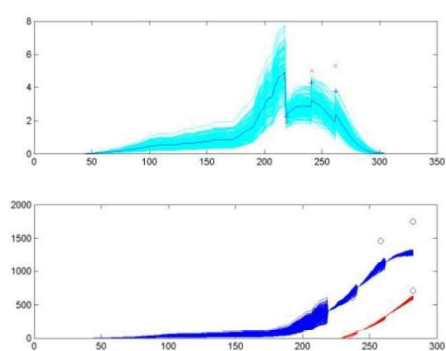
Img 43: WG04



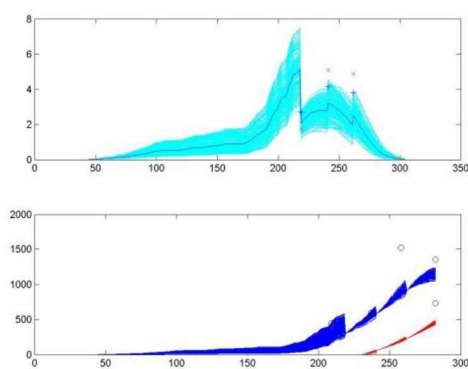
Img 44: WG05



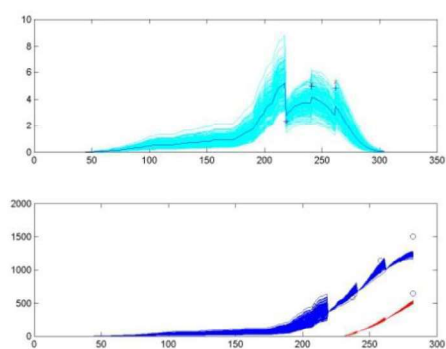
Img 45: WG06



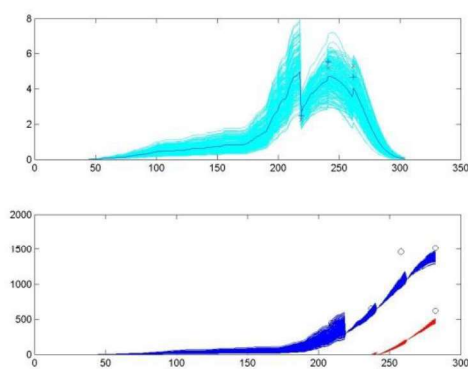
Img 46: WG07



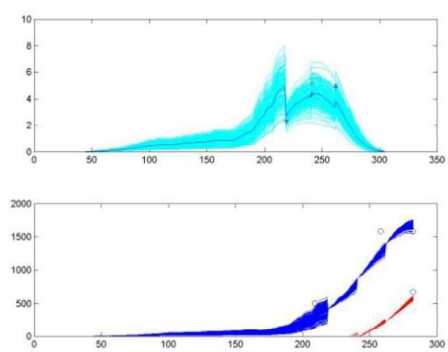
Img 47: WG08



Img 48: WG09

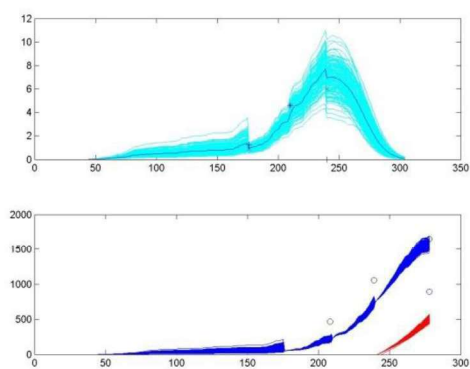


Img 49: WG10

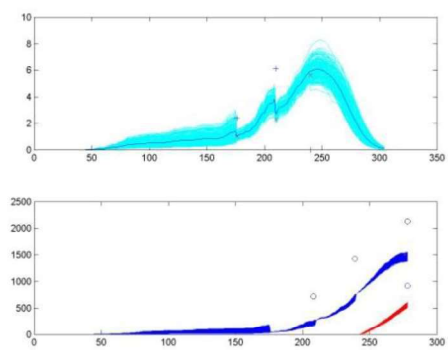


Img 50: WG11

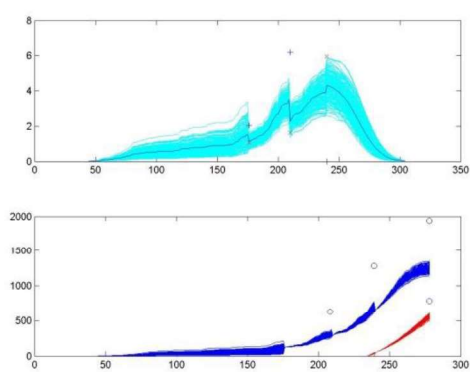
Ciclo Culturale 2014-2015



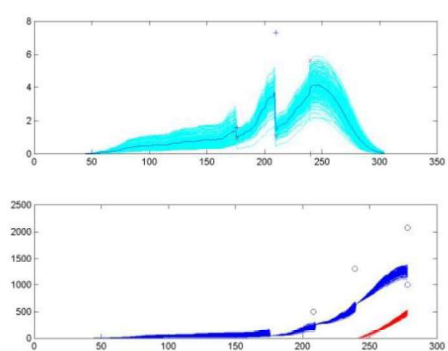
Img 51: RG01



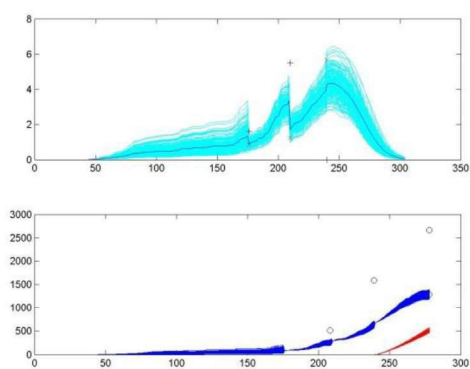
Img 52: RG05



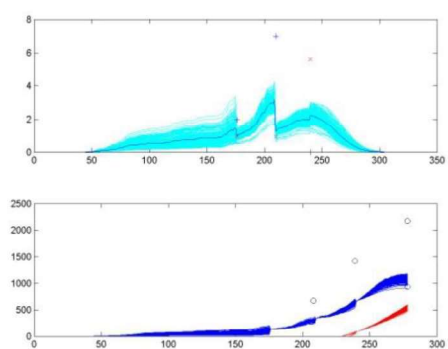
Img 53: RG06



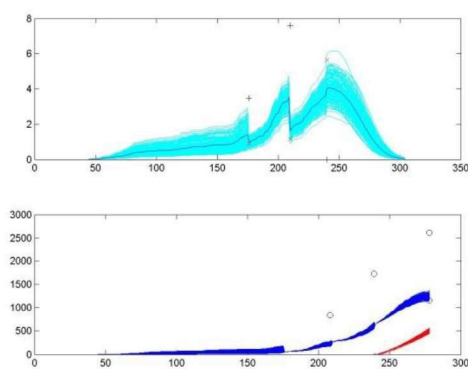
Img 54: RG07



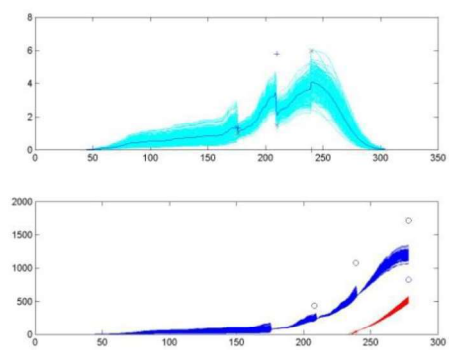
Img 55: RG08



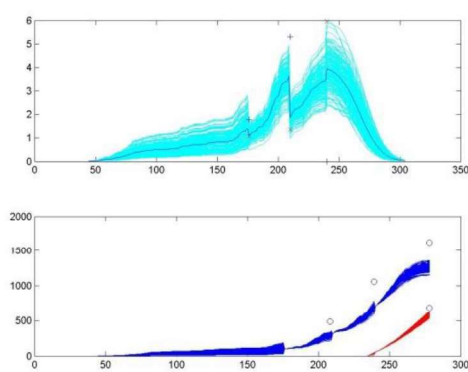
Img 56: RG09



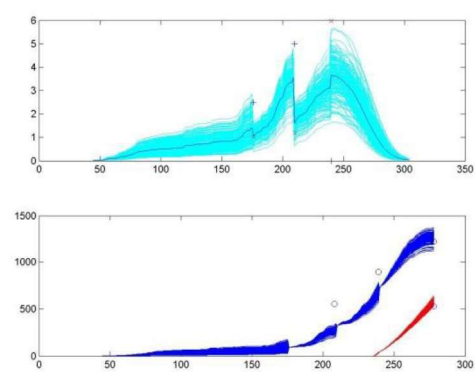
Img 57: RG10



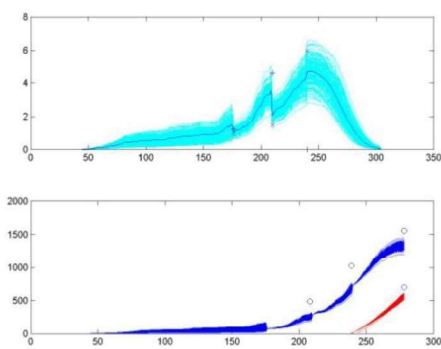
Img 58: RG11



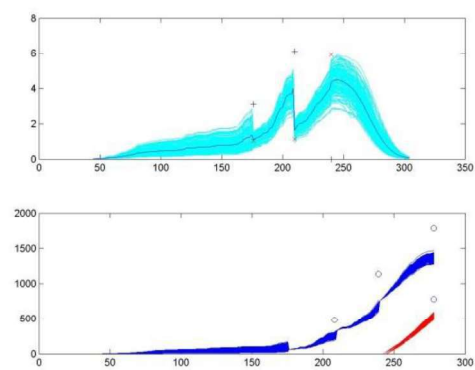
Img 59: RG12



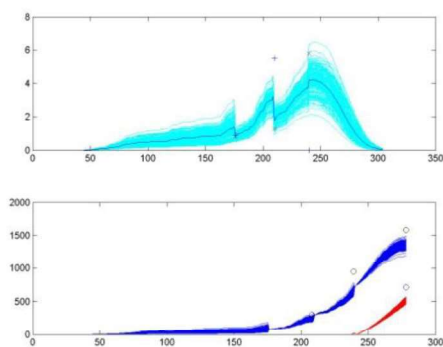
Img 60: RG14



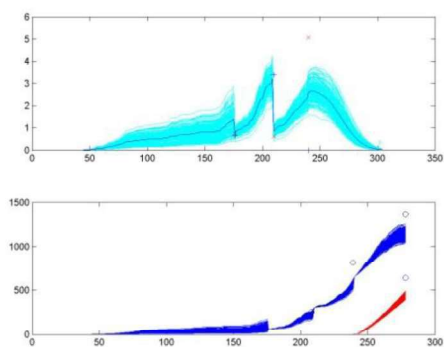
Img 61: RG15



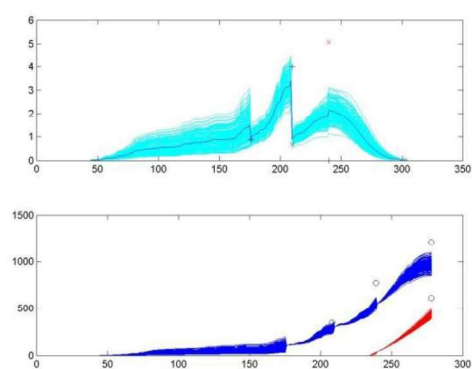
Img 62: RG16



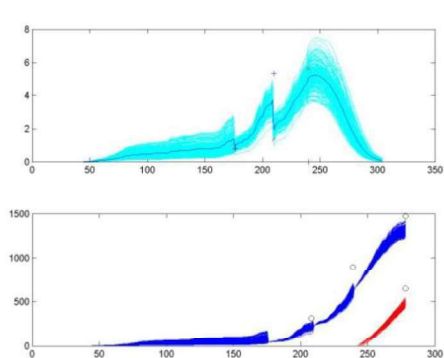
Img 63: RG20



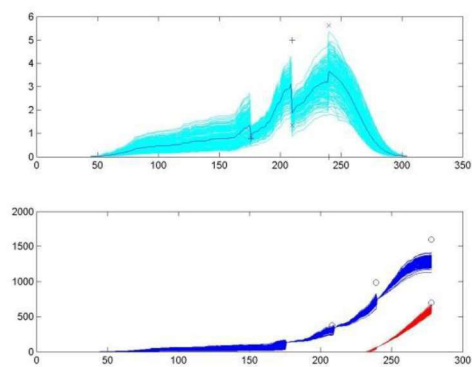
Img 64: RG24



Img 65: RG25



Img 66: RG26



Img 67: RG27

Allegato B

Articoli sottoposti a riviste con impact factor

1. P. C. Silvestro, S. Pignatti, S. Pascucci, H. Yang, Z. Li , G. Yang , W. Huang, R. Casa, 2017. Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and SAFY Models. Sottoposto a *Remote Sensing* (minor revisions al 15/05/2017)
2. Paolo Cosmo Silvestro, Stefano Pignatti, Hao Yang , Guijun Yang, Simone Pascucci, Fabio Castaldi, Raffaele Casa, 2017. Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. Sottoposto a *PLOSONe* (major revisions al 15/5/2017)

Article

Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and SAFY Models

P. C. Silvestro ¹, S. Pignatti ², S. Pascucci ², H. Yang ³, Z. Li ³, G. Yang ³, W. Huang ⁴, R. Casa ^{1*}

¹ DAFNE, Università della Tuscia, Via San Camillo de Lellis, 01100, Viterbo (Italy); p.c.silvestro@unitus.it

² Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy); stefano.pignatti@cnr.it; simone.pascucci@cnr.it;

³ Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097 (P.R. China); yangh@nrcita.org.cn; lizh323@126.com; yanggj@nrcita.org.cn;

⁴ Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094 (P.R.China); huangwj@radi.ac.cn;

* Correspondence: rcasa@unitus.it; Tel.: +39-0761-357555

Academic Editor: name

Received: date; Accepted: date; Published: date

Abstract:

Accurate yield estimation at the field scale is essential for the development of precision agriculture management, whereas at the district level it can provide valuable information for supply chain management. In this paper, HJ1A/B and Landsat 8 OLI images were employed to retrieve leaf area index (LAI) and canopy cover (CC) in the Yangling area (Central China). These variables were then assimilated into two crop models, Aquacrop and SAFY, in order to compare their performances and practicalities. Due to the models specificities and computational constraints, different assimilation methods were used. For SAFY, the Ensemble Kalman Filter (EnKF) was applied using LAI as the observed variable, while for Aquacrop the Particle Swarm Optimization (PSO) was used, using CC. These techniques were applied and validated both at the field and at the district scale. In the field application, the lowest relative mean squared error (RRMSE) value of 18% was obtained using EnKF with SAFY. On a district scale both methods were able to provide production estimates in agreement with data provided by the official statistical offices. From an operational point of view, SAFY with the EnKF method was more suitable than Aquacrop with PSO, in a data assimilation context.

Keywords: leaf area index (LAI); canopy cover (CC); Landsat 8; HJ1A/B; artificial neural networks (ANN); Ensemble Kalman Filter (EnKF); Particle Swarm Optimisation (PSO);

PACS: J0101

1. Introduction

Precision agriculture is being increasingly considered as an optimal technological approach to improve the efficiency and the productivity of farming and cropping systems [1].

Monitoring crop growth using remotely sensed biophysical variables such as leaf area index (LAI) or canopy cover (CC) has proven to be very effective for inferring information on crop biomass and yield, useful for field management [2–4]. Yield estimation and forecasting at the field scale would allow farm managers to plan their agronomic operations, e.g. sowing, tillage or fertilization, on the basis of expected yield potential. At the district level scale, yield estimation and forecasting is useful to public and private organization, for commercial or planning purposes.

The use of remote sensing (RS) in order to monitor crop growth has considerably increased recently, with the development of new technologies such as hyperspectral sensors [5] and easier availability of imagery, allowing to exploit these data in a progressively more effective way [6]. Particularly, satellite RS is increasingly used in monitoring crops and variables that describe their growth, as it allows the monitoring both of large zones on a regional or district scale [7], and of smaller areas on a field scale [6]. The use of these data has proven to be effective in the estimation of crop yield [2; 8–10], by combining the monitoring from remote sensing of crop biophysical variables with crop growth models. Leaf Area Index (LAI), fractional Canopy Cover (CC), the fraction of photosynthetically active radiation absorbed by the canopy (fAPAR) and plant chlorophyll concentration are the most important canopy variables used in this context, being the main state variables used in crop growth modelling [11]. A detailed review of the methods for retrieving canopy variables from remote sensing [11] divided these methods into: 1) statistical approaches, e.g. using vegetation indices and regression methods linking spectral information and biochemical variables [12] and 2) approaches exploiting physically based models [13–15]. The latter are particularly interesting, because they are expected to be of more general validity, without the need of data-specific calibrations as the former methods. Additionally, they often allow to estimate simultaneously several variables such as LAI and CC [14]. A widely used strategy is to train Artificial Neural Networks (ANNs) using the simulations from a canopy radiative transfer model such as PROSAIL [16], which couples the leaf model PROSPECT [17] and the canopy model SAIL [18]. Many studies have shown the effectiveness of this methodology, using both multispectral [14; 19] and hyperspectral [20] sensors.

In order to estimate crop yield, the biophysical variables can be assimilated into crop growth models using different approaches. Following [21], it is possible to distinguish two strategies: calibration and updating. The calibration approach consists of re-calibrating the model parameters in order to minimize the differences between simulated and observed variables. In the updating approach, it is the value of the biophysical variables simulated by the model that is updated, on the basis of the value of the biophysical variables observed e.g. from remote sensing.

Some authors have converted the state variables simulated by the models into vegetation indices (VIs) by means of canopy reflectance models and used VIs, derived from RS data, in the assimilation. For example, [22] developed a methodology in which the LAI simulated by the EPIC crop model was converted into reflectance in the RED and NIR spectral range of Landsat TM or AVHRR data using the SAIL model. NDVI values from these reflectances were then compared with NDVI derived from satellite imagery. The EPIC crop model parameters were re-calibrated for NDVI to be within 20 percent of the observed satellite data [22]. Other authors have estimated biophysical variables from RS and used them directly in the assimilation. For

example, [23] recalibrated the parameters of the SUCROS model by minimizing the difference between LAI estimated from SPOT satellite data and LAI simulated by the model. The updating approach has also been applied in many studies [24–28], in which the effectiveness of this method to improve the accuracy of the crop growth model predictions has always been proven. [24] tested the Ensemble Kalman Filter (EnKF), an updating assimilation method, to assimilate the LAI derived from MODIS into the WOFOST crop model [29], in order to determine the production of wheat at a regional scale. [25] explored the potential of EnKF for simultaneously assimilating two remotely sensed variables, i.e. soil moisture and LAI, demonstrating the possibility to prevent yield losses by monitoring the effects of water stress on crops. The potential of EnKF method assimilation was also compared with the POD4DVAR, a 4 dimensional variational algorithm [27], assimilating LAI estimated by the CCD sensors on board of the Chinese satellite HJ1A/B into the CERES crop model, both at a regional and field scale. The POD4DVAR showed a higher computational efficiency, but was more influenced by measurements uncertainties than EnKF.

Regardless of the approach, in all the studies in which remote sensing data are assimilated into crop models, not all the parameters are recalibrated, because of their large number and the limited information about them available *in situ*. This implies that the choices on which parameters to recalibrate at run-time and of the values to assign to fixed parameters are crucial [23]. A study conducted by [28] highlighted these problems. The Green Area Index (GAI) estimated from MODIS was assimilated in the WOFOST [29] model using a calibration approach, minimizing the error between modelled GAI and MODIS-observed GAI. Only the parameters that most affect GAI were optimized, highlighting the need to calculate the parameters sensitivity for each scenario [28].

The efficiency of the assimilation methods is also strongly dependent on the number of observations ingested into the process [11]. Many studies in the literature (e.g. [24–26]) showed that using a higher number of observations corresponds an improved accuracy of estimation [2]. Generally, the number of observations assimilated into the models is higher than 6. In a context of regional scale, there are many satellites which can provide data frequently, albeit at a low resolution (e.g. MODIS). However, for precision agriculture applications, it is necessary to work at the field scale, with a spatial resolution of at least 30 m or higher. The opportunities to obtain a high number of data acquisitions with such resolution are currently rather limited, therefore, in order to apply these methods at the field scale, it is necessary to use efficient assimilation algorithms, performing well even with a limited number of images, or to extend the number of observations in other ways. For example, [4] tried to solve this problem by using a data fusion algorithm blending MODIS (250 m) and OLI data (30 m) to estimate LAI at the field scale and assimilated the data into the Simple Algorithm For Yield (SAFY) model. The fused data set had the temporal resolution of MODIS data (8 days) and the spatial resolution of OLI (30 m).

In addition to the number of observations, also the choice of the crop model to use is an important factor conditioning the assimilation results. Generally, models that accurately describe the crop growth process have a higher accuracy, but they are more complex, more difficult to integrate with other processes (such as the assimilation methods) and have higher computational costs. Especially the large number of parameters makes the calibration of the model rather laborious, and this is one of the main problems of the application of assimilation

methods. The choice of the model to use, in agreement with the purpose to be achieved, should consider other attributes in addition to model accuracy, such as complexity and plasticity [30–31].

Among all the crop models that have been used in assimilation studies, Aquacrop [32], seems particularly interesting for its relative simplicity, as compared to more complex models, by its emphasis on water limitation and by the explicit use, as main canopy state variable, of CC. This variable is usually easier to estimate from remote sensing than LAI, not being subject to saturation at high values [31]. The first applications of Aquacrop within data assimilation schemes have provided encouraging results [31, 33–35]. However, in Aquacrop there are about 100 parameters related to the characteristics of the crop and the soil, as well as other management and input factors. During the assimilation, [33] recalibrated 3 parameters, [34] and [35] 8 parameters and [31] 13 parameters, leaving all the others fixed. Additionally, in the version currently made available by the developers, it is impossible to implement a re-initialization at the RS observation dates, preventing the application of updating approaches of data assimilation, so that only recalibration approaches have been employed so far.

SAFY [36], a much simpler model than Aquacrop, with less parameters, i.e. 13 in the original version, has been also successfully employed for the estimation of wheat yield, using only recalibration assimilation approaches so far [4, 36]. This model is potentially suitable also for the application of updating assimilation algorithms, given the accessibility of the source code, and it's particularly appealing, in this context, also for its high computational speed [31]. The aim of this work was to develop and compare two methods for the estimation of winter wheat yield, suitable both at field and district level, based on data assimilation algorithms combining remotely sensing crop biophysical variables with the Aquacrop and SAFY crop growth models.

Because of the above mentioned coding constraints, it was necessary to use different assimilation techniques for the two models, i.e. the EnKF was employed for SAFY, while for Aquacrop the Particle Swarm Optimization (PSO) [37] was used. A dataset, composed by a series of images acquired with multispectral satellite sensors, was used to estimate the LAI and the CC using a model-based method employing Artificial Neural Network [14]. Additionally, the purpose of this study was to test the performance of assimilation methods with a reduced number of observations, in order to verify their potential for precision agriculture applications, in case of constraints e.g. due to a limited number of satellite acquisitions available within the crop growth season, a situation often occurring in areas affected by frequent cloud cover.

2. Materials and Methods

2.1. Study area and datasets employed

The study area (34.27°N, 108.09°E, altitude around 460 m) is located in the rural region of Yangling, in the centre of China, Province of Shaanxi (Figure 1). A dataset composed by weather data, field data and remote sensing data, was collected during three different winter wheat growth seasons, between September and June, from 2012 to 2015.

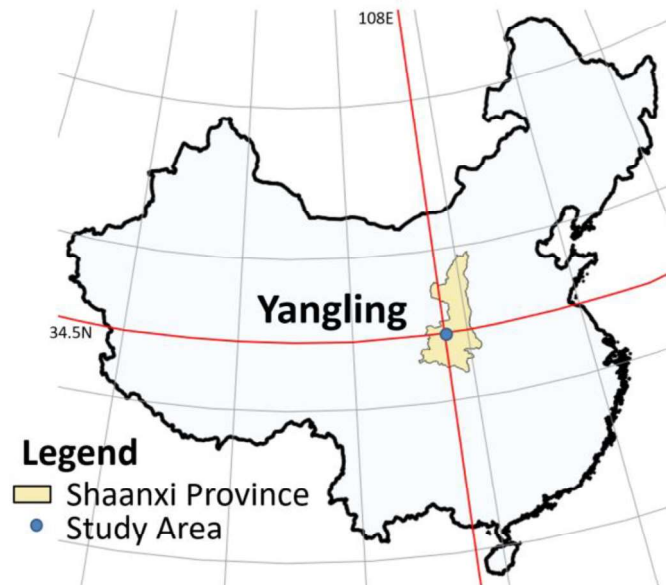


Figure 1. Map of China showing the location of the test area of Yangling (Shaanxi Province)

The weather data were collected at the weather station ID 57034 (34.25°N 108.22°E), by the China Meteorological Administration and made available on the China Meteorological Data Sharing Service System (CMDSSS, <http://cdc.cma.gov.cn>). They included: temperature, rainfall, relative humidity and wind speed. Other climatic data needed to drive the crop models, such as evapotranspiration and solar radiation were derived from these variables. Reference evapotranspiration was calculated using the ET_0 calculator [38], whereas solar radiation was estimated from the relationship proposed by [39]. Temperatures and rainfall for winter wheat growth seasons of 2012-2013, 2013-2014 and 2014-2015 in Yangling are shown in Figure 2.

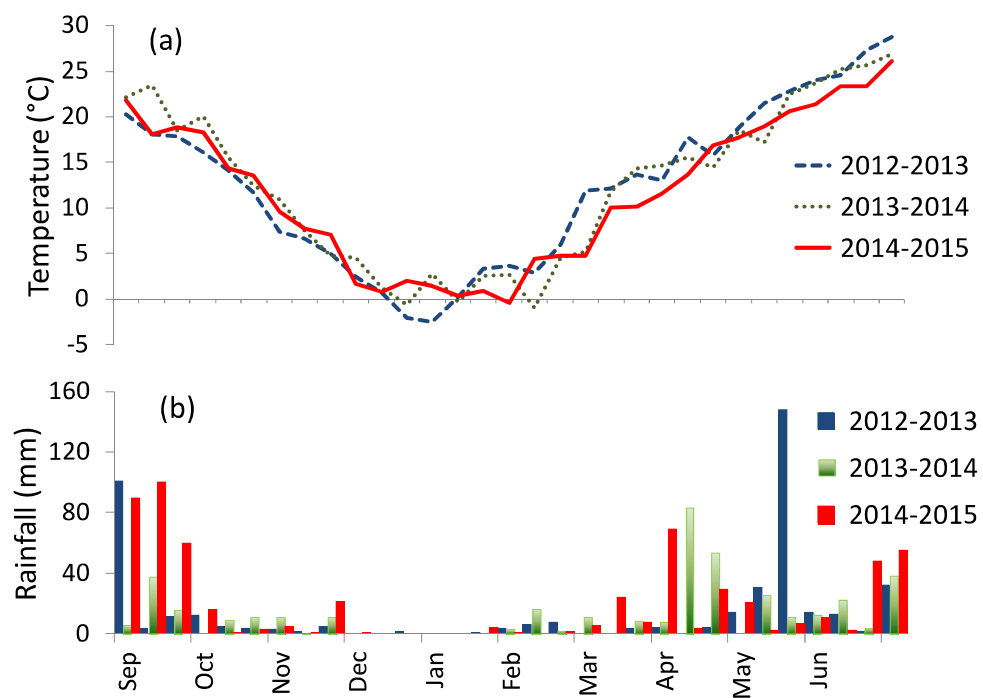


Figure 2. Ten days mean temperatures (a) and rainfall (b) for the Yangling study site, for the wheat crop cycles (1 September to 30 June) of 2012-2013 (blue), 2013-2014 (green), 2014-2015 (red).

The climate of the Yangling area is defined as semi-arid, with precipitations of around 500 mm per year, usually concentrated in the summer, between June and October, while the winter season is almost devoid of precipitation. Because of this rainfall distribution, winter wheat is mandatorily irrigated. Temperatures vary from a minimum of -4°C in winter to a maximum of 30°C in summer, even if temperatures below 0°C are not common, and were limited to a few days in winter for our study period, except for 2012-2013, when the temperatures remained under 0°C for several weeks between December and January. This means that the period of cold stress is limited to few days. The rainfall for the three crop growth seasons considered was almost completely absent between December and January, increasing in spring, concentrating large amounts of rain in a few days, reaching a peak in summer, and the decreasing until early autumn, when the dry season begins.

Field measurements of wheat LAI, biomass and yield, were carried out each year from March to June by staff from the National Engineering Research Centre for Information Technology in Agriculture (NERCITA). Table 1 shows the list of the field measurements dataset.

Table 1. Field measurements of LAI (L), biomass (b) and yield (y) conducted in Yangling from 2013 to 2015. n. pts are the number of sampling points where ground data were collected.

Year	n. pts	March	April	May	June
2013	49	30 Mar	27 Apr		1 Jun
		L, b	L, b		b,y
2014	35	5 Mar, 28 Mar	22 Apr, 27 Apr	16 May	9 Jun
		L, b	L, b	L, b	b,y
2015	28	27 Mar	25 Apr		5 Jun
		L, b	L, b		b,y

The ground measurement points, georeferenced using a GPS, were distributed in different winter wheat fields throughout an area of about 1200 km^2 surrounding Yangling. In addition to the sampling points reported in Table 1, there were additional points, 10 in 2013 and 26 in 2014, for which data were incomplete (i.e. not all sampling dates were present). These points were excluded from the yield and biophysical variables validation, but were used for the validation of the classification of wheat fields (see section 2.2.5). Three local wheat cultivars (Xiaoyan 22, Xinong9871, and Shanbei139) were planted in the first half of October and harvested at the beginning of June. Field management (weed control, pest management, and fertilizer application) followed local standard practices for wheat production. A different number of ground points was sampled in different years (Table 1). For each ground point, LAI was estimated using the LAI-2000 Plant Canopy Analyzer (LI-COR Inc., Lincoln, NE, USA) collecting data from 5 points within a $30\times 30\text{ m}$ square. LAI was converted into fractional

ground canopy cover (CC) using the equation proposed by [40] from field data collected in the USA, who reported a r^2 of 0.96:

$$CC = 94 \cdot (1 - e^{-0.43LAI})^{0.52} \quad (1)$$

Biomass and yield were determined by cutting plants from a 0.25 m² area positioned near the point of LAI measurements. All plant samples were oven dried at 70°C to a constant weight, and final dry weight recorded. Further details on the methodologies of data collection can be found in [33].

The remote sensing dataset was composed by several images acquired by Landsat 8, HJ1A and HJ1B satellites, between February and June of each year (Table 2). All the images were acquired using multispectral sensors, respectively the OLI on board the Landsat 8, and the CCD1 and CCD2 on board the HJ1A and HJ1B.

Table 2. Remote Sensing data set for Yangling, 2013–2015

crop cycles			
Year	Month	Day	Satellite
2013	March	5	HJ1B
2013	March	28	HJ1B
2013	April	26	HJ1A
2013	May	11	HJ1A
2014	March	15	Landsat8
2014	April	7	HJ1A
2014	April	29	HJ1A
2014	May	20	Landsat8
2015	February	23	Landsat8
2015	March	29	HJ1A
2015	April	28	HJ1B

The OLI sensor has 7 bands with a spatial resolution of 30 m, with wavelengths ranging between 0.45 μm and 2.29 μm . Only the 4 bands in the visible and NIR regions were employed as inputs of the Artificial Neural Network algorithm used to calculate the crop biophysical variables (see section 2.2.2). This was done also because the CCD sensors (on board of HJ satellites) have similar 4 bands, in the region of visible and NIR, also with a spatial resolution of 30 m. Details on the spectral characteristics of the OLI and CCD sensors can be found in [41–42].

Surface reflectance Landsat 8 images were provided, geographically and atmospherically corrected, by the United States Geological Survey (USGS). The HJ images needed georectification, radiometric calibration and atmospheric correction. The georectification was done by identifying ground control points in GoogleEarth and warping the images using the Image-to-Map registration tool in ENVI [43]. The radiometric calibration was performed using gain and offset values provided for each band by the metadata of the respective HJ images. The atmospheric correction was performed using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) module included in the ENVI software. The altitude of the

sensor was set at 650 km and the ground elevation at 500 m (Yangling mean elevation). The mid-latitude winter atmospheric model and the rural aerosol model were used, with an initial visibility between 20 and 30 km, depending on the image.

2.2 Data assimilation methods

It is possible to divide the methodology adopted (Figure 3) into four steps: in the first step the biophysical variables LAI and CC were estimated; in the second step the assimilation was carried out: LAI values were assimilated into the SAFY model through an updating method based on the EnKF, whereas CC values were assimilated into Aquacrop using the PSO method; in the third step the two methodologies were applied and validated at the field scale using ground measurements; in the fourth step the method was extended to the district scale and validated using official statistical data. Details on each of these steps is provided in the following paragraphs.

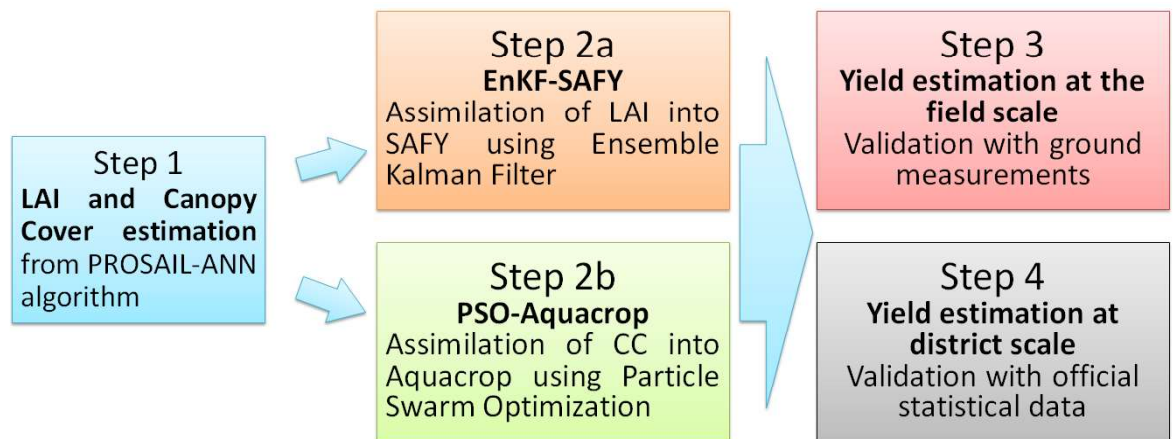


Figure 3: Flowchart of the methodology applied, from RS data to yield estimation.

2.2.1. LAI and Canopy Cover estimation

LAI and CC were retrieved using the algorithm developed by [15]. This algorithm converts measured spectral reflectances into biophysical variables, by means of Artificial Neural Networks (ANNs) trained over a database of canopy reflectances simulated using the radiative transfer model PROSAIL [16]. The PROSAIL input dataset was composed by different kinds of parameters characterizing the crop canopy, sensor characteristics and satellite configuration (Table 3).

In the present study, canopy, leaf and soil parameters were set following the experimental results obtained by [20] for winter wheat, except for the mode of the LAI. The latter was set using as reference the field measurements dataset, setting the LAI mode at 1.5 for the images acquired in February and in the first week of March, at 2.5 for the images acquired between March and the beginning of April and at 4 for the images acquired between April and May.

Table 3. Input parameters of the PROSAIL radiative transfer model used to generate the training database. The distribution used for each parameters is the truncated Gaussian function defined by: lower (min) and upper bounds (max), mode and the standard deviation (Std). N_C is the number of classes in which the distribution is split. The LAI mode was varied according the period of satellite data acquisition.

Variable	min	max	mode	Std	N_C
----------	-----	-----	------	-----	-----

Canopy	LAI	0.0	10.0	1.5 / 2.5 / 4	4.0	6
	Average leaf angle (degrees)	30	80	60	20	4
	N	1.20	2.20	1.50	0.30	4
Leaf	Chlorophyll content ($\mu\text{g m}^{-2}$)	20	90	45	30	4
	Dry matter content (g m^{-2})	0.0030	0.0110	0.0050	0.0050	4
	Relative water content	0.60	0.85	0.75	0.08	4
	Brown Pigment index	0.00	2.00	0.00	0.30	4
Soil	Brightness Parameter	0.50	3.50	1.20	2.00	4

The sensor characteristics input included the relative spectral response of each single band and four error parameters, accounting for instrumental noise, radiometric calibration and atmospheric correction errors.

The satellite configuration inputs include: geographical position of the target, time and day of year of acquisition, sun and view zenith and azimuth angles. This information is different for each image, so the ANN algorithm was run specifically for each acquired image. For each image a total of 98,304 PROSAIL model simulations were used to train the ANNs, in order to associate a given set of input variables and parameters (characterizing leaf, canopy and geometry of observation) to a reflectance spectrum. After the training, the ANNs were employed to "invert" the process and associate a given spectrum observed by the satellite (corresponding to a pixel and resampled according to the bands of the sensor) to a set of biophysical variables. Five ANNs were employed in this study and the median was used as the biophysical variable estimate. To apply the ANN algorithm, the Matlab (MathWorks, MA, USA) tool developed by [19] was used.

2.2.2. Crop growth simulation models

As mentioned, two different crop growth models were used in this study: Aquacrop [32] and SAFY [36].

Aquacrop was developed by the Food and Agriculture Organization of the United Nations (FAO) to describe crop response to water and simulates the aboveground biomass production mainly as a function of Water Productivity (WP):

$$B_n = WP^* \cdot \sum_{i=1}^n \frac{Tr_i}{ET_{0_i}} \quad (2)$$

where B_n is the cumulative aboveground biomass production after n days ($\text{g}\cdot\text{m}^{-2}$), Tr_i is the daily crop transpiration (mm day^{-1}); ET_{0_i} is the daily reference evapotranspiration (mm day^{-1}); WP^* is the normalized crop water productivity ($\text{g}\cdot\text{m}^{-2}$). The canopy cover (CC) is a crucial state variable of Aquacrop through its expansion, ageing, conductance and senescence, since it determines the amount of water transpired, which determines the amount of biomass produced [32]. The final yield is the product of the final biomass multiplied by the harvest index [44]. A detailed description of the Aquacrop crop model can be found in [32] and [44].

The input information necessary to run Aquacrop model consists of: daily weather data (maximum and minimum temperature, rainfall and evapotranspiration), soil composition and agronomic management information. A set of more than 100 parameters is used, particularly linked to crop physiological and soil physical processes. The other model used in this work is the Simple Algorithm For Yield estimate (SAFY), developed specifically for remote sensing applications [36]. It is based on Monteith's concept [45], a simple theory linking the total dry phytomass production and the photosynthetically active solar radiation (APAR) absorbed by

plants. The purpose of the algorithm is to represent the time courses of aboveground biomass (DAM), Green Leaf Area Index (GLAI) and yield. The daily increase of DAM (ΔDAM) is represented by the following equation:

$$\Delta DAM = R_g \cdot \varepsilon_C \cdot \varepsilon_I \cdot ELUE \cdot F_T \quad (3)$$

where R_g is the incoming global radiation, ε_C is the ratio of incoming PAR to global solar radiation, ε_I is the light interception efficiency (fraction of APAR), $ELUE$ ($g \cdot MJ^{-1}$) is the light use efficiency and F_T is a temperature stress function. The ε_I depends on the green leaf area index (GLAI) and a light interception coefficient k through Beer's law (4):

$$\varepsilon_I = 1 - e^{-k \cdot GLAI} \quad (4)$$

GLAI development is divided in two phases, from emergence to the beginning of senescence, it increases proportionally with DAM, from senescence to the end of the growth cycle it decreases. In this way biomass (DAM) depends directly on GLAI, but at the same time the daily development of GLAI ($\Delta GLAI$) depends on DAM (positive feedback). The grain yield is computed from DAM through a partitioning coefficient of the biomass to grain (Pgro_P2G). A detailed description of SAFY was presented by [36].

In this work, a modified version of SAFY was used, in which a simple water balance was introduced, following the FAO guidelines for computing crop water requirements [46]. The same modification had been formerly proposed by [47] and [48], introducing a dependence of biomass yield on crop water stress. This modification was made since it allows an explicit assessment of the crop response to water availability and the characterization of water stress, which is lacking in the original model. In the present work, the ΔDAM expressed in equation (3) is multiplied by a water stress factor K_s , a dimensionless transpiration reduction factor dependent on available soil water, ranging between 0 and 1 [49]. K_s is calculated from the total available soil water in the root zone (TAW) and readily available soil water in the root zone (RAW), resulting from a simplified water balance driven by crop evapotranspiration. The necessary input to run the modified version of SAFY are daily information about: incoming global radiation, average air temperature, rainfall and reference evapotranspiration. Further details on this modified version of SAFY are reported by [49].

Both models were subjected to a sensitivity analysis [49], in order to identify the parameters that most affected the yield estimated by the models. This procedure is essential for the choice of a reduced set of parameters to target during the assimilation algorithms, leaving all the others at fixed values. Especially for the latter, a preliminary calibration is an important step. The calibration of parameters for both models was done starting from reference values for winter wheat taken from the literature: [36] for SAFY and [50] for Aquacrop. The data on soil properties were obtained from samples gathered in the study site, at the same points where the crop data were collected, which were analyzed in the lab by NERCITA (Beijing) for soil texture and organic matter. Crop management input variables were set according to standard local practice. Subsequently, the value of parameters which resulted most influential from the sensitivity analysis were adjusted, by trial and error, so that the time series of simulated biophysical variables (LAI for SAFY and CC for Aquacrop) fitted the time series observed in two of the ground sampling points (Table 1). These two sampling points were randomly chosen from field measurements dataset and were subsequently excluded from the validation tests. In all the applications with the SAFY model, a unique calibration was carried out, using two points randomly selected from those collected in 2014, since that was the year with most observation dates. For Aquacrop a single calibration for all the applications of the model was not satisfactory, as described in the results. For this reason the model was alternatively calibrated also for each year, using a different set of points, chosen from the experimental set collected during the corresponding year. In this case three different calibrations were done for Aquacrop, one for

each year in which the field measurements were made. A unique set of parameters and input variables was used for the simulations carried throughout the whole study area. The spatial variability among parameters and inputs was subsequently introduced by means of the assimilation algorithms (described in the following sections), by adjusting some parameters.

2.2.3. Ensemble Kalman Filter for SAFY model assimilation

An updating assimilation algorithm based on the EnKF was developed, taking advantage of the accessibility and manageability of the SAFY code. Before applying this method, the model parameters that most influenced the yield were identified from the results of the sensitivity analysis study [50]. In Table 4, the list of the 9 most influential parameters and their relative range of variation are shown. The initial calibration values of the parameters were defined as described in section 2.2.2. The range of variation of the selected parameters was chosen by taking the calibrated values as a mean value. The minimum and maximum were then chosen arbitrarily in a surrounding of maximum $\pm 13\%$ of the mean value, according to a realistic range for each parameter based on expert judgement.

Table 4. List of the most influential parameters of the SAFY model allowed to vary during the assimilation, identified from a previous global sensitivity analysis study [50], in order of decreasing importance.

Parameter	Min	Max	Description
Pgro_R2P	0.7	0.9	Climatic efficiency: ratio of incoming PAR to global radiation
Pfen_SenA	1260	1540	Temperature sum threshold to start senescence (Growing Degree Days, GDD)
Ptfn_Topt	20	22	Optimum temperature for plant development ($^{\circ}\text{C}$)
Pfen_MrgD	288	300	Day of the year of emergence
Pgro_Lue	1.75	1.95	Effective light-use efficiency: ratio of energy produced as DAM from APAR (g.MJ^{-1})
Pfen_PrtA	0.16	0.2	Partition to leaf function parameter 1 (PLa)
Pgro_P2G	0.0098	0.012	Partition coefficient to grain
SMT_sen	2070	2530	Temperature sum to complete senescence (GDD)

The EnKF algorithm developed for SAFY in this work is based on the theory of [51], which considers the observations as random variables, therefore adding random perturbations to the observed values. It follows the application to crop models described by [52]. It is possible to divide the assimilation algorithm into the following steps:

- 1) Generation of an ensemble of $n=100$ vectors, each one containing the values of the i parameters P_i , in this case $i=1,2,\dots,9$ (Table 4). To each element, corresponding to a nominal parameter value (fixed as the mean of the range shown in Table 4) an error

value was added, randomly drawn from a truncated normal distribution $N(0, 1)$, with lower and upper limits as shown in Table 4.

- 2) Simulations with the SAFY model in order to obtain a value of LAI for each element of the ensemble at the date when the first satellite image was acquired. An error ε is added to the simulated LAI values. This error ε was randomly generated from a normal distribution $N(0, Q)$, in which the standard deviation Q was arbitrarily chosen as 20% of the LAI value, to take into account the uncertainty of the model. In this way a matrix φ_{t_1} was defined:

$$\varphi_{t_1} = \begin{bmatrix} L^1 & L^2 & \dots & L^{100} \\ P_1^1 & P_1^2 & \dots & P_1^{100} \\ \dots & \dots & \dots & \dots \\ P_9^1 & P_9^2 & \dots & P_9^{100} \end{bmatrix} \quad (5)$$

where each column is an element of the ensemble and represents a random configuration of the model using parameters P_i and the corresponding LAI value obtained L^n at the time t_1 of the first satellite acquisition.

- 3) Generation of the vector M_{t_1} , where each element is composed by the LAI observed at time t_1 , i.e. retrieved from remote sensing data, plus an error $\tau_{t_1}^j$ drawn from $N[0, \text{var}(\tau_{t_1}^j)]$, where $\text{var}(\tau_{t_1}^j)$ is a variance expressing the measurement error, which in our case was inferred from the comparison of the LAI values estimated from remote sensing, with the ground measurements. In this study the measurement error variance was chosen as $\sqrt{\text{var}(\tau_t)} = 0.3 \cdot M_t$, taking into account the RRMSE found for LAI estimation from our measurements (see section 3.1).
- 4) Computation of the variance-covariance matrix of φ_{t_1} for the 100 ensemble elements.
- 5) Calculation of the Kalman gain using the variance-covariance matrix :

$$K_{t_1} = \frac{\sum_{t_1} R^T}{R \sum R^T + \text{var}(\tau_{t_1})} \quad (6)$$

where: $R=(1,0,0,\dots,0)$ with number of elements equal to the number of ensemble elements.

- 6) Update of φ_{t_1} as follows:

$$\varphi_{t_1,K}^j = \varphi_{t_1}^j + K_{t_1} (M_{t_1}^j - R \varphi_{t_1}^j)$$

where j is the j -th column of φ_{t_1} .

- 7) Replacement of LAI and parameters values (the elements of φ_{t_1}) with those calculated at step 6 (the elements of $\varphi_{t_1,K}$).

- 8) Repetition from step 3 for each satellite observation date. When the last observation has been assimilated, SAFY runs to the end of the crop growth cycle and outputs the yield.

2.2.4. PSO algorithm for Aquacrop model assimilation

The source code of the Aquacrop model was not available to the authors and this prevented the addition of changes in order to use an updating assimilation method, such as the EnKF used for SAFY. For this reason a method of assimilation by calibration [11], PSO, was employed. This method is based on a comparatively simple principle, needs few input parameters, can be easily applied to other models and was shown to have good computation efficiency [37]. The basic assumption of PSO is that a group of m particles ($m=15$ in this study) flies with certain speeds without quality and size in a d -dimensional search space. In this study m was chosen as equal to 15, which is lower than the number of particles suggested by [54], i.e. 20-50, since these require considerable computing resources, which were not available for the present study. However, it has been reported that there are no drawbacks from using a lower number of particles provided that this is higher than the number of parameters [54-55], which was 9 in our case. Each particle can modify its position and velocity based on both the best point in current generation (p_{id}) and the best point of all particles in the swarm (p_{gd}). The PSO assimilation method is described by [54] and its application for Aquacrop, that can be found in [33-35], is summarized in the following steps:

- 1) the initial position (value) and velocity of each particle was determined. The adjusted parameters were selected from the sensitivity analysis study previously carried out [49] and are listed in Table 5 .
- 2) Aquacrop was run using the executable file (ACsaV40) released by FAO and controlled by a Matlab script. The simulated CC (CC_s) was calculated from the Aquacrop input including weather data, soil data, crop genotype parameters and management data.
- 3) A cost function (J) was constructed to assess the discrepancy between CC_s and the remote sensing estimated CC (CC_r) as shown in equation (9). The minimization of the cost function determined whether the optimization algorithm had reached the optimum input parameter values.

$$J = \sqrt{\sum_{i=1}^N \left(\frac{CC_{s_i} - CC_{r_i}}{CC_{r_i}} \right)^2 / N} \quad (9)$$

where CC_{s_i} and CC_{r_i} are respectively the simulated and remote sensing values of canopy cover at observation date i , N is the number of dates and J is expressed as % (i.e. the unit of CC).

- 4) Seek the best point of the i -th particle (p_{id}) and of all particles in swarm (p_{gd}). The p_{id} and p_{gd} were searched at each iteration.
- 5) Update the position and velocity of particles as shown in the equation [10]:

$$\begin{aligned} v_{id}^{k+1} &= v_{id}^k + c_1 \zeta (p_{id}^k - x_{id}^k) + c_2 \eta (p_{gd}^k - x_{id}^k) \\ x_{id}^{k+1} &= x_{id}^k + v_{id}^{k+1} \end{aligned} \quad (10)$$

where v_{id} is the best velocity of the i -th particle; k is the index of the iteration; c_1 and c_2 are constants which were assigned a value of 2; ξ and η are random values between 0 and 1; x_{id} is the best position of the i -th particle.

- 6) If the iteration target was not reached, the updated positions were replaced and the next step was carried out. In this study the iteration target was fixed at 30 because it is the minimum iteration number indicated in literature [54]. Choosing a higher number does not allow to run the script within an acceptable time.
- 7) If the iteration target was reached, the CC, biomass and yield were calculated.

Table 5. List of the most influential parameters of the Aquacrop model allowed to vary during the assimilation, identified from a previous global sensitivity analysis study [49], in order of decreasing importance.

Parameter	Min	Max	Mean	Description
dos	274	284	279	Day of sowing (Day of year)
mat	2200	2600	2400	Length of the crop cycle (GDD)
flo	1500	1700	1600	Growing Degree Days (GDD) from sowing to flowering
rtx	1	1.4	1.2	Maximum effective rooting depth (m)
wp	16	20	18	Water productivity normalized for ETo and CO ₂ (g m ⁻²)
hi	35	45	40	Reference Harvest Index (HI ₀) (%)
polmn	6	10	8	Minimum air temperature below which pollination starts to fail (°C)
kc	0.08	1.3	1.1	Crop coefficient when canopy is fully covering the ground but prior to senescence
stbio	5	16	11	Minimum GDD required for full biomass production

2.2.5 Spatialized Application

The application at the district level was done for the winter wheat growth seasons of 2013 and 2014. These two years were chosen because official yield data, used for validation, were made available. Information was provided by two official statistical sources: the National Bureau of Statistics of the People's Republic of China (which provided data for the year 2013) and the Shaanxi Provincial Bureau of Statistics (which provided those for 2014).

For each year, the biophysical variables values retrieved from the images acquired between March and May were used in the assimilation, which was run on a pixel-by-pixel basis. For each image, only the pixels belonging to the wheat fields included in the rural area of Yangling were considered. To identify the pixels in which wheat was present, 5 Landsat 8 images acquired between February and June of each year were used. The images were converted into LAI maps through the PROSAIL-ANN algorithm (section 2.2.2). A mask was built using agricultural fields

boundaries obtained from a vector file provided by the National Engineering Centre for Information Technology in Agriculture. In this way all the pixels that fell out of the agricultural areas were excluded.

The remaining pixels of each image were classified as wheat if the following conditions occurred, representing the typical LAI time trends of winter wheat:

- LAI of the 1st date smaller than LAI of 2nd date and the latter smaller than LAI of 3rd date;
- LAI of 3rd date higher than 3;
- LAI of 4th date higher than 3.4
- LAI of 5th date smaller than 0.5 (after harvest).

Not all the pixels included within the boundaries of some fields were classified as wheat, but when the majority, i.e. 80%, of pixels included within the boundaries of a field were classified as wheat, the whole field was considered as such. The classification results were checked against ground truth data. There were 59 points (i.e. ground verified wheat fields) available in 2013 and 61 points in 2014, considering also incomplete data sampling points, in addition to those reported in Table 1. Of these, 57 fields were correctly identified as wheat in 2013 and 58 in 2014, with an overall accuracy of 96%.

To reduce the computational cost of the spatialized assimilation algorithms, a data binning procedure based on LAI value was employed so that the model was run only once for similar pixels. The LAI values in the images, retrieved from remote sensing, were rounded to the first decimal place and, if the number in that place was odd, it was rounded to the nearest lower even number. In this way, the range between the minimum and maximum values of LAI was divided into intervals of 0.2. The total number of LAI classes is the product of the number of intervals between the minimum and the maximum feature value of LAI for each image. Each class is defined by the combination of 3 numbers, 0.2 multiples, belonging to the range min LAI-max LAI of each image, for the 3 dates. In this way, after running the algorithm for all the combination of LAI, it is possible to assign a value of yield to each class. Classifying the points of the map at each pixel, the corresponding yield value was assigned. In this way it was possible to obtain yield maps of the region, by running the algorithm only for the number of identified classes (around 1800) and not for the roughly two million pixels of the region of interest. A similar data binning procedure was employed for the images representing the CC, but in this case the range of variation between a class and the next was not constant, because of the non-linearity of the relationship between LAI and CC (eq. 1). Furthermore the ranges were enlarged, in order to reduce the number of classes, because the computational cost of PSO-Aquacrop imposed a limited number of iterations. The 1800 bin classes identified for LAI were thus reduced to 100 for the CC. A look-up-table procedure was finally employed to assign the results of the simulations to the respective pixels after the assimilation.

3. Results

3.1. LAI and CC estimation and validation

LAI and CC were estimated using the ANN algorithm described in section 2.2.2 for all the images acquired by HJ-1A, HJ-1B and Landsat 8 (Table 2) in the 3 years (from 2013 to 2015). Some acquisition dates of the images were near, i.e. within 1 to 4 days, the dates when ground

measurements were performed. For these images, the ground sampling points were identified according to their GPS coordinates, and measured and estimated LAI and CC were compared.

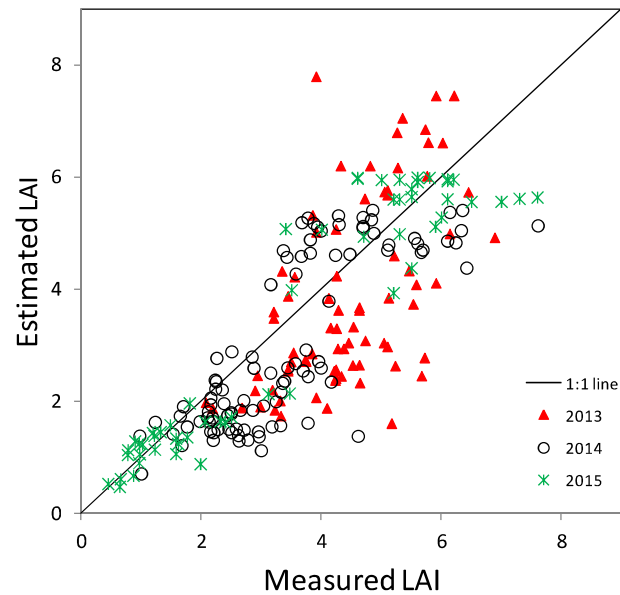


Figure 4. Validation results for the retrieval of LAI from HJ1A, HJ1B and Landsat 8 images, using field measurements of 3 years in Yangling rural area.

Figure 4 shows the comparison between LAI estimated from the satellite images employing the ANN algorithm and LAI measured during the field campaigns. The overall Relative Root Mean Square Error (RRMSE) for LAI for the three years was 30%. RMSE and RRMSE for the three years examined, both for CC and LAI, are shown in detail in Table 6. The estimation error is larger for higher values of LAI, in fact for LAI between 0 and 2 the difference between estimates and measures is small (Figure 4), while particularly for LAI values between 3 and 4 the error tends to be higher. These results indicate that the algorithm tends to underestimate LAI, especially for 2013, for which the results were worse than for other years (Table 6).

Table 6. Statistics resulting from the estimation of LAI and canopy cover (CC) from Landsat and HJ1 images assessed against ground measurements

		RMSE	RRMSE
Dataset		(t ha ⁻¹)	(%)
LAI	2013	1.48	33
	2014	1.03	30
	2015	0.75	21
	All years	1.15	30
CC	2013	7.27	8
	2014	8.05	10
	2015	5.06	7
	All years	7.17	9

Figure 5 shows the results obtained for CC estimation. In this case the overall RRMSE is 9%, i.e. lower than for LAI (Table 6). Saturation affects less the estimation of CC, even if the ANN algorithm tends also in this case to underestimate this biophysical variable. The results obtained for 2015 present the smallest RRMSE, while the results for 2013 and especially 2014 provide a much worse estimation.

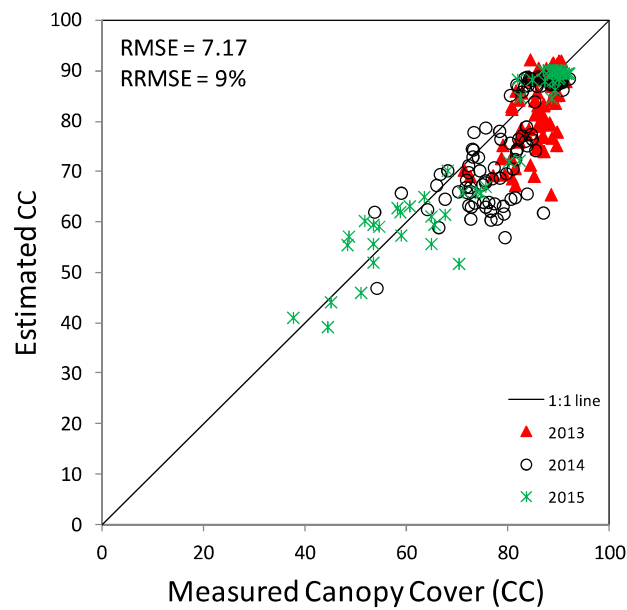


Figure 5. Validation results for the retrieval of canopy cover (CC) from HJ1A, HJ1B and Landsat 8 images, using field measurements of 3 years in the Yangling rural area. The field measured LAI was converted into CC using equation (1).

3.2. Assimilation results: field scale application

EnKF for SAFY and PSO for Aquacrop were tested at the field scale, validating the results with the ground yield measurements. Figure 6a shows the comparison between measured and simulated yield, resulting from the application of the EnKF method to the SAFY model, for which a unique preliminary calibration of model parameters had been done, as described in section 2.2.3. An overall RRMSE of 18% was obtained for SAFY, with values ranging from 15% to 20% in the different years (Table 7). The RRMSE for EnKF with the SAFY model was 18%, while for PSO with the Aquacrop model it was 24%, for the data of all years pooled together. The performance of Aquacrop was particularly poor, as compared to SAFY, for 2013 data (Table 7).

Table 7: Statistics resulting from the estimation of yield using 2 different assimilation strategies: the EnKF for SAFY and the PSO for Aquacrop (only the case of yearly calibration results are shown).

		RMSE	RRMSE
		(t ha ⁻¹)	(%)
EnKF	2013	1.23	20
	2014	1.19	17
	2015	0.78	15
	All years	1.09	18
PSO	2013	1.24	36
	2014	1.19	18
	2015	0.72	12
	All years	1.12	24

Figure 6b and Figure 6c both show the comparison between measured and simulated yields resulting from the application of the PSO assimilation method to the Aquacrop model. In Figure 6b the simulations were preceded by a single calibration of the model (using two data points from 2014), while in Figure 6c the model had been previously calibrated for each year considered.

The application of PSO with the Aquacrop model with only one calibration, showed a rather limited range of simulated yield values. i.e. between 4 and 6 t·ha⁻¹, with a RRMSE of 0.27, whereas when a calibration was carried out for each year, the range of simulated yield variation was wider, i.e. between 3 and 7 t·ha⁻¹, with a small decrease of RRMSE. In the case of a single calibration no correlation was found ($r=-1.04$, n.s.) between simulated and measured values, whereas when yearly calibrations were performed a low, albeit statistically significant, correlation ($r=0.34$, $P<0.01$) was found between measurements and estimates. For SAFY the correlation ($r=0.53$, $P<0.01$) was higher.

Overall, for the scenarios analyzed, the assimilation using the EnKF method with the SAFY model estimated the yield similarly or better than the PSO with Aquacrop. The computational time required by the two methods is very different, though. Using a PC with an Intel Core i7-4770 CPU at 3.4 GHz, the run time of PSO with Aquacrop was around 40 minutes for each simulated point, while for EnKF with SAFY it was 30 seconds for each point.

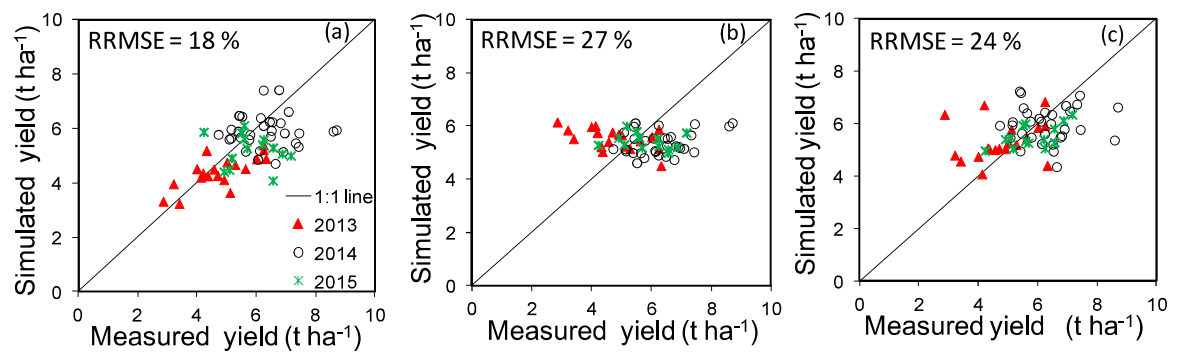


Figure 6. Comparison between measured and simulated wheat yield using a) the EnKF method with the SAFY model after a unique general calibration for all years; b) the PSO method with the Aquacrop model after a unique general calibration for all years; and c) the PSO method with the Aquacrop model after specific calibrations performed for each year.

3.3. Assimilation results: spatialized application

The district scale yield estimation was limited to part of the Yangling administrative area for which it was possible to obtain official data about the production per hectare, the total production and the surface area sown with winter wheat. The district scale application was limited to fields classified as winter wheat in the different years as described in section 2.2.5. From Fig. 7 and 8, it is apparent that a large part of the area in which winter wheat was cultivated in 2013 was occupied by other crops in 2014, as a result of the crop rotation practices adopted in the area.

Figure 7 shows the winter wheat yield maps estimated for 2013 and 2014 using the EnKF method with the SAFY model.

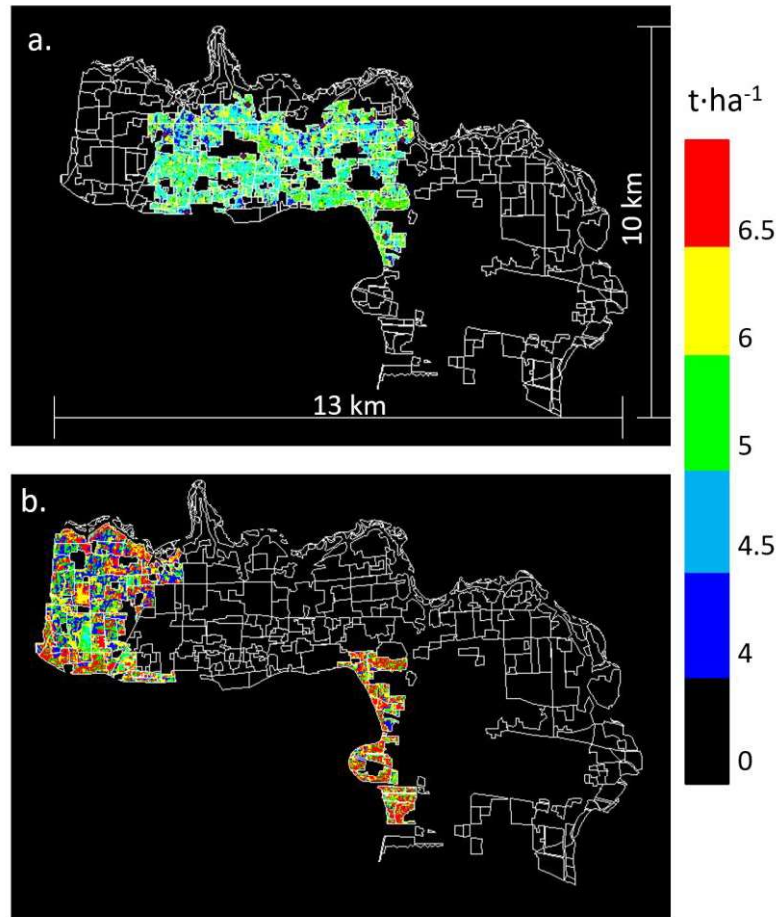


Figure 7. Wheat yield map ($t \cdot ha^{-1}$) for Yangling estimated using the EnKF assimilation method with the SAFY model for 2013 (a) and 2014 (b).

The official statistical data (provided by Shaanxi Provincial Bureau of Statistics) for 2013 reported an average yield of $5.5 t \cdot ha^{-1}$ for the area and a total production of $1.09 \cdot 10^4 t$ for the whole wheat surface area, estimated at 1980 ha. The method used for the classification (section 2.2.5) of wheat cultivation area, exploiting also the knowledge of agricultural field boundaries, provided an estimate of wheat sowing area of 1983 ha, i.e. very accurate. The average yield per unit surface estimated using the EnKF method with the SAFY model, resulted to be $4.98 t \cdot ha^{-1}$, and the total production for the Yangling administrative area was of $0.99 \cdot 10^4 t$, i.e. about a 9.4% less than the official statistics yield. For 2014, the official statistical data provided by the National Bureau of Statistics of the People's Republic of China reported a mean yield of $6.13 t \cdot ha^{-1}$, for a total production of $0.93 \cdot 10^4 t$ in the whole wheat cultivation area, estimated at 1470 ha. For the 2014 the wheat area classification results were less accurate as for the 2013: the wheat sowing area estimated was of 1564 ha, overestimated by about 7% compared to official statistical data. The estimated mean yield was of $5.67 t \cdot ha^{-1}$, and the total production estimated was $0.82 \cdot 10^4 t$, so for 2014 there was an underestimation relative to the official statistics of about 12.2%. For the

same reference wheat sowing area, the yield was also estimated using the PSO assimilation method with the Aquacrop model (Figure 8).

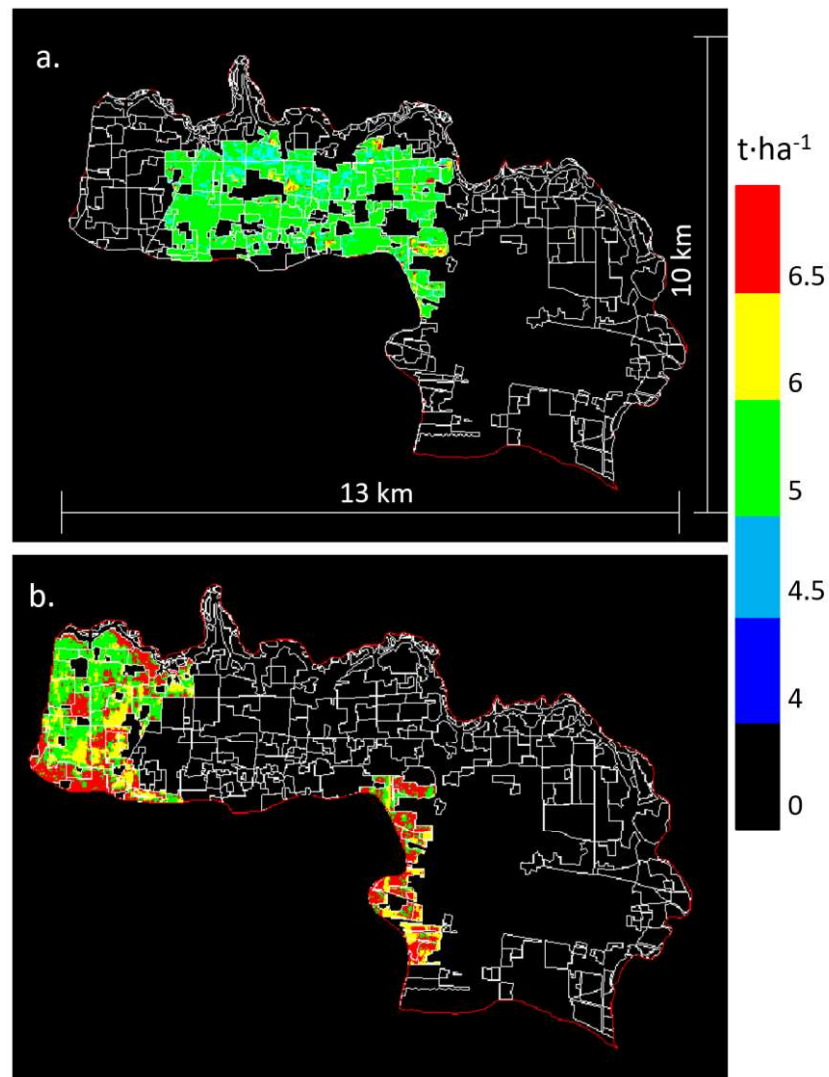


Figure 8. Wheat yield map ($t \cdot ha^{-1}$) for Yangling, estimated using the PSO assimilation method with the Aquacrop model for 2013 (a) and 2014 (b).

All the estimates obtained with Aquacrop shown in Figure 8 were obtained with the model calibrated for each year before the application of the assimilation algorithm. Using the PSO with Aquacrop, the average yield per unit on the studied area was of $5.56 t \cdot ha^{-1}$, very close to the official statistics value, whereas the total production estimated was of $1.10 \cdot 10^4 t$, with an estimation error of about 118 t. For 2014 the same procedure was applied, and the yield per unit area estimated with PSO and Aquacrop was of about $5.97 \cdot 10^4 t$, while the total production estimated for this year was of $0.93 \cdot 10^4 t$. Figure 9 summarizes the comparison between assimilation results and official statistics for the Yangling administrative area.

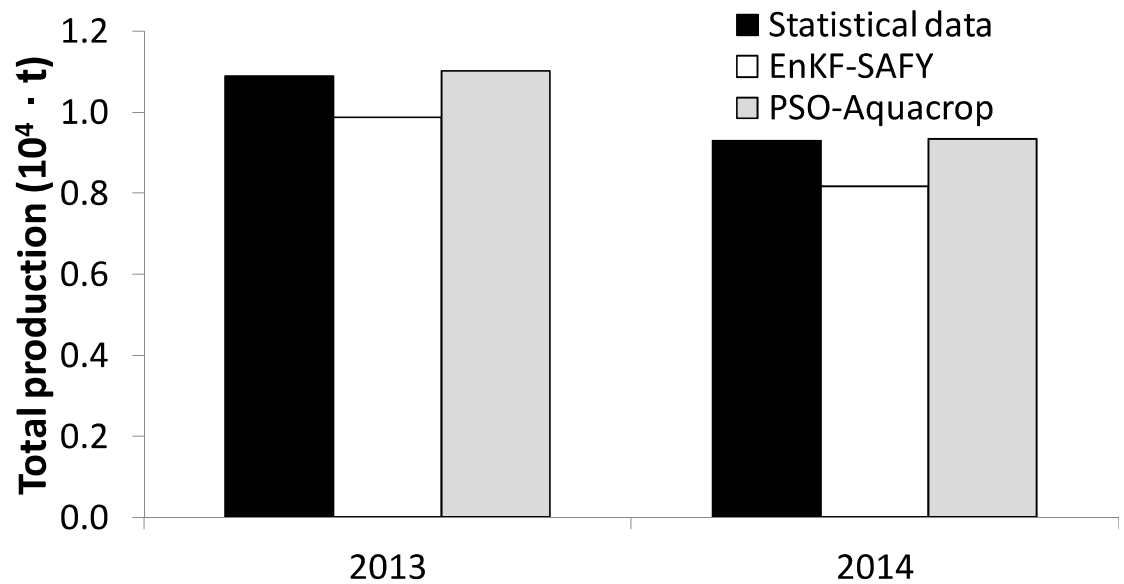


Figure 9. Comparison of wheat production estimated by official statistical surveys (black), and from assimilation using EnKF with SAFY (white) and PSO with Aquacrop (grey).

4. Discussion

The first step of this work was the retrieval, from remote sensing data, of the biophysical variables, to be used in the assimilation procedures. The biophysical variables considered were LAI and CC, both estimated using the ANN algorithm as described in section 2.2.2. The results shown in Figure 4 indicate that overall, the algorithm tended to underestimate the value of LAI. The comparison between estimated values and ground measurements highlighted that, with the exception of small values (between 0 and 2), the points underestimated were more than those overestimated, particularly for values of LAI between 2 and 4, range in which the error was higher than the average RRMSE. This error distribution was partly due to the characteristics of ANN algorithm, which tended to worsen the estimation of winter wheat LAI with increasing values, because of the well known issue of saturation of the canopy reflectance [19]. Using HJ-1 data, [56] obtained better results than ours, for a larger ground dataset collected in the same area of Yangling in 2014 ($n=80$), with RRMSE values as low as 21% for optical vegetation indices. However, it should be noted that the methods used by these authors are based on empirical calibrations, with a prevailing local validity, whereas, in our case, the ANN method is of general application not requiring preliminary calibrations [19–20]. [57] assessed the performance of the same ANN-based algorithm used in the present study, with SPOT4 HRVIR and Landsat 8 data, obtaining a RMSE of 0.74 for winter wheat, better than the RMSE values found in our study (Table 6) for 2013 and 2014, respectively of 1.48 and 1.03, but similar to the one we obtained in 2015, i.e. 0.75. It should be noted that in the dataset used by [57], only 2 data points had LAI values higher than 3 and the maximum LAI was 4.5. Conversely, in our dataset, many points had LAI values well above 3, reaching up to a maximum value of almost 8 (Figure 4).

A considerable impact on the LAI estimation accuracy might have been brought about also by the quality of the remote sensing data. The region of interest is strongly characterized by the presence of clouds and haze, so the probability to acquire clear images is very low, especially during spring, when rainfalls are frequent. The atmospheric correction for these images was difficult and it strongly influenced the reflectance and consequently the estimation of LAI. With such restrictive atmospheric conditions, it is not surprising that radar data can provide better estimates of LAI than optical data [56]. Indeed [56], using empirical calibrations, achieved better results when they included as regressors, in the estimation of LAI, radar polarimetric parameters obtained from RADARSAT-2, alone or coupled with vegetation indices from HJ-1. In the latter case they achieved a RRMSE of 17.4% when using Partial Least Squares Regression.

The consideration about the influence of the atmosphere on the quality of input data for the Yanling area is also true for CC, but in this case the RRMSE between RS estimation and ground measurements was smaller than for LAI, with a RMSE of 7.2% and a RRMSE of 9% for the three years. This is due to the fact that CC is less subject to a saturation at high values than LAI. The observation period was between March and May, when the variation of LAI was between 0.5 and 7, corresponding to a variation of CC between 40% and 91%. Already by the end of March, LAI was higher than 3, which corresponded to a CC of 80%, so for most measurements dates CC varied between 80% and 91%. In such a small range of values, the ANN algorithm could estimate CC changes with a higher accuracy. In this case our results were better than those achieved by [57], who obtained a RMSE of 19% for winter wheat, whereas in our study RMSE values ranged between 5.1% and 7.3% in the three years. Our results are also better than those of [33], who achieved a RMSE of 15.8% and an RRMSE of 19.6% when using only HJ-1 data. These authors obtained better results, comparable to ours, when they used, in empirical regression models, coupled optical and radar data as regressors, reaching a RMSE of 8.8% and a RRMSE of 10.9%, confirming the above considerations on atmospheric effects. Again, it should be remembered that we used a general purpose model-based algorithm, i.e. not requiring a local calibration, unlike [33]. Part of the differences in the accuracy of estimation of CC, could also be attributed to the equations used to convert ground measurement of LAI into CC. We used an equation proposed by [40] after a specific study on wheat, who showed that it performed better than the equation proposed by [58] for maize, which was also employed by [33].

The estimation of LAI and CC was aimed at the application of assimilation methods for estimating yield, thus the error in their estimate could influence the results of the assimilation. In particular, the EnKF method takes explicitly into account the error in the observations to be assimilated into the model, which was set in our case to a value of 30%, in agreement with the RRMSE actually obtained for LAI (Figure 4). This sort of error is expected to have a remarkable effect on the accuracy of the assimilation results [2].

SAFY and Aquacrop are two growth crop models with rather different characteristics. The former is a simple algorithm, developed specifically for remote sensing applications, with a reduced number of influential parameters and strongly dependent on the scenario characteristics [49]. The latter is more complex than SAFY, the number of influential parameters is higher and the influence of scenario characteristics is lower than for SAFY. The strength of Aquacrop is its capability of taking into account crop physiological processes related to water

stress, which are ignored by SAFY, but of great interest for the estimation of crop water requirements and yield response to drought from remote sensing [7, 31]. However, despite being less complex than other crop models such as e.g. EPIC [22] or CERES [26], the larger number of parameters of Aquacrop makes it harder to calibrate and slower to run than SAFY, therefore making it less suitable for an assimilation application.

The application of two different assimilation algorithms to the two models, EnKF on SAFY and PSO on Aquacrop, allowed to obtain reasonable yield estimates both at the field and at the district scale. In the field application, the accuracy of SAFY with the application of EnKF was similar or higher than for Aquacrop with PSO. The estimated yields obtained with EnKF and SAFY varied in a range between 3 and 8 t·ha⁻¹, while for PSO with Aquacrop they varied between 4 and 7 t·ha⁻¹. The first method estimated with a good approximation low yield values, but the estimation accuracy decreased with the increasing yield and above 6 t·ha⁻¹ an underestimation was apparent. Instead, the second method overestimated values of yield lower than 4 t·ha⁻¹ and underestimated values over 6 t·ha⁻¹. Both methods showed the tendency to saturate at values higher than 6.5 t·ha⁻¹.

Both methods were initially applied by using a single preliminary calibration for the entire dataset, i.e. the three different winter wheat growth cycles. However, in such case, the yields estimated from the PSO with Aquacrop had a rather small range of variation (Figure 6b), looking rather insensitive to scenario changes. A different calibration for each climate dataset was necessary (Figure 6c), in order to have comparable results with those obtained applying the EnKF to SAFY, the latter being calibrated only once for all the years. [34] used the PSO method for the assimilation of crop biomass, estimated from field based spectral reflectance measurements, into the Aquacrop model, employing a quite extensive winter wheat dataset collected during 4 years. During the PSO assimilation they recalibrated only three parameters (canopy growth coefficient, maximum canopy cover and maximum effective rooting depth), but no information is provided on how the other parameters were set. They achieved a RMSE of 0.65 t ha⁻¹ for the whole validation dataset.

[35] used the same field dataset and methodology as [34], but during the PSO assimilation they recalibrated 8 Aquacrop parameters: cgc (increase of canopy cover per GDD), ccx (maximum canopy cover), cdc (decrease of canopy cover per GDD), eme (GDD from sowing to emergence), num (n of plants ha⁻¹), psen (soil water depletion factor for canopy senescence), pstoshp (shape factor for water stress limiting stomatal conductance), and rootdep (maximum effective rooting depth). Also in this case no information was provided on the other parameters and input variables. They achieved RMSE values for yield estimates ranging from 0.51 to 0.61 t ha⁻¹ for the different years and RRMSE from 8.8 to 10.7%, i.e. much better than our RRMSE value of 24%. However these results were obtained using ground-based hyperspectral measurements. [33] used HJ-1 data, acquired in 2014 in Yangling, for estimation of biomass or canopy cover, which was then used for the assimilation into the Aquacrop model using the PSO method. In this case, the yield estimation accuracy was lower, with a RMSE of 1.24 t ha⁻¹ and a RRMSE of 26%, comparable with our results. Since, for these authors, the estimation of biomass was more accurate than that of CC, because of the above mentioned issues, they obtained better results when they used biomass for the assimilation, with a RMSE of 0.92 t ha⁻¹ and a RRMSE of 19.4%. Even better results were achieved by [33] when they combined optical and radar data for

the estimation of the biophysical variables to be used for the assimilation, reaching a RMSE of 0.81 t ha^{-1} and a RRMSE of 17%. This confirms that the use of radar data could partially overcome the above mentioned difficulties with the atmospheric effects on optical data for the Yangling area.

In the regional scale application, both methods simulated the mean and total yield of the Yangling administrative area with a good approximation, as compared to official statistics. The wheat classification procedure was quite successful and it showed the spatial distribution of growing areas, interestingly revealing the effect of year-to-year variation due to crop rotation practices. Yield estimation results were generally consistent with those obtained with the field scale application. For PSO with Aquacrop the accuracy in the spatialized application was higher than expected from field scale results. It should be noted that for this method it was necessary to recalibrate the fixed parameters and the ranges of parameters optimized during the assimilation, for each observed crop growth cycle. For EnKF with SAFY conversely, the same calibration was used for all the analyzed crop growth cycles. Because of the computational limits of Aquacrop, the classes in which the pixels were binned, were reduced from about 1800 used in EnKF-SAFY to 100 for PSO-Aquacrop. This reduction was one of the causes of the lower spatial variation of yield for Aquacrop (Figure 8) as compared to SAFY (Figure 7). Another cause was the low variation of CC, the input data for PSO-Aquacrop, which affected the output.

Despite the error in the input data, found for both LAI and CC, and the limited number of remote sensing observation dates, the estimation of yield was reasonable. For EnKF with SAFY this depended on the characteristics of the Kalman Filter. This method, in fact, “filtered” the error on the measurements, introduced as a random effect, reducing its influence on the output. The effect of the error can be also reduced by increasing the number of ensemble, but this inevitably increases the computational time. For this method the choice of measurements error variance and the number of ensembles is fundamental [51].

For PSO with Aquacrop, the input error of CC was lower than for SAFY, but the yield estimation accuracy seemed to be more affected by the calibration of the model for each climatic scenario. This method thus resulted to be strongly dependant on the initial calibration, which resulted rather complex because of the high number of parameters which described the model. This dependence from calibration combined with the inaccessibility of the Aquacrop code and with the high computational cost, made the PSO method with Aquacrop less suitable for an operational application as an assimilation algorithm.

The value of RRMSE obtained with the assimilation based on the EnKF with SAFY ranged, in the present study, between 15% and 20% among the years, with a value of 18% for the whole dataset.

[48], using a modified SAFY version including CO_2 fluxes, and recalibrating 7 parameters assimilating LAI retrieved from remote sensing, obtained RRMSE values between 12% and 24% for wheat yield estimation. It should be noted that a much larger number of observations of LAI from remote sensing acquisitions were available to [48] as compared to the present study.

In the literature [22–28] the RRMSE value found by applying different methods of assimilation in different models was reported to be between 5–16%, therefore quite lower than ours. But it is important to consider that some of these studies [24–25, 27] employed more than 20 observations of biophysical variables in the assimilations for each crop growth cycle, while

other studies [4, 26] used less images (around 10) at an higher resolution and with a lower error on the biophysical variables estimations. The method tested in this study, instead, assimilated a reduced number of images (just 3 or 4) for each crop growth cycle and with an estimation error of about 20%, reaching levels of accuracy in the estimation of the yield comparable with the studies encountered in the literature. [2] showed that the frequency of available observations, but also their distribution along the crop growth cycle has an important influence on the accuracy of model estimation, when used for model forcing.

5. Conclusions

This study demonstrated the possibility to estimate the winter wheat yield, at field and district level scale, through the use of assimilation methods. Two different approaches were tested, one based on an updating assimilation method, the EnKF, employed for the assimilation of LAI into SAFY, and another based on a calibration assimilation method, the PSO, used to assimilate the CC into Aquacrop. The tests at the field scale showed the feasibility of using medium resolution (30 m GSD) satellite data, such as from HJ1A/B or Landsat 8 OLI images, to estimate yield, with potential applications for precision agriculture.

However, the results obtained with Aquacrop were less accurate as compared to other methods of assimilation for calibration encountered in the literature [22-23, 28]. This is mainly due to the characteristics of Aquacrop, for which is indispensable to carry out a very accurate calibration. Nevertheless, the application at the district scale of PSO assimilation with Aquacrop, was in agreement with the yield provided by official statistics. This occurred because Aquacrop was recalibrated for each climate dataset, furthermore to meet the high computational requirements of this assimilation method for calibration, a coarser binning of pixels, according to the CC temporal series, was applied, forcing a reduced number of classes. In this way the range of variation of simulated yields was rather limited, leading to accurate results, but less responding to spatial or temporal variation of yields. The high computational cost, the difficult calibration and the need to recalibrate for each year of climate datasets makes, therefore, the PSO method applied to Aquacrop less efficient.

The performance of the EnKF with the SAFY model, was comparable with the results of others upgrading assimilation method encountered in literature [24-28], which is especially encouraging when considering that, despite the limited number of remote sensing images, i.e. 3 or 4, with an error of LAI estimation of 30%, the error on the yield estimation was around 18%. This allows the application of this upgraded assimilation method in areas of the world where it is not possible to obtain a large number of images for each crop cycle.

The encouraging results of the present work obtained with SAFY, should be confirmed with further studies with other validations, possibly with experiments that provide more frequent field measurements, a larger variety of climatic and environmental datasets and a higher quality of remote sensing images collection.

Acknowledgments: the authors acknowledge with gratitude funding from the European Space Agency (ESA) and the Ministry of Science and Technology (MOST) of the People's Republic of China, in the context of the Dragon-3 Programme, project ID 10448 "Farmland Drought". P.C.S. was supported during his PhD work by co-funding from a Young Scientist Dragon 3 Grant from ESA and from a University of Tuscia PhD scholarship. P.C.S. is thankful to NERCITA for hosting him in China during 2014 and 2016. The authors are grateful to Fred Baret and Marie Weiss (INRA Avignon, France) for the provision of the Matlab code of the ANN algorithm and for their kind assistance; to CESBIO and Benoit Duchemin for providing the SAFY code and to Dirk Raes and FAO for providing the command-line version of Aquacrop.

Author Contributions: R.Casa and S.Pignatti conceived and designed the study methodologies and experiments; P.C.Silvestro, H.Yang, S.Pascucci and Z.Li performed the experiments; P.C.Silvestro and R.Casa analyzed the data; G.Yang and H.Yang contributed materials, data and analysis tools; P.C.Silvestro, R.Casa, S.Pignatti and W.Huang wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Mondal, P.; Basu, M.; Bhadoria, P. B. S. Critical Review of Precision Agriculture Technologies and Its Scope of Adoption in India. *Am. J. Exp. Agric.* **2011**, *1*, 49–68.
2. Casa, R.; Varella, H.; Buis, S.; Guérif, M.; Solan, B. De; Baret, F. Forcing a wheat crop model with LAI data to access agronomic variables : Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *Eur. J. Agron.* **2012**, *37*, 1–10.
3. Castaldi, F.; Casa, R.; Pelosi, F.; Yang, H. Influence of acquisition time and resolution on wheat yield estimation at the field scale from canopy biophysical variables retrieved from SPOT satellite data. *Int. J. Remote Sens.* **2015**, *36*, 2438–2459.
4. Dong, T.; Liu, J.; Qian, B.; Zhao, T.; Jing, Q.; Geng, X.; Wang, J.; Huffman, T.; Shang, J. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 63–74.
5. Pignatti, S.; Acito, N.; Amato, U.; Casa, R.; De Bonis, R.; Diani, M.; Laneve, G.; Matteoli, S.; Palombo, A.; Pascucci, S.; Romano, F.; Santini, F.; Simoniello, T.; Ananasso, C.; Zoffoli, S.; Corsini, G.; Cuomo, V. Development of algorithms and products for supporting the Italian hyperspectral PRISMA mission: The SAP4PRISMA project. International Geoscience and Remote Sensing Symposium (IGARSS) 2012, Munich; Germany; 22 - 27 July 2012, Article n. 6351620127-130. Mulla, D. J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371.
6. Casa, R.; Rossi, M.; Sappa, G.; Trotta, A. Assessing crop water demand by remote sensing and GIS for the Pontina Plain, Central Italy. *Water Resour. Manag.* **2009**, *23*, 1685–1712.
7. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981.
8. Kalpana, R.; Natarajan, S.; Mythili, S. Remote sensing for crop monitoring - a review. *Agric. Rev.* **2003**, *24*, 31–39.
9. Lobell, D. B. The use of satellite data for crop yield gap analysis. *F. Crop. Res.* **2013**, *143*, 56–64.
10. Dorigo, W. a.; Zurita-Milla, R.; de Wit, a. J. W.; Brazile, J.; Singh, R.; Schaepman, M. E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 165–193.
11. Huang, Z.; Turner, B. J.; Dury, S. J.; Wallis, I. R.; Foley, W. J. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.* **2004**, *93*, 18–29.
12. Atzberger, C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sens. Environ.* **2004**, *93*, 53–67. Bacour, C.; Baret, F.; Béal, D.; Weiss, M.; Pavageau, K. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens. Environ.* **2006**, *105*, 313–325.
13. Baret, F.; Pavageau, K.; Béal, D.; Weiss, M.; Berthelot, B.; Regner, P. Algorithm Theoretical Basis Document for MERIS Top of Atmosphere Land Products (TOA _ VEG). **2006**, 1–37.
14. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P. J.; Asner, G. P.; François, C.; Ustin, S. L. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66.

15. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91.
16. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.* **1984**, *16*, 125–141.
17. Baret, F.; Hagolle, O.; Geiger, B.; Bicheron, P.; Miras, B.; Huc, M.; Berthelot, B.; Niño, F.; Weiss, M.; Samain, O.; Roujean, J. L.; Leroy, M. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens. Environ.* **2007**, *110*, 275–286.
18. Verger, A.; Baret, F.; Camacho, F. Optimal modalities for radiative transfer- neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS / PROBA observations. *Remote Sens. Environ.* **2011**, *115*, 415–426.
19. Delécolle, R.; Maas, S. J.; Guérif, M.; Baret, F. Remote sensing and crop production models: present trends. *ISPRS J. Photogramm. Remote Sens.* **1992**, *47*, 145–161.
20. Doraiswamy, P. C.; Moulin, S.; Cook, P. W.; Stern, A. Crop Yield Assessment from Remote Sensing. *Photogramm. Eng. Remote Sensing* **2003**, *69*, 665–674.
21. Launay, M.; Guerif, M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric. Ecosyst. Environ.* **2005**, *111*, 321–339.
22. Wu, S.; Huang, J.; Liu, X.; Fan, J.; Ma, G.; Zou, J. Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield. *IFIP Adv. Inf. Commun. Technol.* **2012**, *370 AICT*, 410–418.
23. Ines, A. V. M.; Das, N. N.; Hansen, J. W.; Njoku, E. G. Remote Sensing of Environment Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* **2013**, *138*, 149–164.
24. Huang, Y.; Zhu, Y.; Li, W.; Cao, W.; Tian, Y. Assimilating Remotely Sensed Information with the Wheat Grow Model Based on the Ensemble Square Root Filter for Improving Regional Wheat Yield Forecasts. *Plant Prod. Sci.* **2013**, *16*, 352–364.
25. Jiang, Z.; Chen, Z.; Chen, J.; Liu, J.; Ren, J.; Li, Z.; Sun, L.; Li, H. Application of Crop Model Data Assimilation With a Particle Filter for Estimating Regional Winter Wheat Yields. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4422–4431.
26. De Wit, A.; Duveiller, G.; Defourny, P. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agric. For. Meteorol.* **2012**, *164*, 39–52.
27. van Diepen, C. A.; Wolf, J.; van Keulen, H.; Rappoldt, C. WOFOST: a simulation model of crop production. *Soil Use Manag.* **1989**, *5*, 16–24.
28. Confalonieri, R.; Bregaglio, S.; Acutis, M. Quantifying plasticity in simulation models. *Ecol. Modell.* **2012**, *225*, 159–166.
29. Casa, R.; Silvestro, P. C.; Yang, H.; Pignatti, S.; Pascucci, S.; Yang, G. Assimilation of remotely sensed canopy variables into crop models for an assessment of drought-related yield losses: A comparison of models of different complexity. In *IEEE 36th International Geoscience and Remote Sensing Symposium, IGARSS 2016*; Beijing, 2016; p. 5925-5928..
30. Steduto, P.; Hsiao, T. C.; Raes, D.; Fereres, E. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agron. J.* **2009**, *101*, 426. Jin, X.; Li, Z.; Yang, G.; Yang, H.; Feng, H.; Xu, X.; Wang, J.; Li, X.; Luo, J. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 24–37.
31. Jin, X.; Yang, G.; Li, Z.; Xu, X.; Wang, J.; Estimation of water productivity in winter wheat using the AquaCrop model with field hyperspectral data. *Precis. Agric.* (in press).
32. Jin, X.; Kumar, L.; Xu, X.; Yang, G.; Wang, J. Estimation of winter wheat biomass and yield by combining the Aquacrop model and field hyperspectral data. *Remote Sens.* **2016**, *8*, 972
33. Duchemin, B.; Maisongrande, P.; Boulet, G.; Benhadj, I. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environ. Model. Softw.* **2008**, *23*, 876–892.
34. Eberhart, R.; Kennedy, J. A New Optimizer Using Particle Swarm Theory. *Micro Mach. Hum. Sci.* **1995**, *MHS '95., Proc. Sixth Int. Symp.* **1995**, 39–43.

35. Raes, D. The ETo Calculator Table of Contents - Reference Manual. **2009**, 1–38.
36. Bristow, K. L.; Campbell, G. S. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric. For. Meteorol.* **1984**, *31*, 159–166.
37. Nielsen, D. C.; Miceli-Garcia, J. J.; Lyon, D. J. Canopy Cover and Leaf Area Index Relationships for Wheat, Triticale, and Corn. *Agron. J.* **2012**, *104*, 1569.
38. Zanter, K.. Landsat 8 (L8) Data Users Handbook. <http://www.webcitation.org/6mu9r7riR> (Archived by WebCite on 20 12 2016)
39. Zhao, K.; Xu, J.; Zhao, Z.; Song, L.; Xiao, K. Cross Comparison of HJ-1A/B CCD and Landsat TM/ETM+ Multispectral Measurements for NDVI, SAVI and EVI Vegetation Index. *Remote Sens. Technol. Appl.* **2013**, *28*, 674–680.
40. ENVI User's Guide. <http://www.webcitation.org/6muBGHxL0> (Archived by WebCite on 20 12 2016)
41. Raes, D.; Steduto, P.; Hsiao, T. C.; Fereres, E. AquaCropThe FAO Crop Model to Simulate Yield Response to Water: II. Main Algorithms and Software Description. *Agron. J.* **2009**, *101*, 438.
42. Monteith, J. L.; Moss, C. J. Climate and the Efficiency of Crop Production in Britain [and Discussion]. *Philos. Trans. R. Soc. B Biol. Sci.* **1977**, *281*, 277–294.
43. Allen, R. G.; Pereira, L. S.; Raes, D.; Smith, M. *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*; FAO - Food and Agriculture Organization of the United Nations: Rome, 1998.
44. Duchemin, B.; Boulet, G.; Maisongrande, P.; BenHadj I, Hadria R, Khabba S et al. Un modèle simplifié pour l'estimation du bilan hydrique et du rendement de cultures céréalières en milieu semi-aride. In: Proceedings of "Deuxième Congrès Méditerranéen Ressources en Eau dans le Bassin Méditerranéen - WATMED ", Marrakech, 14-17 November 2005, 14-17.
45. Veloso, A. Modélisation spatialisée de la production, des flux et des bilans de carbone et d'eau des cultures de blé à l'aide de données de télédétection : application au sud-ouest de la France. PhD Thesis, Université Toulouse III- Paul Sabatier, Toulouse, France, 23 June 2014
46. Silvestro P.C., Pignatti S., Yang H., Yang G., Pascucci S., Castaldi F., Casa R. Sensitivity analysis of the Aquacrop and SAFY crop models for the assessment of water limited winter wheat yield in regional scale applications . *PLoS One* (under review).
47. Jin, X.; Feng, H.; Zhu, X.; Li, Z.; Song, S.; Song, X.; Yang, G.-J.; Xu, X.; Guo, W. Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PLoS One* **2014**, *9*, e8693848.
48. Burgers, G.; van Leeuwen, P. J.; Evensen, G. Analysis Scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.* **1998**, *126*, 1719–1724.
49. Makowski, D.; Guérif, M.; Jones, J. W.; Graham, W. Data assimilation with crop models. In *Working With Dynamic Crop Models*; Wallach, D.; Makowski, D.; Jones, J. W., Eds.; Elsevier, 2006; pp. 151–170.
50. Li, Z.; Jin, X.; Zhao, C.; Wang, J.; Xu, X. Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *Eur. J. Agron.* **2015**, *71*, 53–62.
51. Poli, R.; Kennedy, J.; Blackwell, T. Particle swarm optimization. *Swarm Intell. US*, **2007**, 33–57.
52. Lin, W.C.; Yin, Y.; Cheng, S.R.; Cheng, T.C.E.; Wu, C.H.; Wu, C.C. Particle swarm optimization and opposite-based particle swarm optimization for two-agent multi-facility customer order scheduling with ready times. *Appl. Soft. Comput.*, **2017**, *52*, 877–884.
53. Jin, X.; Yang, G.; Xu, X.; Yang, H.; Feng, H.; Li, Z.; Shen, J.; Zhao, C.; Lan, Y. Combined Multi-Temporal Optical and Radar Parameters for Estimating LAI and Biomass in Winter Wheat Using HJ and RADARSAR-2 Data. *Remote Sens.* **2015**, *7*, 13251–13272.
54. Li, W.; Weiss, M.; Waldner, F.; Defourny, P.; Demarez, V.; Morin, D.; Hagolle, O.; Baret, F. 2015. A generic algorithm to estimate LAI, FAPAR and FCOVER variables from SPOT4_HRVIR and Landsat sensors: evaluation of the consistency and comparison with ground measurements. *Remote Sens.* **2015**, *7*, 15494–15516.

Hsiao, T.C.; Heng, L.; Steduto, P.; Roja-Lara, B.; Raes, D.; Fereres, E. AquaCrop - The FAO model to simulate yield response to water: parametrization and testing for maize. *Agron. J.*, **2009, *101*, 448–459.**



© 2016 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications

Paolo Cosmo Silvestro^{1¶}, Stefano Pignatti^{2&}, Hao Yang^{3&}, Guijun Yang^{3&}, Simone Pascucci^{2&}, Fabio Castaldi^{1&}, Raffaele Casa^{1¶*}

¹Department of Agriculture Forestry and Nature (DAFNE), University of Tuscia, Viterbo, Italy

²Institute of Methodologies for Environmental Analysis (IMAA), Consiglio Nazionale delle Ricerche (CNR), Tito Scalo (Potenza), Italy

³National Engineering Research Center for Information Technology in Agriculture (NERCITA), Beijing, China

*Corresponding author:

E-mail: rcasa@unitus.it (RC)

¶These authors contributed equally to this work: conceptualisation, methodology, software, data curation, writing-original draft preparation, visualisation and supervision, writing-review and editing

& These authors contributed equally to this work: visualisation and supervision, resources, project administration and funding acquisition.

Abstract

Process-based models can be usefully employed for the assessment of field and regional-scale impact of drought on crop yields. However, in many instances, especially when employing models at the regional scale, it is necessary to identify the parameters and input variables that most influence the output and to assess how their influence varies when climatic and environmental conditions change. In this work, two different crop models able to represent yield response to water, Aquacrop and SAFYE, were compared, with the aim to quantify their complexity and plasticity through Global Sensitivity Analysis (GSA), using Morris and EFAST (Extended Fourier Amplitude Sensitivity Test) techniques, for moderate to strong water limited climate scenarios. Although the rankings of the sensitivity indices for the different factors was clearly influenced by the scenarios used, the correlation among the rankings, higher for SAFYE than for Aquacrop, assessed by the top-down correlation coefficient (TDCC), revealed clear patterns. This allowed to identify the most relevant factors and those negligible for both models. In general parameters and input variables related to phenology and to water stress physiological processes were found to be the most influential for Aquacrop. For SAFYE, it was found that the water stress could be inferred indirectly from the processes regulating leaf growth, described in the original SAFY model. It was found that SAFYE has a lower complexity and plasticity than Aquacrop, making it more suitable to less data demanding regional scale applications, in case the only objective is the assessment of crop yield and no detailed information is sought on the mechanisms of the stress factors affecting its limitations.

Keywords: Morris; EFAST; global sensitivity analysis; SAFY; complexity; plasticity.

Introduction

Drought is one of factors which most influences agricultural production worldwide, causing significant and occasionally dramatic harvest losses [1]. In addition to the areas typically affected by low or erratic rainfall, areas experiencing water deficits have increased due to the growing water demand for non-agricultural uses and because of climate change [2]. There is a growing need to expand currently available techniques to monitor the impact of drought on crops, e.g. by predicting the yield loss in the presence of water shortages, both at the field and at the regional scale. First requirement herewith, is a crop model suitable to study the impact of drought, i.e. one that can correctly simulate physiological processes related to the soil water status, such as crop transpiration and yield responses to water. Many models are available for this purpose, with varying degree of complexity and predictive performance. For example, the Aquacrop model was developed by the Food and Agriculture Organization of the United Nations specifically for the purpose of assessing crop response to water and it is increasingly used by scientists and agronomists [3-14]. In a comparison study of different wheat crop models [15], Aquacrop was found to be the most sensitive to water stress and the simplest when compared to other models, with comparable accuracy. Aquacrop was shown to simulate correctly winter wheat yield under non-limiting [10] and water deficit [7-8, 11-12] conditions, generally providing better results in wet or moderate water stress years than in very dry years [7], thus requiring additional investigations before applying it to severe drought conditions. After appropriate calibration, Aquacrop could achieve low root mean square errors (RMSE) between measured and estimated yields, in the order of 0.27 [12] to 1.29 [8] t ha⁻¹.

Although most process-based crop models like Aquacrop were originally conceived for field scale applications, they are increasingly employed for spatialized regional scale studies [11, 14]. Crop models are usually quite demanding in terms of data requirements and when applied to large areas, they can be affected by many sources of uncertainty, due to the poor quality of input data (e.g. weather, soil), to the lack of information on management (e.g., sowing dates, fertilization practices, cultivars grown), which can be also variable in space and time, as well as to the model structure [16, 17], to the experience of the model users [18] and to the uncertainty in the data used for their calibration [19-20].

Because of these constraints, there is a growing interest, for large scale applications, in the assimilation of remote sensing data into dynamic process-based crop models. The objective of data assimilation is to characterize the state of an agroecosystem by combining information from various sources such as models and observations in space and time of the variables of interest [21]. This is a promising methodology that can be used to improve regional scale model simulations, reducing model errors, by ingesting into the process biophysical variables observed during the running crop season. This makes the methodology apt for the assessment, for instance, of the impact of drought on crop yields, suitable for large scale high spatial and temporal resolution spatialized applications [14, 16, 21-22]. For example Ines et al. [22] developed a procedure to assimilate data (remotely sensed soil moisture and leaf area index), into a crop model to improve the estimation of crop yield and growth. By applying the data assimilation-crop modeling framework, the authors claim to improve yield prediction performance substantially, which can contribute to prevent actual yield losses due to water stress. Other successful applications on wheat were also provided, for example by [13-14] using the Aquacrop model.

Various methods have been developed to integrate remotely sensed data into crop models, according to two main strategies: calibration and updating [21].

Calibration methods encompass adjustment of parameters in order to obtain an optimal agreement between simulated and observed state variables. Here, model parameters or input variables are re-calibrated at each assimilation step. The other assimilation strategy consists of updating model state variables and parameters, whenever an observation is available. Because of the high number of parameters and inputs in most crop models, assimilation methods require the set of calibrated, or updated, parameters or inputs to be a subset of the total number of those included in the model, in order to simplify the process and reduce the computational cost of the algorithm.

To solve this issue, it is possible to follow two different strategies: 1) use a simpler model with a reduced number of parameters and inputs, or 2) fix the values of the parameters and input variables which are less influential, in order to reduce as much as possible the number of factors to vary. In all circumstances, however, a preliminary identification of the factors that most affect the targeted outputs of the model is a requirement.

The use of a simple model, with a reduced number of parameters, was proposed, for example, by Duchemin et al. [23], who developed the Simple Algorithm For Yield (SAFY), in order to investigate the perspectives offered by coupling a simple vegetation growth model and ground-based remotely-sensed data for the monitoring of wheat production. The model simulates dynamically leaf area index (LAI) and dry above-ground biomass and crop yield. Being quite simple, it is very attractive for operational applications at a regional scale. It was employed by Chahbi et al. [24] to estimate the dynamics and yields of cereals in semi-arid, low productivity regions in North Africa, confirming the ability of the model for yield prediction. However, the model does not take into account the effects of water or nutrient limitations on plant growth and it does not simulate physiological processes related to the soil water status, such as crop transpiration and yield responses to water. The impact of water deficit is expected to be accounted for by the variation of the effective light-use efficiency, with the idea that this parameter can be calibrated from the time course of the LAI observed from remote sensing. [23] highlights the limits of this assumption, since LAI is an indicator of all agro-environmental stresses considered together, with no possibility to scrutinize the underlying causes for yield reductions. For these reason Duchemin et al. [25] and Veloso [26] added a water balance and evapotranspiration component to the model, thus renamed SAFYE.

Alternatively, more complex models can be used, such as Aquacrop. The latter is quite interesting in the context of remote sensing data assimilation, because it employs canopy cover (CC) as a key state variable. CC is easier to retrieve from remote sensing than e.g. LAI, which is difficult to estimate at high values, being subject to saturation of the reflectance signal [27].

The identification of parameters and input variables that most affect the output of a model is a fundamental problem for all the applications of crop models whenever large uncertainty on their values is expected, such as at the regional scale, and when assimilation algorithms are employed. In general, this is also important for model calibration studies, since accurate calibration of a minimum number of parameters is an requirement for all crop models.

A suitable technique for such purpose is sensitivity analysis (SA). The aim of a sensitivity analysis is to determine how sensitive the outputs of a crop model are, with respect to the elements of the model which are subject to uncertainty or variability, i.e. typically input variables and parameters [16]. It is possible to distinguish two different strategies: local and global SA [28]. Local Sensitivity Analysis (LSA) is focused on

the local impact of the factors on the model outputs and is carried out by computing partial derivatives of the output variables with respect to the input factors [16]. Each input parameter is varied in a neighborhood of a nominal value in order to calculate numerically the derivatives. During the variation of a single parameter, all others are held constant at a value representative of the situation to be simulated. This one-at-a-time (OAT) approach is suitable when the variation around the nominal value is small enough, so that the relationship between inputs and outputs can be considered as linear. LSA methods are characterized by a small computational cost, but they are inadequate to evaluate the effect of simultaneous variation of all input factors on the model output in case of non-linear biophysical relationships [28].

An extensive overview of SA methods was done by Cariboni et al. [29], who show the inadequacy of local methods for crop models, because of the complexity of the latter and the necessity to know the interaction between parameters. For these reasons GSA is considered more adequate in this context. Global Sensitivity Analysis (GSA) methods evaluate the effect that the simultaneous change of several or all the input factors have on the output of the model in wide ranges of variation [28]. The estimate of sensitivity of the model to each parameter is obtained by varying at the same time many or all the input factors, measuring the combined effect on model outputs. The inconvenience of this method is the high computational cost, and in case of an excessive number of factors, the difficulty of convergence of the algorithm.

It is possible to distinguish three different groups of GSA: 1) variance-based methods, such as SOBOL [30], Fourier Amplitude Sensitivity Test (FAST) [31], Extended FAST (EFAST) [32]; 2) screening methods, e.g. Morris [33]; 3) regression-based methods, e.g. Latin Hyper Cube Sampling, Random and Quasi-Random LpTau [34]. Confalonieri et al. [34] performed a comparison of several SA techniques, applying them to the rice crop model WARM, demonstrating the agreement between rankings of crop parameters from the most to the last relevant. They evaluated the accuracy and the efficiency of each SA method, concluding that resource intensive SA methods might not be needed to identify the most relevant parameters. In fact the Morris method, the simplest amongst GSA methods assessed, produced results comparable to those obtained by more computationally expensive methods. If the model is described by a limited number of parameters or it is possible to exclude a priori some parameters, it becomes convenient to use a variance-based method. The most frequently used methods are Sobol and EFAST. The efficiency of these methods is similar, in terms of computational time, but EFAST highlights better than Sobol the influence of interactions between parameters on the variance model [32]. In the literature both are used to analyze the sensitivity of a limited number of parameters. Chen et al. [35] preferred to use Sobol indices while in other studies (e.g. [16], [32]) the EFAST method is preferred. Confalonieri [34] proposed a combination of Morris and Sobol methods, to exploit the computational advantages of Morris method to detect the non influential parameters and to reduce the number of parameters when using Sobol. Vanuytrecht et al. [9] used the same strategy, applying at first a screening technique using the Morris method and subsequently EFAST as variance-based method to examine sensitivity of the yield output of the Aquacrop model for different crops and climates. They showed that the sensitivity to important parameters depends strongly on environmental conditions and it was not possible to establish a ranking of parameters valid for each climatic scenario. This signifies that it would be necessary to perform a sensitivity analysis, possibly a simple screening method, for the scenarios in which the model will be used. Here, the plasticity of the model, i.e. the tendency to change its behavior and sensitivity under different conditions, is thus determining the need for an extent of SA. From the operational point of view, output sensitivity and plasticity of a

crop model are essential informative characteristics when one wants to assimilate remote sensing data into crop models at the regional scale.

In the present study, the sensitivity of two models, of great interest in the assessment of wheat yield in water limited conditions for regional scale studies, was assessed. These models are: Aquacrop [3], including a larger number of parameters and input factors, but usually achieving accurate estimates of wheat yield [7-14], and SAFY, having a reduced number of parameters, but generally providing less accurate yield estimates [23-26].

The objective of the present work is to study the complexity and plasticity of these two models in a comparative manner, more specifically focusing on the global sensitivity of the winter wheat yield to model parameters and input variables, in a wide range of water limited conditions occurring during the wheat growth season. A previous global sensitivity analysis was carried out for Aquacrop by Vanuytrecht et al. [9], but no assessment of its plasticity, i.e. of how variable the sensitivity of the model is to environmental changes, has been performed. This is a crucial aspect, relevant for all the users of the model in large scale spatialized applications. For SAFY no previous global sensitivity analysis has been reported yet.

Materials and Methods

Crop models used

Aquacrop

Aquacrop is a widely used "water driven" productivity model that simulates canopy cover (CC), biomass and yield of a crop mainly as a function of the water productivity, i.e. the biomass produced per unit of water transpired by the vegetation, normalized for atmospheric evaporative demand and air CO₂ concentration [3-14]. Input variables include weather data, consisting of daily maximum and minimum temperature, evapotranspiration and rainfall, as well as soil properties and agronomic management information, including some related to the crop (e.g. sowing date). In the version of Aquacrop used in the present work (version 4.0) there are 45 parameters (S1 Table), defining the crop physiological and developmental responses to environmental factors and to soil water and salinity stresses. Of all these parameters, 29 are considered as conservative, i.e. crop specific, but not changing with cultivar, time, management practices, geographic location or climate [38]. These parameters are not supposed to require a local calibration for a well calibrated crop such as wheat, but would need to be calibrated using data from multiple location for a species new to Aquacrop. Crop development is limited by upper and lower temperature thresholds and is determined by a set of phenological parameters (e.g. *eme*, *flo*, *mat*, *sen*) which specify the length (in days or in growing degree days) of each development phase, e.g. from sowing to emergence (*eme*), flowering (*flo*), maturity (*mat*) or senescence (*sen*).

Canopy cover (CC) is the main state variable describing the growth of the foliage. Its dynamics, in non limiting conditions, are regulated by parameters setting the initial canopy cover, its growth rate coefficient (CGC), the maximum CC potentially achievable (CCx) and its decline rate coefficient (CDC). Water stress, as well as soil salinity or limited fertility, limits or delays the CC development through stress coefficients.

The amount of water transpired by the crop is a function of CC . Cumulative biomass production is obtained as the sum of the daily ratio between crop transpiration and ET_0 for each day of the crop cycle. The proportional factor between biomass and transpiration ratio is the normalized Water Productivity (pwp) (Eq. 1):

$$B_n = pwp \cdot \sum_{i=1}^n \left(\frac{Tr_i}{ET_{0i}} \right) \quad (1)$$

where B_n is the cumulative aboveground biomass production at day n ($g\ m^{-2}$); Tr_i is the daily crop transpiration ($mm\ day^{-1}$); ET_{0i} is the daily reference evapotranspiration ($mm\ day^{-1}$); i is the day of crop cycle (from 1 to n) and pwp is the normalized crop water productivity.

The yield is calculated by multiplying the final biomass by a harvest index [38]. The influence of water stress on crop development is represented by a number of stress coefficients, which describe the impact of water stress on canopy development, transpiration and harvest index due to limited water availability. Additionally, temperature limitations on biomass production and pollination are considered. A more detailed description of the Aquacrop model is presented by Raes et al. [38].

SAFYE

The original Simple Algorithm For Yield (SAFY) proposed by [23] is based on the Monteith concept [39], which links the production of total dry phytomass to the photosynthetically active portion of solar radiation (PAR) absorbed by the crop. SAFY simulates, on a daily time step, three main state variables from emergence of the crop to the end of its senescence. These are: Dry Above-ground Mass (DAM), Leaf Area Index (LAI) and Grain Yield (GY).

The daily production of dry above-ground phytomass (ΔDAM) depends on the incoming global radiation (R_g) through three parameters: 1) climatic efficiency ($Pgro_R2P$) which is the ratio of incoming photosynthetically active to global radiation; 2) light-interception efficiency ($Pgro_Kex$), which affects the fraction of photosynthetically active radiation absorbed by the canopy (APAR) and 3) effective light-use efficiency $Pgro_Lue$ ($g\ MJ^{-1}$), which is the ratio of photosynthetically energy produced as DAM from APAR.

Furthermore, a temperature stress function $F_T(Ta)$ guarantees that the daily rate of biomass production of vegetation (ΔDAM) increases as the air temperature is closer to the optimum temperature ($Ptfn_Topt$), and goes to zero out of the range between minimum and maximum critical temperatures. The final equation of ΔDAM is:

$$\Delta DAM = R_g \cdot Pgro_R2P \cdot Pgro_Kex \cdot Pgro_Lue \times F_T(Ta) = APAR \times Pgro_Lue \times F_T(Ta) \quad (2)$$

The light-interception efficiency depends on the green leaf area index (GLAI) and a light interception coefficient $Pgro_Kex$ according to Beer's law. The development of GLAI is split into two phases. In the first phase GLAI increases, starting at crop emergence and ending at the beginning of senescence. In the second phase, between the beginning and the end of senescence, GLAI decreases with a rate defined by senescence parameters.

The third state variable simulated by SAFY is the grain yield. It is calculated from DAM by means of a proportionality factor $Pgro_P2G$, corresponding to the harvest index. A detailed description of SAFY was presented by Duchemin et al.[19]. In the present work, a modified version of SAFY, called SAFYE was used, introducing a dependence of biomass yield on crop water stress, adopting the same modification as used by [25] and [26]. These modifications were introduced since they allow an explicit assessment of the crop response to water availability and the characterization of water stress, which was lacking in the original model. The dry above-ground phytomass rate of change (ΔDAM) is multiplied by a water stress factor K_s , a dimensionless transpiration reduction factor dependent on available soil water, ranging between 0 and 1 [37]. K_s is calculated from the total available soil water in the root zone (TAW) and readily available soil water in the root zone (RAW), resulting from a simplified water balance driven by crop evapotranspiration, as described by the FAO irrigation and drainage paper n. 56 [37]. ΔDAM used in the modified version of SAFY is computed as:

$$\Delta DAM = APAR \cdot Pgro_Lue \cdot F_T(Ta) \cdot K_s \quad (3)$$

The version of SAFYE used in this work has 23 parameters in total, of which 10 were added in the present modified version to the 13 of the original SAFY [23]. Of these 23 parameters, 13 can be considered as conservative, i.e. theoretically not needing a local calibration [26], whereas 9 are cultivar specific and 1 depends on management factors, as listed in S2 Table. Additionally there are 8 input variables related to soil properties and management factors and 7 daily weather variables required: solar radiation, minimum and maximum air temperature, minimum and maximum relative humidity, mean wind speed and reference evapotranspiration.

SAFYE was tested with field data collected on winter wheat in Xiaotangshan (40.17°N, 116.43°E), near Beijing (China) during four years (2008-2011), within an experimental trial in which different sowing dates were used, according to a randomized block design with three replicates. In the first three years, three sowing dates were adopted: 28 September, 7 and 20 October in 2008-2009; 25 September, 5 and 15 October in 2009-2010; 25 September, 5 and 15 October in 2010-2011. The last year (2011-2012) only one sowing date was used: 25 September. The area of each plot was 100 m² in 2008-2010 and 300 m² in 2011. Weed control, pest management and fertilizer application were performed according to the local standard practices. The biomass and the leaf area index (LAI) were sampled 5-6 times during each growing season. The LAI-2000 Plant Canopy Analyzer (LI-COR Inc., Lincoln, NE, USA) was used for the determination of LAI. Aboveground biomass was collected using random sampling of a 0.25-m² area, with four replicates per plot. All samples were oven dried at 70 °C to a constant weight. The grain yields of each plot with three replicates for each treatment were obtained by randomly sampling a 1.5 m² area. Further details on this experimental dataset are provided by [8].

The performance of the model, previously validated by [25] and [26], was assessed by comparing simulated and measured data by computing the Root Mean Squared Error (RMSE) and the relative RMSE, i.e. RRMSE the ratio between RMSE and the mean of the measured data.

Sensitivity analysis strategy

Following a widely used procedure [9, 34], the SA strategy used in this study was to apply preliminarily a screening method, focusing on the computation speed

rather than on the assessment of higher order effects, in order to be able to exclude non-influential parameters of the models. In the second step, a variance based method is applied, which requires a longer computational time, but allows the assessment of interactions and higher order effects. Using only the ranking of influential parameters, it is possible to establish accurately and with acceptable computational costs, the effect that the simultaneous change of multiple input factors have on the output of the model. The two global sensitivity methods chosen in this work are the screening method of Morris [33] and the variance based method Extended Fourier Amplitude Sensitivity Test (EFAST) [28]. The latter method complements the first, which does not allow to quantify non linear and second order effects. Besides, although considered a GSA method, the experimental part of the Morris method is composed of individually randomized one at a time (OAT) experiments and sensitivity measures are typically considered qualitative (i.e., ranking significant input factors), but not necessarily quantitative in regard to the degree of significance [40].

Morris Method

The Morris Method [33] is designed for the quantification of elementary effects that the variation of input factors (parameters and input variables) produce on model outputs. More specifically, the aim of the method is to determine whether the effect of input factors variation are: negligible, linear and additive, non-linear or involving interactions with other factors [41].

The experimental design is composed of individually randomised one-at-time experiments. Morris design has the following characteristics [33]: (1) the points are sampled from a region (Ω) structured as a p-dimensional regular grid, (2) the design is structured in groups of points, called trajectories. In this study the number of points was chosen as equal to the number of factors, i.e. model input variables or parameters, so that $k=p$. In this case the number of trajectories (r) has been set to 20, as a compromise between the minimum value suggested by Campolongo et al. [41], i.e. 10 and the value used by [9], i.e. 25, in order to have limited computational cost. Each trajectory is composed by k points, each point is an input variable or parameter (also called factor) of the model. The number of trajectories simulated in this work is 500 ($Q=500$). In this set of simulated trajectories, r trajectories with the highest spread are chosen: in this case $r=20$. The spread is defined by Campolongo et al. [41].

The elementary effect (d_i) of i -th factor (X_i) due to a factor perturbation Δ is the relative difference between the output of the model obtained with and without perturbation:

$$d_i(X_1, \dots, X_n, \Delta) = \frac{y(X_1, \dots, X_{i-1}, X_i + \Delta, X_{i+1}, \dots, X_n) - y(X_1, \dots, X_n)}{\Delta} \quad (4)$$

where Δ is a value in $\{0, 1/(p+1), 2/(p+1), \dots, 1\}$, p is the number of intervals into which the parameter range is subdivided, $\mathbf{X} = (X_1, X_2, \dots, X_k)$ is any selected value in Ω such that the transformed point $(\mathbf{X} + \mathbf{e}_i \Delta)$ is still in Ω for each index $i=1, \dots, k$ and $\mathbf{e}_i$ is a vector of zeros but with a unit as its i -th component.

The random sampling of different $\mathbf{X}$ from Ω allows to obtain the finite distribution of elementary effects d_i associated with the i -th input parameter (F_i). Morris [33] suggested to use the mean (μ) and the standard deviation (σ) of the distribution F_i as sensitivity indices.

Empirical studies [42] have shown that the absolute value of mean (μ^*) can be considered as a “total sensitivity index” because a low μ^* is sufficient to define a parameter as negligible. It is sufficient that μ^* is below a threshold value to consider a parameter as negligible. In this study, a parameter or input variable was considered as non influential when, in all the scenario considered, μ^* was below a threshold value of 0.1 t ha⁻¹ of grain yield, which was the only output variable examined in this study. This threshold value was chosen arbitrarily, assuming to represent a yield error well below what has been reported to be the typical range of error found when employing the models [7-8, 10-12, 25-26]. .

Extended Fourier Amplitude Sensitivity Test (EFAST)

The EFAST method is an evolution of the Fourier Amplitude Sensitivity Test (FAST), which is a SA method based on the use of the variance. It employs the decomposition of the Sobol variance [30] and proposes to calculate sensitivities indices using the total variance of the output of the model and the conditional variances depending on the parameters.

The principle of FAST [31] is to change the inputs x_i at different frequencies and to examine the importance of the function (f) components at these frequencies.

The FAST method established a unique correspondence between the Fourier coefficients for ω_i frequency associated with the factor x_i and the sensitivity of 1st order of model outputs relative to x_i factor. Using the FAST method it is possible to calculate only the main effect on the output of the input variation, because it neglects the effects of the 2nd order, from which it is possible to quantify the sensitivity of the interaction between the variation of parameters. It is complex to calculate in detail the terms of interaction, because all the linear combinations of the frequencies ω_i must be considered. Saltelli et al. [32] show a method to quantify rather quickly the influence of each interaction, calculating a main sensitivity index (S_i) and an index of total sensitivity (ST_i), i.e. the sum of the main effect of x_i plus the interaction terms to all orders including x_i .

The sensitivity indexes are:

$$S_i = \frac{D_i}{D} \quad (5)$$

$$ST_i = 1 - \frac{D_{\sim i}}{D} \quad (6)$$

where D_i is the main effect of parameter i on the output variance, $D_{\sim i}$ is the sum of all the variance terms that do not include the index i and D is the variance of the output (in our case grain yield).

The innovation of EFAST, besides considering higher order effects and introducing the total sensitivity index, is in the way the n dimensional parameter space is explored. It uses a search curve defined by the following set of parametric equations:

$$x_i = \frac{1}{2} + \frac{\arcsin[\sin(\omega_i s)]}{\pi} \quad (7)$$

where s is a scalar ranging from $-\pi$ to π and ω_i is the frequency associated with the parameter i . Saltelli et al. [32] suggested optimal values of ω_i between $16 N_r$ and $(16 N_r + 48 N_r)$ where N_r is the number of different search curves within the hyperspace.

In the present study, the EFAST method has been used subsequently to the application of the Morris method, analyzing only the parameters found to be non-negligible. The sensitivity indices of EFAST were calculated for individual parameters as first-order effects S_i , i.e. the main sensitivity index, and total effects ST_i , first order plus interactions of higher order with other parameters, i.e. the total sensitivity index. S_i and ST_i range between 0 and 1, with higher values indicating more important effects. There is not an objective way of establishing a threshold value for considering a parameters as negligible or influential, but we decided to adopt a threshold of ST_i of 0.1 (i.e. 10% of the output variance), on the basis of an analysis of the literature, since this was the most frequently used value [16, 29, 34, 40].

For this study a Matlab (Mathworks Inc., MA, USA) implementation of the EFAST and Morris algorithms was used [42].

Range of variation of model parameters

The range within which the parameters and input variables were allowed to vary during the GSA was defined in order to represent the uncertainty on their values in the context of a regional scale model application. For both Aquacrop and SAFYE, the parameters considered as conservative, i.e. not suppose to be varied for a given crop species, were allowed to vary within a maximum range of $\pm 33\%$ of a mean nominal value obtained from the literature as detailed in S1 and S2 Tables. In facts, the calibration results of different studies on wheat [7-8, 10-12] suggest that even for these parameters a limited range of variation is needed, thus we adopted for these parameters a range resulting from the values used these studies, including them in the SA, as also done by Vanuytrecht et al. [9]. The range of variation of the other parameters and input variables was also defined, as specified in S1 and S2 Table, according to the literature or expert knowledge on the variability of each factor likely to be faced when applying the models in a regional scenario. It should be noted that this application context is rather different from the typical field scale model application, in which knowledge of many soil input variables (e.g. soil texture) or management factors (e.g. sowing date) is generally available. The GSA for Aquacrop was carried out for 9 input variables and 45 parameters, of which 29 were considered as conservative, as presented in the S1 Table.

For SAFYE, the range of variation of the parameters and input factor was set using the same criteria used for Aquacrop, adopting whenever possible the same ranges for equivalent parameters in the two models, as detailed in S2 Table. The GSA was carried out for 8 input variables and 23 parameters, of which 13 were considered as conservative.

Assessment of model complexity and plasticity

The two models, Aquacrop and SAFYE, were compared in terms of two indicators derived from the GSA results: complexity and plasticity. Complexity is the parsimony of the model in representing the biophysical system, providing information about the amount and relevance of model parameters and inputs [43]. Plasticity is defined as the aptitude of a model to change the sensitivity to its parameters and inputs when the conditions of application change [44].

In this study we choose to represent the complexity through the parameter ratio (R_p) introduced by Confalonieri et al. [44]. It is the proportion of relevant parameters (S) over the total number of model parameters (T):

$$R_p = \frac{S}{T} \quad (8)$$

All the parameters and input variables for which the ST_i was higher than 0.1 were considered as relevant (S), whereas T was the total number of factors tested in the SA. According to Confalonieri et al. [44], plasticity can be quantified using an index L:

$$L = TDCC \cdot e^{\sigma_{SAM}^{-1}} \quad (9)$$

where TDCC is the top-down concordance coefficient presented by Iman and Conover [45] and σ_{SAM} is the standard deviation of a normalized agrometeorological indicator (SAM) proposed by Confalonieri et al [46]. TDCC, ranging between 0 and 1, is considered suitable for comparing parameters rankings obtained from SA carried out under different conditions. It has the capability of emphasizing the agreement among rankings assigned to relevant parameters and of deemphasizing the disagreement among those assigned to less important parameters [43]. Considering the value of a sensitivity index SI_{ij} for parameter X_i and replicate R_j (each replicate corresponds to a SA ranking), ranks assigned to the sensitivity index for replicate R_j can be denoted as $r(SI_{ij})$, $i=1, 2, \dots, k$ for the k parameters. A rank of 1 is assigned to the parameter X_i with the largest value of SI_{ij} , a rank of 2 is assigned to the parameter with the second largest SI_{ij} , and so on. Then, Savage scores $ss(SI_{ij})$ can be calculated according to Eq. (10):

$$ss(SI_{ij}) = \sum_{i=r(SI_{ij})}^k \frac{1}{i} \quad (10)$$

for all the parameters X_i and replicates R_j . The TDCC is then calculated as:

$$TDCC = \frac{\sum_{i=1}^k \left[\sum_{j=1}^{n_{SA}} ss(SI_{ij}) \right] - n_{SA}^2 \cdot k}{n_{SA}^2 \cdot \left[k - \sum_{i=1}^k \frac{1}{i} \right]} \quad (11)$$

where n_{SA} is the number of sensitivity analysis replicates.

The normalized synthetic agrometeorological indicator SAM [46], ranges between -1 (corresponding to maximum drought) and +1, is computed as:

$$SAM = \frac{Rain - ET_0}{Rain + ET_0} \quad (12)$$

where $Rain$ (mm) and ET_0 (mm) represent respectively the cumulative rainfall and reference evapotranspiration. Confalonieri et al. [46] showed that SAM is a very useful standardized indicator for characterizing the conditions of applications of crop models. The use of an exponential function of the variability of SAM in the index L was chosen by [44] to improve its discriminating capability at plasticity values close to zero and to reduce asymmetries in the distribution of L.

The plasticity index L proposed by Confalonieri et al. [44] ranges from 0 to about 1.51, with the highest plasticity at 0. This indicator couples model response, represented by TDCC, with the variability of environmental conditions, represented by the standard deviation of SAM. Since plasticity is related to the capability of simulation models to change the way they react to a changing environment, by changing the relevance of their parameters, [44] consider a high plasticity as optimal, because of the pronounced phenotypic plasticity intrinsic in the modelled cropping systems. However in the

practical terms, a high plasticity would mean that the results of a sensitivity analysis would not have a general validity.

Climatic scenarios

The meteorological data used in this study to drive the model simulations for the sensitivity analysis were obtained from three sites representing contrasting environments, in terms of temperature extremes and water availability during the winter wheat growth season (Table 1). They were selected in order to encompass a large variability of water limitation patterns during the winter wheat growth season, either in the early phase (autumn to beginning of winter) or in the full wheat development phase (end of winter to spring).

The Xiaotanshan site (lat. 40.17°N, long. 116.43°E, alt. 57 m), near Beijing (China), is characterized by a continental climate, with a cold dry winter and a hot wet summer, belonging to the Koppen-Geiger [47] class Dwa (cold, dry winter, hot summer), i.e. monsoon-influenced, hot-summer humid continental climate.

The Yangling site (lat. 34.27°N, lon. 108.09°E, alt. 460 m), in the Chinese Province of Shaanxi, is located within a Monsoon-influenced humid subtropical area, falling within the Cwa Koppen-Geiger class (temperate, dry winter, hot summer), characterized by high evapotranspirative deficit in spring.

Viterbo (lat. 42.72°N, long. 12.12°E, alt. 310 m), Central Italy, has a typical Mediterranean climate with wet autumn and dry hot summer, falling into the Koppen-Geiger class Csa (temperate, dry summer, hot summer).

The dataset consisted of daily rainfall, temperatures (mean, maxima and minima), relative humidity and wind speed. Solar radiation was calculated from the temperature range using the relationship proposed by Bristow and Campbell [36]. Reference evapotranspiration was calculated using the FAO Penman-Monteith method [37], using all the available variable (temperature, humidity, solar radiation and wind speed).

To quantify the degree of crop water stress in the three sites, we calculated the evapotranspirative deficit, i.e. the difference between rainfall and reference evapotranspiration (ET_o) and the normalized synthetic agrometeorological indicator (SAM) [46]. The advantage of using this index is that it expresses the evapotranspiration deficit in relative terms, allowing an easy comparison among sites and years, though it is only a partial indicator of the seasonal conditions impacting crop growth, since it does not take into account temperature extremes.

Based on such indicators, we selected, for each site, two winter wheat growth seasons (Fig 1), spanning from the 1st of September to 31st of July, characterized respectively by relatively dry or wet conditions. These were selected on the basis of the evapotranspiration deficit and SAM occurring in the periods considered as most important for winter wheat growth (Table 1), i.e. from October to December (autumn to beginning of winter) and from February to June (end of winter to spring).

Fig. 1. Meteorological datasets used in the sensitivity analysis study. Time trends of daily maximum and minimum air temperature (left), rainfall and reference evapotranspiration (right) for the sites of Xiaotanshan (top), Yangling (centre) and Viterbo (bottom), for the two years considered in each study site.

Table 1. Summary of the climatic data sets used in this study. Cumulative rainfall and reference evapotranspiration (ET_o), average maximum and minimum temperature, evapotranspiration deficit (rain-ET_o) and synthetic agrometeorological indicator (SAM) [46], in the most crucial periods of wheat growth, for the years used in the sensitivity analysis study and for the long term climatic data (30 years).

Site	Year	Autumn-Winter (Oct-Dec)							Winter-Spring (Feb-Jun)							
		Rain (mm)	ET _o (mm)	Tmin (°C)	Tmax (°C)	Deficit ET (mm)	SAM		Year	Rain (mm)	ET _o (mm)	Tmin (°C)	Tmax (°C)	Deficit ET (mm)	SAM	
Yangling	1994	236.6	392.7	10.3	18.4	-156.1	-0.2	Wet	1995	189.5	817.1	9.9	22.0	-627.6	-0.6	Dry
	1986	197.1	405.5	8.8	18.0	-208.4	-0.3	Dry	1987	442.3	726.6	8.7	19.9	-284.3	-0.2	Wet
	long term								long term							
	mean	266.0	391.7	9.3	18.3	-125.8	-0.2		mean	306.6	761.9	9.1	20.3	-455.4	-0.4	
Viterbo	2002	210.4	138.3	9.2	18.3	72.1	0.2	Dry	2003	106.2	538.1	8.1	22.7	-431.9	-0.7	Dry
	2009	350.6	113.1	6.3	15.9	237.5	0.5	Wet	2010	508.2	422.1	7.2	18.3	86.1	0.1	Wet
	long term								long term							
	mean	270.8	129.4	6.5	15.6	141.3	0.4		mean	276.9	460.8	7.1	18.8	-183.9	-0.2	
Xiaotangshan	2010	139.8	122.8	5.3	15.0	17.0	0.1	Wet	2011	432.5	508.5	8.2	18.9	-76.0	-0.1	Wet
	2011	116.8	250.2	5.9	14.9	-133.4	-0.4	Dry	2012	481.9	785.1	8.3	18.4	-303.2	-0.2	Dry
	long term								long term							
	mean	37.59	128.7	0.5	10.9	-91.1	-0.5		mean	152.7	491.2	7.1	18.9	-338.5	-0.5	

Results

Assessment of the SAFYE model

The version of the SAFYE model used in the present work was able to simulate quite accurately the winter wheat grain yield of the Xiaotangshan dataset (Fig 2), with an overall RMSE of 0.25 t ha⁻¹ and a RRMSE of 5% for the final grain yield estimation across all years and sowing dates. The time trends of simulated above ground biomass and LAI were in agreement with measured data for most of the sowing dates, although the time of simulated grain maturity was slightly delayed as compared to the actual harvest dates (S1 File).

SAFYE lead to a small improvement in the estimation of grain yield as compared to the original version of the SAFY model [23] for which the RMSE was 0.27 t ha⁻¹, for this dataset.

Fig 2. Relationship between measured and simulated grain yield in winter wheat with the SAFYE model, across 4 years at Xiaotangshan.

Application of Morris methods

The results of the Morris method (both for Aquacrop and SAFYE) showed that the value of μ^* was strongly influenced by the climatic scenarios employed. In facts, values of μ^* were rather different not only between the sites, but they also varied significantly among the years analyzed (Fig 3 and Fig 4). However, there was a number of factors that were consistently non-influential, i.e. with μ^* values lower than 0.1 t ha⁻¹, that could be identified for both models, which was the main objective of the application of this screening method.

Fig 3. Values of Morris Mean Effect (μ^*) for the Aquacrop model using climatic scenarios for Xiaotangshan (a), Yangling (b) and Viterbo (c) for two crop growth seasons each. Bars in red identify drier years, bars in blue wetter years. Parameters abbreviations are given in S1 Table. Parameters with mean μ^* values lower than 0.1 t ha⁻¹ have been omitted.

Fig 4. Values of Morris Mean Effect (μ^*) for the SAFYE model using climatic scenarios for Xiaotangshan (a), Yangling (b) and Viterbo (c) for two crop growth seasons each. Bars in red identify drier years, bars in blue wetter years. Parameters abbreviations are given in S2 Table. Parameters with μ^* values lower than 0.1 t ha⁻¹ have been omitted.

In general, higher values of μ^* were observed for Aquacrop than for SAFYE. In particular, very high values of μ^* were found for Aquacrop in Viterbo, for which the highest sensitivity was for the day of sowing (*dos*), and for the temperature sum (GDD) until senescence (*sen*), grain physiological maturity (*mat*) and until flowering (*flo*). The latter two parameters seemed especially influential in the wetter year. Also for Yangling these parameters resulted quite important and the influence of *flo* and *sen* was similarly higher in the wetter year, whereas they resulted less influential but still not negligible in Xiaotangshan. For the Xiaotangshan site though, the sensitivity of

yield to these parameters was still noticeable, but less strong than that of other factors such as soil field capacity (*fc*) and wilting point (*pwp*). The latter two input variables were also very influential for Yiangling and in both sites their μ^* values were higher in the relatively wetter years. It should be noted that these two sites had a remarkable evapotranspirative deficit also in the relatively wetter years (Table 1) and it could be expected that in these conditions the influence of soil water holding characteristics, regulating rain water storage in the root zone, is more important when there is sufficient rain during the growth season, than when rainfall is reduced. A similar pattern, i.e. stronger influence in wetter years in these two sites, appeared also for two other parameters linked to soil water stress: the soil water depletion factor for canopy senescence (*psen*) and for stomatal control (*psto*).

Overall, the SA for Aquacrop shows that all the factors categorized by [38], as related to the development of green canopy, as well as flowering, can be considered as influential. These include both phenological parameters, such as the temperature sum to emergence (*eme*), to crop maturity (*mat*) and the length of flowering (*flolen*), in addition to *sen* and *flo* already mentioned, as well as parameters describing the increase (*cgc4gdd*) or decline (*cdc4gdd*) of the canopy cover. Three parameters related to root development resulted non negligible (i.e. with μ^* higher than 0.1 t ha^{-1}): most notably maximum rooting depth (*rtx*) and the growing degree days (GDD) to reach maximum root depth (*root*), especially for the Xiaotangshan and Yangling sites. All the parameters related to air temperature stress were non-negligible, with the base temperature below which crop development stops (*To_crop*) and below which pollination starts to fail (*polmn*), especially influential in Viterbo in 2002-03 where low temperatures occurred in March (Fig 1). It is interesting to note that *To_crop* was not influential in Xiaotangshan in which temperatures were well below the range used for this parameter, whereas in Viterbo the temperatures recorded in winter tended to be close to the range used for *To_crop*. Also parameters linked to crop transpiration (*kc* and *kcdcl*), to crop water productivity (*wp* and *wp_yfp*) and to the harvest index (*hi* and *hilen*) were found to be influential. The climate scenarios (sites and years) influenced the rankings of the parameters on the basis of the μ^* , as measured by the top-down concordance correlation (TDCC) which overall had a value of 0.47 (in a range 0 to 1). The correlation between the results of the sensitivity analysis of the different scenarios was slightly lower among the wet years (TDCC=0.49) than among the dry years (TDCC=0.51), highlighting more differences among the previous. The correlation within sites was higher for Xiaotangshan and Yangling, TDCC of 0.7 and 0.75, than for Viterbo which had a correlation of 0.63 among the sensitivity analysis results of the dry and wet year.

Despite the differences of climate and sites, it is possible to distinguish a group of factors which was constantly negligible, i.e. for which μ^* was always less than the 0.1 t ha^{-1} threshold (Table 2).

Table 2. List of Aquacrop factors that were found to have a negligible influence on grain yield, i.e. with Morris total sensitivity index $\mu^* < 0.1 \text{ t ha}^{-1}$, in all the climatic scenarios of the present study.

Name of parameter or input variable	Type	Process affected (from [38])	Description
<i>polmx</i>	conservative parameter	Air temperature stress	Maximum air temperature above which pollination starts to fail

<i>evardc</i>	conservative parameter	Crop transpiration	Effect of canopy cover in reducing soil evaporation in late season stage
<i>cpc2</i>	conservative parameter	Crop water productivity	Crop performance under elevated atmospheric CO ₂ concentration
<i>rtmin</i>	management input	Development of root zone	Minimum effective rooting depth
<i>rtexup</i>	cultivar specific parameter	Development of root zone	Maximum root water extraction in top quarter of root zone
<i>rtxlw</i>	cultivar specific parameter	Development of root zone	Maximum root water extraction in bottom quarter of root zone
<i>exc</i>	conservative parameter	Harvest Index	Excess of potential fruits
<i>hipsveg</i>	conservative parameter	Harvest Index	Coefficient describing positive impact on HI (harvest index) of restricted vegetative growth during yield formation
<i>hingsto</i>	conservative parameter	Harvest Index	Coefficient describing negative impact on HI of stomatal closure during yield formation
<i>hipsflo</i>	conservative parameter	Harvest Index	Possible increase of HI due to water stress before flowering
<i>hinc</i>	conservative parameter	Harvest Index	Allowable maximum increase of specified HI
<i>Ssf</i>	management input	Soil fertility stress	Soil fertility/salinity stress coefficient
<i>ecss</i>	conservative parameter	Soil salinity stress	Electrical Conductivity of soil saturation extract at which crop can no longer grow
<i>ecsss</i>	conservative parameter	Soil salinity stress	Electrical Conductivity of soil saturation extract at which crop starts to be affected by soil salinity
<i>fk</i>	management / environment	Soil water stress	Evaporation decline factor for stage II
<i>Ksat</i>	management input	Soil water stress	Saturated hydraulic conductivity
<i>psenshp</i>	conservative parameter	Soil water stress	Shape factor for water stress coefficient inducing early senescence
<i>ppol</i>	conservative parameter	Soil water stress	Soil water depletion factor for pollination: upper threshold (fraction TAW)
<i>rew</i>	management input	Soil water stress	Readily evaporable water from top layer
<i>pexup</i>	conservative parameter	Soil water stress	Soil water depletion factor for canopy expansion: upper threshold, fraction of total available water (TAW)
<i>anaer</i>	cultivar specific / environment	Soil water stress	Anaerobic point below saturation limiting aeration
<i>cn</i>	management input	Soil water stress	Soil Curve Number

These include all the factors linked to soil salinity or fertility stress, as well as some parameters related to the harvest index (*exc*, *hipsveg*, *hingsto*, *hipsflo* and *hinc*) and *cpc2*, i.e. crop performance under elevated atmospheric CO₂, which was obviously irrelevant since CO₂ was kept constant in the simulations. Some soil water stress

factors were negligible, though others of this category were influential as discussed above.

The SA for SAFYE highlighted a set of factors with μ^* values higher than the threshold of 0.1 t ha^{-1} established for considering them as influential. This set was approximately the same in all experimental sites, although the values of μ^* were generally higher in Yangling and in Xiaotangshan, as compared to Viterbo, for which the highest values were found in the driest year. This was in contrast with the other two sites, for which higher μ^* values were generally found in the wetter years, especially in Yangling.

Similarly to what was found for Aquacrop, higher values of μ^* for field capacity (*FC*) and wilting point (*WP*) were found in the wetter years in Xiatangshan and Yangling, though in this case they were comparatively less influential than other parameters. The emergence date (*Pfen_MrgD*), which can be considered equivalent to the sowing date in Acquacrop, since that input is not used in SAFYE, was the most influential factor for the dry year in Viterbo, similarly to what found for Aquacrop. In that year the drought and high temperatures occurring in the late spring caused an earlier senescence and termination of the growth cycle, thus sowing date would have been crucial, allowing a longer or shorter time period for photosynthetic assimilation. Specific leaf area (*Pgro_Sla*), used in the model to convert biomass into leaf area, was found to be very influential in all sites. The same was true for the other parameters related to green canopy development, such as the two parameters appearing in the function regulating the partitioning of biomass to leaves (*Pfen_PrtA* and *Pfen_PrtB*), as well as parameters regulating the senescence (*Pfen_SenA* and *Pfen_SenB*). The conversion factor of solar radiation into PAR (*Pgro_R2P*) had a very strong influence on grain yield, especially at Xiaotangshan and Yangling where more variability in irradiance levels were found than at Viterbo (data not shown, but it can be inferred from ETo patterns in Fig 1). All the parameters related to air temperature stress were also influential (*Ptfn_Topt*, *Ptfn_TpSn*, *Ptfn_Tmin*, *Ptfn_Tmax*). As expected the two key parameters affecting PAR absorption by the canopy (*Pgro_Kex*) and its conversion into biomass (*Pgro_Lue*) were very important, similarly to the parameter regulating the partitioning of biomass to grain (*Pgro_P2G*).

The factors that were found to have a negligible impact, i.e. with μ^* values lower than 0.1, were all those introduced in SAFYE (with the exception of soil field capacity and wilting point), i.e. factors linked to transpiration, root development and water stress (Table 3).

Table 3. List of SAFYE factors that were found to have a negligible influence on grain yield, i.e. with Morris total sensitivity index $\mu^* < 0.1 \text{ t ha}^{-1}$, in all the climatic scenarios of the present study.

Name of parameter or input variable	Type	Process affected	Description
<i>Kc_tab_ini</i>	Conservative parameter	Crop transpiration	Crop transpiration coefficient at the initial development stage
<i>Kc_tab_mid</i>	Conservative parameter	Crop transpiration	Crop coefficient at mid stage
<i>h_max</i>	Non conservative parameter	Crop transpiration	Maximum crop height

<i>Kc_tab_end</i>	Conservative parameter	Crop transpiration	Crop coefficient at final stage
<i>SMT_sen</i>	Non conservative parameter	Development of green canopy cover	Temperature sum to complete senescence
<i>Pfen_stel</i>	Non conservative parameter	Development of green canopy cover	Temperature sum at 10% canopy cover, i.e. stem elongation
<i>Pfen_flw</i>	Non conservative parameter	Flowering	Temperature sum to flowering
<i>Rd_tab_end</i>	Input variable	Root development	Root depth at end of growth
<i>Rd_tab_ini</i>	Input variable	Root development	Root depth at emergence
<i>p_tab_end</i>	Conservative parameter	Soil water stress	Fraction of readily available water at the final stage
<i>p_tab_mid</i>	Conservative parameter	Soil water stress	Fraction of readily available water at the mid stage
<i>De</i>	Input variable	Soil water stress	initial cumulative depth of evaporation
<i>p_tab_ini</i>	Conservative parameter	Soil water stress	readily available water as fraction of totally available water at the initial stage
<i>Dr</i>	Input variable	Soil water stress	initial root zone depletion

The correlation among the results obtained in the different scenarios was much higher with SAFYE than for Aquacrop, with an overall TDCC of 0.92. Although values of TDCC were generally much higher, similarly to what was found for Aquacrop the results of the wet years had a lower correlation (TDCC=0.9) than those of dry years (TDCC=0.95). Again, the dry and wet scenarios at Viterbo (TDCC=0.90) were more different than at Xiaotangshan (TDCC=0.93) and Yangling (TDCC=0.98).

Application of EFAST method

Fig 5 shows the two EFAST Sensitivity indices, calculated for Aquacrop for the three locations, for each crop season. The indices represent individual first-order effects, i.e. main sensitivity index (S_i), and first order plus interactions of higher order with other parameters, i.e. the total sensitivity index (ST_i) on the variance of model output, i.e. grain yield. Some differences in the degree of sensitivity to what was found using the Morris method are apparent. It should be noted that the Morris sensitivity index μ^* includes both first order effects and interactions, so it should be compared to ST_i rather than to S_i . TDCC was computed, to assess the correlation (for common factors) among the rankings of μ^* provided by Morris and those of EFAST based on ST_i , revealing similarities between the results of the two methods. It was found that the correlation between Morris and EFAST results was higher among the dry years for Yangling and Viterbo (TDCC respectively of 0.86 and 0.91), whereas for Xiaotangshan it was higher in the wet year (0.84) than in the dry year (0.56).

EFAST revealed that interactions and second order effects were overall very high for Aquacrop, as shown by the fact that few factors had strong first order effect, revealed by high S_i (Fig 5). These included soil water holding properties (field capacity and wilting point, fc and pwp) especially at Xiaotangshan and Yangling, in which much drier conditions prevailed as compared to Viterbo. The length of the crop cycle (mat)

and the maximum effective rooting depth (*rtx*) had also a quite important first order effect. The temperature sum to senescence (*sen*) had a relevant first order effect only for Viterbo. Considering the total sensitivity index (ST_i), three parameters were found to be influential ($ST_i > 0.1$) across all sites and years. These were the day of sowing (*dos*) and the phenological parameters indicating the growing degree days from sowing to maturity (*mat*) and to senescence (*sen*). Another important phenological parameter, GDD to flowering (*flo*) was also very influential except for the dry year at Xiaotangshan. Soil field capacity (*fc*) and wilting point (*pwp*) were influential in all scenarios considered except for the wetter year in Viterbo. The maximum effective rooting depth (*rtx*) contributed markedly on the output variance for all sites and years except again for the wet year at Viterbo. These results seem to confirm that even in the years selected as relatively wet for the other two sites, conditions of crop water stress occurred.

Fig 5. EFAST results for Aquacrop. Main Sensitivity index (Si) (left panels) and Total Sensitivity index (ST_i) (right panels), for the three sites, Xiaotangshan (top), Yangling (centre) and Viterbo (bottom), for the two wheat growth seasons examined (wet years bars are blue and dry years red). Abbreviations of parameters and input factors are reported in S1 Table.

Adopting the threshold of the total sensitivity index to identify influential factors ($ST_i > 0.1$), considering the maximum values found across all the scenarios, 4 input variables and 13 parameters, of which 5 were conservative, were selected (Table 4).

Table 4. List of factor of the Aquacrop model resulting highly influential on the grain yield, according to the main sensitivity index ($ST_i > 0.1$) from the EFAST analysis, ranked from the most to the least influential.

Name of parameter or input variable	Type	Process affected	Description
fc	Input variable	Soil water stress	Soil Water Content at Field Capacity
dos	Input variable	all	day of sowing
sen	Non conservative parameter	Development of green canopy cover	GDD from sowing to start senescence
flo	Non conservative parameter	Flowering	GDD from sowing to flowering
pwp	Input variable	Soil water stress	Soil Water content at Wilting Point
mat	Non conservative parameter	Development of green canopy cover	Length of the crop cycle
rtx	Input variable	Development of root zone	Maximum effective rooting depth
hilen	Non conservative parameter	Harvest Index	Period of Harvest Index build up during yield formation starting at flowering
psen	Conservative parameter	Soil water stress	Soil water depletion factor for canopy senescence: upper threshold

Kex	Non conservative parameter	Soil water stress	Soil evaporation coefficient for fully wet and non-shaded soil surface
psto	Conservative parameter	Soil water stress	Soil water depletion fraction for stomatal control: upper threshold
wp	Conservative parameter	Crop water productivity	Water productivity normalized for ETo and CO ₂
hi	Non conservative parameter	Harvest Index	Reference Harvest Index
polmn	Conservative parameter	Air temperature stress	Minimum air temperature below which pollination starts to fail
cdc4ggd	Conservative parameter	Development of green canopy cover	CDC for GGD: decrease in canopy cover

However it can be observed that additional parameters had a smaller, but still possibly non negligible influence on yield (Fig 6). The correlation (TDCC) among the sensitivity analysis results obtained in the different scenarios for Aquacrop was higher for EFAST, with a TDCC of 0.71, than for Morris (0.47). Contrarily to what observed with the results of Morris, a higher correlation was found among the results of the wet years (0.86) than of the dry years (0.65), but also in this case it was found that Viterbo (TDCC=0.68) was the site in which the two years were the most different as compared to the other two sites (0.89 and 0.93).

Fig 6. First-order and total sensitivity indices estimated by the EFAST method for the Aquacrop model. The first part of the bars (dark grey) corresponds to the average estimate of the first-order index (S_i) over all the sites and scenarios, the full bars indicate average estimates of total index (ST_i), while the lines indicate extreme estimates of total indices.

The EFAST results for SAFYE (Fig. 7) revealed that for this model, first order effects were relatively more important than for Aquacrop. The similarity between the rankings obtained from Morris and EFAST was higher for SAFYE as compared to Aquacrop. For SAFYE the TDCC values between the results of the two methods ranged between 0.83 (Viterbo wet year) to 0.96 (Yangling wet year). Factors showing noticeable EFAST main sensitivity index values concerned especially the climatic conversion factor from solar radiation to PAR (*Pgro_R2P*) and the specific leaf area (*Pgro_Sla*) already shown to be very influential from the Morris method. A high total effect on the output variance, i.e. including also interactions and second order effects, was shown by 4 parameters. In addition to *Pgro_R2P* and *Pgro_Sla*, these included also the parameter regulating the partitioning of biomass to leaves (*Pfen_PrtB*) and the temperature threshold to start senescence (*Pfen_SenA*). The light extinction coefficient in the canopy (*Pgro_Kex*) was also found to be generally influential, except for the dry year in Viterbo and the wet year in Xiaotangshan. The emergence date (*Pfen_MrgD*) was only relevant for the dry year in Xiaotangshan. These different results were nevertheless more correlated among them than those of Aquacrop, yielding in this case a TDCC of 0.93. For SAFYE the correlation among rankings of ST_i was roughly similar for wet and dry scenarios, with TDCC values of 0.94 and 0.93 respectively. Also in this case it was found that the results of Viterbo (TDCC=0.81) were less similar among themselves than those of the two other sites (TDCC of 0.98 and 0.97).

Fig 7. EFAST results for SAFYE. Main Sensitivity index (S_i) (left panels) and Total Sensitivity index (ST_i) (right panels), for the three sites, Xiaotanshan (top), Yangling (centre) and Viterbo (bottom), for the two wheat growth seasons examined (wet years bars are blue and dry years red). Abbreviations of parameters and input factors are reported in S2 Table.

When taking into account the threshold set for considering a factor as influential, applied to the maximum ST_i across all climate and site scenarios, only 5 parameters and one input variable were selected (Table 5). The influential factors were the same for both dry and wet scenarios, with the exception of *Pfen_MrgD* which had a strong effect on the output variance ($ST_i > 0.1$) only in the dry year in Xiaotangshan.

Table 5. List of factor of the SAFYE model resulting highly influential on the grain yield, according to the main sensitivity index ($ST_i > 0.1$) from the EFAST analysis, ranked from the most to the least influential.

Name of parameter or input variable	Type	Process affected	Description
Pgro_Sla	Conservative parameter	Development of green canopy cover	Specific Leaf Area ($\text{m}^2 \text{g}^{-1}$)
Pgro_R2P	Conservative parameter	Radiation environment	Climatic efficiency: ratio of incoming photosynthetically active radiation (PAR) to global radiation
Pfen_SenA	Non conservative parameter	Development of green canopy cover	Temperature threshold to start senescence ($^{\circ}\text{C}$)
Pfen_PrtB	Non conservative parameter	Development of green canopy cover	Partition to leaf function parameter 2 (PLb)
Pgro_Kex	Non conservative parameter	Development of green canopy cover	Light extinction coefficient in canopy
Pfen_MrgD	Input variable	Development of green canopy cover	Day of the year of emergence

Other parameters also important were found to be in particular two parameters linked to temperature stress (*Ptfn_Topt* and *Ptfn_TpSn*) and the light use efficiency (*Pgro_Lue*) (Fig. 8).

Fig 8. First-order and total sensitivity indices estimated by the EFAST method for the SAFYE model. The first part of the bars (dark grey) corresponds to the average estimate of the first-order index (S_i) over all the sites and scenarios, the full bars indicate average estimates of total indices (ST_i), while the lines indicate extreme estimates of total indices.

Complexity and plasticity of Aquacrop and SAFYE

The results of the EFAST sensitivity analysis allowed to quantify the complexity and the plasticity of Aquacrop and SAFYE. The complexity of these two crop models was

measured using the parameter ratio (R_p) (Eq. 9) i.e. the ratio of influential factors over the total. The factors, input variables and parameters, that were considered relevant, were those which had a total sensitivity index ST_i greater than the arbitrary threshold value fixed at 0.1. For Aquacrop R_p had a value of 0.28, indicating that 28% of the factors tested (15 out of 54), i.e. including both parameters and input variables, were found to be influential. For SAFYE a lower complexity index R_p of 0.19, was found, since 6 factors out of the 31 tested were found to be influential.

The plasticity of the models was calculated by the index L (eq. 9). This index is obtained from the combination of the TDCC, expressing the correlation among rankings of the total sensitivity index ST_i in the different scenarios, with the standard deviation of the agrometeorological indicator SAM, expressing the variability of the environmental conditions. The results are summarized in Table 6, which shows that all the scenarios employed in this study were characterized by an evapotranspiration deficit, as testified by the SAM values lower than zero. The largest variability in the climatic conditions between the dry and wet years occurred at Viterbo and the lowest at Yangling. In general, the climatic variability across sites was higher among the wet years than among the dry years.

Table 6. Synthesis of the assessment of the plasticity of the SAFYE and Aquacrop models. The plasticity index L is calculated according to Eq. 9 from the normalized synthetic agrometeorological indicator (SAM) (eq. 12) and its standard deviation σ_{SAM} and the top-down concordance coefficient (TDCC) (eq. 11).

	SAM	σ_{SAM}	TDCC		L	
			SAFYE	Aquacrop	SAFYE	Aquacrop
Site						
Xiaotanshan	-0.17	0.18	0.98	0.89	0.43	0.39
Yangling	-0.34	0.15	0.97	0.93	0.41	0.40
Viterbo	-0.02	0.26	0.81	0.68	0.38	0.32
Year						
wet	-0.04	0.20	0.94	0.86	0.42	0.39
dry	-0.32	0.12	0.93	0.65	0.39	0.27
Total	-0.18	0.21	0.93	0.71	0.42	0.32

As described in the previous sections, the TDCC, expressing the correlation between rankings of the total sensitivity index ST_i among different scenarios, was always higher for SAFYE than for Aquacrop. This meant that a higher plasticity, indicated by lower L values, was always found for the latter model. The largest differences between the two models in terms of TDCC values, and hence of L, were found between the two years at Viterbo and among the three scenarios of the dry years, where the Aquacrop model demonstrated a much higher plasticity than SAFYE.

Discussion

The results obtained in this study, performed using a contrasting range of water limited climatic scenarios of very different winter wheat growing areas, allowed to obtain useful information on the sensitivity and plasticity of two models, SAFYE and

Aquacrop, increasingly used in regional scale studies [14, 24, 26, 48-51]. In these conditions, there is typically greater uncertainty on parameters and input factors, as compared to field based application of crop models. Therefore it is important to know on which factors data collection and calibration efforts should concentrate. Global sensitivity analysis is considered as the state of the art technique for such purpose [28]. It is well known that the results of sensitivity analysis studies depend on to the "boundary conditions" chosen [52]. In this specific case, these conditions are the climate dataset, actual data from three different sites, and the range of variation of parameters and input variables. The range of variation of these factors was chosen to be the same for all the tests of a given model and, as much as possible, a similar (or the same) range of variation was adopted for analogous factors of the two models. This was done in order to reduce the variability of the boundary conditions of the SA and provide some possible element of comparison between the models. The climatic conditions are therefore the only variables that differentiate the scenarios between the sites. Thus it is possible to distinguish more simply if the parameters are always influential or always negligible, with similar degrees of influence for any scenario, or if their degree of influence varies with the climatic dataset. The climatic data using in this study were chosen not to be representative of all the possible wheat growing areas, but only of situations of moderate to strong water deficit, since the main interest was that of providing tools for the assessment of the impact of such conditions on wheat yield, at the regional scale. Thus the results of this study could be only valid for similar climatic scenarios, which nevertheless characterize quite a large extent of the wheat growing areas. The overall value of the σ_{SAM} , as reported in Table 6, indicates that the variability among the climatic scenarios used in the present work was very high, being higher for instance than the overall value of the dataset used by [44] which included locations all over Europe (e.g. from Ukraine to Greece) for 10 years.

The different climatic scenarios led to differences in the ranking of the sensitivity of the two models for both GSA techniques, Morris and EFAST. It is to be expected that the most important factors can differ between climatic scenarios, since for example, a parameter that determines the effect of water stress will be important in situations where there is actually water stress, but not where water is non-limiting. This is an important aspect that affect the results of all sensitivity analysis studies [52]. However, some common patterns do appear, allowing to classify groups of parameters as influential or negligible, even in highly contrasting climatic scenarios, as illustrated in our study. We used an objective measure, the top down correlation coefficient (TDCC) to assess the agreement among the ranking of factors assessed across different climates. It was found that the correlations (TDCC values) were always statistically significant ($P < 0.01$), using the methodology to testing its significance reported by Confalonieri et al. [44]. The correlation of the results of the two EFAST rankings of the wet and dry years in a given location ranged between 0.68, for Aquacrop in Viterbo, to 0.98, for SAFYE in Xiaotangshan (Table 6). The correlation among rankings across sites ranged from 0.65 for Aquacrop in the dry years, to 0.94 for SAFYE in the wet years. Overall, it emerged a clear picture, showing that SAFYE had a lower plasticity than Aquacrop, for which the sensitivity analysis results were more affected by the differences in the climatic scenarios.

The impact of the climate scenarios on the sensitivity analysis results, as quantified by the TDCC values, was much higher for Morris than for EFAST, thus the latter was confirmed to be a more robust method [29, 40], providing more stable rankings of the sensitivity index across scenarios. The overall TDCC (across all years and sites) was 0.47 for Aquacrop with Morris and 0.71 for EFAST, for SAFYE it was 0.92 with Morris and 0.93 with EFAST. The agreement between the rankings of the sensitivity

indices obtained with the two methods, respectively μ^* and ST_i , for the same set of parameters (i.e. only those in both analyses), was further assessed by using the TDCC across methods for each scenario. It emerged, as also illustrated by comparing Fig 3 with Fig 5 and Fig 4 with Fig 6, that some agreement existed between the rankings of the Morris and EFAST methods. The TDCC varied between a minimum of 0.56 for the dry year at Xiaotangshan to a maximum of 0.96 for SAFYE in the wet year at the Yangling site. On average the agreement between the rankings of the two methods was higher for SAFYE, with a mean TDCC of 0.91, than for Aquacrop which had a mean TDCC of 0.74.

However, it should be noted that it can be expected that the rankings obtained by using the Morris method are not stable. As illustrated by [9], a much higher number of trajectories than the 25 used in this study, would be required to have more stable results, possibly more than 400, making the computing excessively time consuming. For these reasons, the Morris method, as reported by [9] and [40], can only be considered useful for screening out negligible factors, whereas its sensitivity rankings can only be considered as qualitative and not quantitative.

In this context, with reference to the Aquacrop model, the set of negligible parameters, identified in this study by the Morris method, coincides almost completely with the one identified by Vanyutrecht et al. [9]. The most notable differences concern the soil curve number (*CN*) and the shape factor for water stress coefficient inducing early senescence (*psenshp*), which were classified as negligible in the present study, whereas they were found to be very influential by [9]. The *CN* is used by Aquacrop for the estimation of the amount of rainfall lost by surface runoff [38], so it is understandable to have had a limited impact, in the water limited situations of the scenarios used in the present study. In Aquacrop the shape of the water stress coefficient curve can be selected as linear or concave by changing *psenshp*. As observed by [9], shape parameters are highly intertwined with the critical level parameters of the stress coefficients. However, the former have no physical

meaning, whereas the latter have physical sense and can more easily be calibrated. In our case the parameter regulating the specific stress coefficient, i.e. soil water depletion threshold for triggering early canopy senescence (*psen*), was found to be very influential, similarly to what found by [9].

The set of Aquacrop negligible parameters listed in Table 2 have, in all the scenarios tested in the present study, a limited impact on the variation of the output, so it is possible to exclude them from calibration or assimilation methods, or sensitivity analysis for climatic scenarios similar to those analyzed in this study.

For SAFYE, the results of the Morris method indicated very clearly that all the parameters and input factors describing the water balance and related stress function, introduced into the original SAFY model (to turn it into SAFYE), were negligible, except for soil moisture at field capacity and wilting point (Table 3). The latter two parameters, however, were not in the top ranking of the EFAST analysis (Fig. 8). The implication of this finding could be that, if the purpose of the study is only the assessment of wheat yield, even in conditions of severe water stress the original SAFY model could be appropriate. Indeed it would be even a better choice than SAFYE, since it relies on a more limited number of parameters and input factors. If, on the contrary, information is sought on the incidence and patterns of water stress, separately from other stress factors (e.g. nitrogen), SAFYE should be preferred, though a simplification of its water stress components could be suggested.

In agreement with the literature [9, 52], the results of this study suggest confirm that the application of the Morris method proved to be efficient for the identification of the negligible parameters, but not equally so for the classification of the parameters'

sensitivity. Looking at the ranking of the parameters with respect to the Morris μ^* (Fig 3 and 4) it is possible to note that the weather conditions had a more significant effect than for EFAST (Fig 5 and 6), as confirmed by the higher TDCC of the former tests. The ranking of parameters sensitivity is expected to be more reliable for EFAST, as previously mentioned, since the TDCC values across sites and years were much higher than for Morris. For both models it was possible to establish a set of influential parameters for all considered scenarios (Tables 4 and 5). EFAST allowed also to discriminate the effect of first order effects of single factors from second order effects and interactions on the output variance, revealing interesting difference among the models.

For Aquacrop it was found that very few factors had a strong first order effect, so that their impact on the grain yield was mainly due to second order effects and interactions with other parameters (Fig. 6), revealing a rather complex model behavior. A higher predominance of first order effects was found by [9], however with different crop species and with more diversified climate scenarios from those we used, possibly highlighting more clearly first order effects. For SAFYE first order effects had a stronger impact on yield variance (Fig. 8).

The ranking of the total sensitivity index for Aquacrop (Fig. 6) was different from the one obtained by [9], since as mentioned earlier, this reflects the relative importance of specific factors addressing processes occurring in the different climatic scenarios. In our case, the factors that had the highest impact on yield variance were mainly those related to water stress as well as phenological parameters regulating the development of the green canopy cover.

For SAFYE, the most influential factors were those related to the development of leaf area and leaf biomass. This would confirm that, in this model, the impact of water stress on crop yield is modeled indirectly from its impact on LAI, rather than explicitly through the water stress parameters introduced into the original SAFY model.

For both models used, Aquacrop and SAFY, it appeared that the parameters describing the phenological cycle of the crop played a decisive influence on the estimation of crop yield, being influential for both models. The coincidence between phenological stages and periods of the year with certain temperature and rainfall seems to be among the causes that most affect the variation in model output, especially for Aquacrop.

From the literature [44] it is known that choices made on the ranges of parameters and input variables influence the ranking of the sensitivity indices. For both Aquacrop and SAFYE these ranges were chosen after a careful inspection of the literature (S1 Table and S2 Table), within physically plausible ranges, expected to be found when the models are applied to regional scale studies. In this case, it is important to include also input variables in the SA, in addition to model parameters, because it is interesting to know, especially in a regional scale model application, what is their impact on the output. Many authors have made it clear that SA should not be necessarily strictly limited to parameters, but can also include input variables (see e.g. Wallach et al. [16] or Saltelli et al. [28]).

For Aquacrop we decided to test in the SA also conservative parameters [38], that are supposed to be fixed for a given species. Actually, the calibration results of different studies on wheat [7-8, 10-12] suggest that even for these parameters a limited range of variation is needed, thus we adopted for these parameters a range resulting from the values used these studies (with a maximum of $\pm 33\%$), including them in the SA, as also done by Vanuytrecht et al. [9].

As expected, the SA carried out for Aquacrop was more complex than that done for SAFYE. Aquacrop describes more accurately than SAFYE the crop growth processes, using a higher number of equations and input variables. Therefore the SA performed for

the former model had a much higher computational cost than for the latter. Using a computer with a Intel(R) Core(TM) i7-4770 CPU processor at 3.40 GHz, and 16 GB of RAM, the computation time for applying EFAST to Aquacrop was in the order of hours, while for SAFYE in the order of minutes. SA is a step required for any application of assimilation or calibration, thus the computational cost could be a factor that affects the choice of the model to use.

The SA carried out by this study allowed to calculate the plasticity and the complexity of the two crop models. Both complexity and plasticity were found to be higher for Aquacrop than for SAFYE (Table 6). The plasticity quantifies how much the climatic conditions change the influence of specific groups of parameters on the output. Confalonieri et al. [44], in a comparison of three crop models, found that the model having the highest complexity (WOFOST) was the one with the highest plasticity, but also the lowest robustness. They considered that a high value of plasticity was a positive aspect for a crop model, since it reflects its capability to change behavior in response to diverse environmental conditions, seeming to reflect the phenotypic plasticity typical of crops. However, in the context of data assimilation, or of applications of crop models at a regional scale, a high plasticity would be a nuisance, since it would not allow to clearly identify for which parameters the effort to reduce their uncertainty needs to be focused. A high plasticity would limit the generalization capability of a crop model, strengthening the need for site-specific calibrations and sensitivity analyses. In the present work, the plasticity for Aquacrop ($L=0.32$) was even higher than that found by [44] for the model having the highest plasticity among those tested, i.e. WOFOST ($L=0.37$). Instead, SAFYE had a lower plasticity, with an L of 0.42, i.e. intermediate between that of the models Cropsyst ($L=0.40$) and WARM ($L=0.44$).

Conclusions

The results reported in this study contribute to the knowledge on the behavior, in water limited climatic scenarios, of two models, Aquacrop and SAFY, increasingly used for large scale studies [14, 24, 26, 48-51]. Under these contexts, higher uncertainties exist on the values of parameters and input factors needed to run these models, as compared to more usual field based applications of crop models. In the case of large scale applications, it is very useful to know, not only the sensitivity of the targeted model outputs, grain yield in our case, to the parameters and input variables, but also model complexity and plasticity. Sensitivity analysis results could provide guidance on the data gathering and calibration efforts, but the knowledge of complexity and plasticity could even provide elements on the choice of the most suitable model, based on the information available and on the processes that need to be described.

In our case, it was apparent that Aquacrop, despite being simpler than most other crop models, is nevertheless more complex than SAFYE and it has also a higher plasticity. The implications would be that in a regional scale application of Aquacrop, more attention should be paid to its calibration than for SAFYE, which has a smaller number of influential factors. Furthermore, given the higher plasticity of Aquacrop, it would be advisable, more than for SAFYE, to carry out a preliminary sensitivity analysis, in case the scenario in which it would be used differs markedly from those of sensitivity studies reported in the literature. The present study employed quite a diversified range of climatic scenarios, characterized by moderate to severe water stress, so the results should valid for similar situations.

SAFYE was found to be less complex and with lower plasticity than Aquacrop. It was found that the parameters that were added to the original SAFY model, to include

water balance and crop water stress processes, were mostly negligible, with the notable exception of soil moisture at field capacity and wilting point. This suggests that a simplification of the description of these processes in SAFYE could be made. The limited number of parameters and the lower plasticity, makes this model perhaps more suitable than Aquacrop to large scale applications, whenever limited information on the input variables is available and an estimate of crop yield is the only information sought. However it should be mentioned that this advantage is counterbalanced by the fact that this model would lack the ability of providing insight on the processes that contributed to yield reduction, that would only be made clear by models incorporating the description of influence of the different abiotic (or even biotic) stress factors on crop growth.

Acknowledgements

This study was supported by a grant to PCS PhD scholarship under ESA (European Space Agency) Contract No. 4000110311/14/I-BG, in the framework of Dragon 3 Project ID 10448 "Farmland Drought" and by the CNR contribution in the framework of the Agreement on scientific cooperation between CNR and CAS China 2014-2016. The authors would like to thank Dirk Raes for providing the command-line version of Aquacrop and Samuel Buis and other colleagues at INRA Avignon (France) for providing the Matlab code for the GSA.

References

1. Mishra AK, Singh VP. A review of drought concepts. *J Hydrol.* 2010; 391(1-2):202–216. doi:10.1016/j.jhydrol.2010.07.012
2. Anderson-Wilk M. Rising expectations for conservation in water-scarce times. *J Soil Water Conserv.* 2008; 63:4A.
3. Steduto P, Hsiao TC, Raes D, Fereres E. AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agron J.* 2009; 101(3):426-437. doi:10.2134/agronj2008.0139s
4. Foster T, Brozović N, Butler AP, Neale CMU, Raes D, Steduto P, Fereres E, Hsiao TC. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agric Water Manag.* 2017; 181:8-22
5. Todorovic M, Albrizio R, Zivotic L, Abi Saab M, Stöckle C, Steduto P. Assessment of AquaCrop, CropSyst, and WOFOST models in the simulation of sunflower growth under different water regimes. *Agron J.* 2009; 101:509–521
6. Abi Saab MT, Todorovic M, Albrizio R. Comparing AquaCrop and CropSyst models in simulating barley growth and yield under different water and nitrogen regimes. Does calibration year influence the performance of crop growth models? *Agric Water Manag.* 2015; 147:21–33
7. Xiangxiang W, Qianjiu W, Jun F, Qiuping F. Evaluation of the AquaCrop model for simulating the impact of water deficits and different irrigation regimes on the

- biomass and yield of winter wheat grown on China's Loess Plateau. *Agric Water Manag.* 2013; 129:95–104. doi:10.1016/j.agwat.2013.07.010
8. Jin X, Feng H, Zhu X, Li Z, Song S, Song X, et al. Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PLoS ONE* 2014; 9(1):e86938. doi:10.1371/journal.pone.0086938
 9. Vanuytrecht E, Raes D, Willems P. Global sensitivity analysis of yield output from the water productivity model. *Environ Modell Softw.* 2014; 51:323–332. [doi:10.1016/j.envsoft.2013.10.017](https://doi.org/10.1016/j.envsoft.2013.10.017)
 10. Mkhabela MS , Bullock PR. Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. *Agric Water Manag.* 2012; 110:16–24.
 11. Iqbal MA, Shen Y, Stricevic R, Pei H, Sun H, Amiri E et al. Evaluation of the FAO AquaCrop model for winter wheat on the North China Plain under deficit irrigation from field experiment to regional yield simulation. *Agric Water Manag.* 2014; 135:61-72.
 12. Andarzian B, Bannayan M, Steduto P, Mazraeh H, Barati MA, Rahnama A. Validation and testing of the AquaCrop model under full and deficit irrigated wheat production in Iran. *Agric Water Manag.* 2011; 100:1–8.
 13. Jin, X.; Kumar, L.; Xu, X.; Yang, G.; Wang, J. Estimation of winter wheat biomass and yield by combining the Aquacrop model and field hyperspectral data. *Remote Sens.* 2016, 8, 972
 14. Jin, X.; Li, Z.; Yang, G.; Yang, H.; Feng, H.; Xu, X.; Wang, J.; Li, X.; Luo, J. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS J. Photogramm. Remote Sens.* 2017, 126, 24–37.
 15. Castañeda-Vera A, Leffelaar PA, Álvaro-Fuentes J, Cantero-Martínez C, Mínguez MI. Selecting crop models for decision making in wheat insurance. *Eur J Agron.* 2015; 68:97–116. doi:10.1016/j.eja.2015.04.008
 16. Wallach D, Makowski D, Jones JW editors. Working with dynamic crop models. evaluation analysis, parameterization and applications. Amsterdam: Elsevier Science; 2006.
 17. Confalonieri R, Bregaglio S, Adam M, Ruget F, Li T, Hasegawa T et al. A taxonomy-based approach to shed light on the babel of mathematical models for rice simulations. *Envir. Modell. Softw.* 2016; 85:332-341.
 18. Confalonieri R, Orlando F, Paleari L, Stella T, Gilardelli C, Moredi E. et al. Uncertainty in crop model predictions: What is the role of users? *Environ. Modell. Softw.* 2016; 81:165-173.

19. Kersebaum KC, Boote KJ, Jorgenson JS, Nendel C, Bindi M, Frühauf C, et al. Analysis and classification of data sets for calibration and validation of agroecosystem models.. *Environ. Modell. Softw.* 2015; 72:402-417.
20. Confalonieri R, Bregaglio S, Acutis M. Quantifying uncertainty in crop model predictions due to the uncertainty in the observations used for calibration. *Ecol. Model.* 2016; 328:72-77.
21. Dorigo W, Zurita-Milla R, de Wit JW, Brazile J, Singh R, Schaepman ME. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int J Appl Earth Obs Geoinf.* 2007; 9(2):165–193. doi:10.1016/j.jag.2006.05.003
22. Ines AVM, Das NN, Hansen JW, Njoku EG. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens Environ.* 2013; 138:149–164. doi:10.1016/j.rse.2013.07.018
23. Duchemin B, Maisongrande P, Boulet G, Benhadj I. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environ Model Softw.* 2008; 23(7):876–892.
[doi:10.1016/j.envsoft.2007.10.003](https://doi.org/10.1016/j.envsoft.2007.10.003)
24. Chahbi A, Zribi M, Lili-Chabaane Z, Duchemin B, Shabou M, Mougenot B, et al. Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the SAFY growth model. *Int J Remote Sens.* 2014; 35(3):1004–1028. doi: 10.1080/01431161.2013.875629
25. Duchemin B, Boulet G, Maisongrande, BenHadj I, Hadria R, Khabba S et al. Un modèle simplifié pour l'estimation du bilan hydrique et du rendement de cultures céréalières en milieu semi-aride. In: *Proceedings of "Deuxième Congrès Méditerranéen Ressources en Eau dans le Bassin Méditerranéen - WATMED "*, Marrakech, 14-17 November 2005, 14-17.
26. Veloso A. Modélisation spatialisée de la production, des flux et des bilans de carbone et d'eau des cultures de blé à l'aide de données de télédétection : application au sud-ouest de la France. PhD Thesis, Université Toulouse III- Paul Sabatier, Toulouse, 2014. Available from <http://www.sudoc.fr/188547614>
27. Baret F, Guyot G. Potentials and limits of vegetation indices for LAI and APAR assessment *Remote Sens Environ.* 1991; 35: 161-173.
28. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis: The Primer*. New York: Wiley; 2008.
29. Cariboni J, Gatelli D, Liska R, Saltelli A. The role of sensitivity analysis in ecological modelling. *Ecol Model.* 2007; 203(1-2):167–182.
30. Sobol IM. Asymmetric convergence of approximations of the Monte Carlo method. *Comput Math Math Phys.* 1993; 33(10):1391–1396.

31. Cukier R., Levine H., Shuler K. Nonlinear sensitivity analysis of multiparameter model systems. *J Comput Phys.* 1978; 26(1):1–42. doi: 10. 1016/0021-9991(78)90097-9
32. Saltelli A, Tarantola S, Chan K. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics.* 1999; 41(1):39–56.
33. Morris MD. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* 1991; 33(2):161–74.
34. Confalonieri R, Bellocchi G, Bregaglio S, Donatelli M, Acutis M. Comparison of sensitivity analysis techniques: a case study with the rice model WARM. *Ecol Model.* 2010; 221(16):1897–1906. [doi:10.1016/j.ecolmodel.2010.04.021](https://doi.org/10.1016/j.ecolmodel.2010.04.021)
35. Chen Y, Cournède PH. Data assimilation to reduce uncertainty of crop model prediction with Convolution Particle Filtering. *Ecol Model.* 2014; 290:165–77. [doi:10.1016/j.ecolmodel.2014.01.030](https://doi.org/10.1016/j.ecolmodel.2014.01.030)
36. Bristow KL, Campbell GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric For Meteorol.* 1984; 31(2):159–166. doi:10.1016/0168-1923(84)90017-0
37. Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration: guidelines for computing crop water requirements. *Irrig Drain Pap.* 56; 1998; Rome: Food and Agriculture Organisation of the United Nations (FAO).
38. Raes D, Steduto P, Hsiao TC, Fereres E. AquaCrop Version 5.0 Reference Manual. Annex I. Food and Agriculture Organization of the United Nations, Rome, Italy, 2016. Available from: <http://www.fao.org/land-water/databases-and-software/aquacrop/en/>
39. Monteith JL, Moss CJ. Climate and the Efficiency of Crop Production in Britain. *Philos. Trans. R. Soc. London Ser. B* 1977; 281:277–294.
40. DeJonge KC, Ascough JC, Ahmadi M, Andales A, Arabi M. Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecol Model.* 2012; 231:113–125. doi:10.1016/j.ecolmodel.2012.01.024
41. Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. *Environ Modell Softw.* 2007; 22(10):1509–1518. [doi:10.1016/j.envsoft.2006.10.004](https://doi.org/10.1016/j.envsoft.2006.10.004)
42. Buis S, Varella H, Gue M. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. *Environ Modell Softw.* 2010; 25:310–9.
43. Confalonieri R, Acutis M, Bellocchi G, Donatelli M. Multi-metric evaluation of the models WARM , CropSyst , and WOFOST for rice. *Ecol Model.* 2009; 220:1395–1410.

44. Confalonieri R, Bregaglio S, Acutis M. Quantifying plasticity in simulation models. *Ecol Model.* 2012; 225:159–166.
45. Iman RL, Conover WJ, Conover WJ. A measure of top-down correlation. *Technometrics* 1987; 29(3):351–357.
46. Confalonieri R, Bregaglio S, Acutis M. Short communication: A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. *Ecol Model.* 2010; 221(6):960–964. [doi:10.1016/j.ecolmodel.2009.12.003](https://doi.org/10.1016/j.ecolmodel.2009.12.003)
47. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 2007; 11:1633-1644.
48. Casa R, Silvestro PC, Yang H, Pignatti S, Pascucci S, Yang G. Assimilation of remotely sensed canopy variables into crop models for an assessment of drought-related yield losses: A comparison of models of different complexity. *International Geoscience and Remote Sensing Symposium (IGARSS)* 2016,7730548, pp. 5925-5928.
49. Silvestro PC, Pignatti S, Pascucci S, Yang H, Li Z, Yang G, Huang W, Casa R. Estimating winter wheat yield at the field and district scale from the assimilation of biophysical variables from satellite data into the Aquacrop and SAFY crop models. *Remote Sensing.* Forthcoming 2017.
50. Lorite IJ, García-Vila M, Santos C, Ruiz-Ramos M, Fereres E. AquaData and AquaGIS: Two computer utilities for temporal and spatial simulations of water-limited yield with AquaCrop. *Comput. Electron. Agric.* 2013; 96: 227-237.
51. Kim, D., Kaluarachchi, J. Validating FAO AquaCrop using Landsat images and regional crop information. *Agric Water Manag.* 2015; 149: 143-155.
52. Paleari L, Confalonieri R. Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions. *Ecol Model.* 2016; 340, 57-63.

Supporting Information

S1 Fig. Simulated and measured above ground biomass, grain yield and LAI for the Xiaotangshan winter wheat dataset [23] with different sowing dates in the years 2008-2011. Lines are data simulated with the SAFYE model, points are field measurements.

S1 Table. Parameters and input variables of the Aquacrop model considered in the sensitivity analysis, with their minimum and maximum values and the range of variation employed and the references used for setting these values. Conservative and non-conservative parameters and input variables are indicated.

S2 Table. Parameters and input variables of the SAFYE model considered in the sensitivity analysis, with their minimum and maximum values and the range of variation employed and the references used for setting these values. Conservative and non-conservative parameters and input variables are indicated.

Allegato C

Articoli pubblicati (atti di convegni)

1. Casa, R.; Silvestro, P. C.; Yang, H.; Pignatti, S.; Pascucci, S.; Yang, G. Assimilation of Remotely Sensed Canopy Variables Into Crop Models for an. In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International; IEEE: Beijing, 2016; pp. 5925–5928.
2. Silvestro, P. C.; Casa, R.; Pignatti, S.; Castaldi, F.; Yang, H.; Yang, G. SPATIALIZED APPLICATION OF REMOTELY SENSED DATA ASSIMILATION METHODS FOR FARMLAND DROUGHT MONITORING. In Proceedings of Dragon 3 Final Results and Dragon 4 Kick-Off; Ouwehand, L., Ed.; ESA-SP, 2016.
3. Guijun, Y.; Hao, Y.; Xiuliang, J.; Silvestro, P. C.; Stefano, P.; Raffaele, C. FARMLAND DROUGHT EVALUATION BASED ON THE ASSIMILATION OF MULTI- TEMPORAL MULTI-SOURCE REMOTE SENSING DATA INTO AQUACROP MODEL. In Proceedings of Dragon 3 Final Results and Dragon 4 Kick-Off; L. Ouwehand, Ed.; ESA-SP, 2016; Vol. 739, pp. 4–8.
4. Casa, R.; Silvestro, P. C.; Yang, H.; Pignatti, S.; Pascucci, S.; Yang, G.; Imaa, C. N. R.; Scalo, T. DEVELOPMENT OF FARMLAND DROUGHT ASSESSMENT TOOLS BASED ON THE ASSIMILATION OF REMOTELY SENSED CANOPY BIOPHYSICAL VARIABLES INTO CROP WATER RESPONSE MODELS Department of Agriculture, Forests, Nature and Energy (DAFNE), Università della Tuscia , Via San. In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International; IEEE: Milan, 2015; pp. 4005–4008.
5. Silvestro, P. C.; Casa, R.; Pignatti, S.; Castaldi, F.; Yang, H.; Yang, G. Development of an assimilation scheme for the estimation of drought-induced yield losses based on multi-source Remote Sensing and the Aquacrop model. In Dragon 3 Mid Term Results, Proceedings of the conference held 26-29 May, 2015 at Chengdu, China. Edited by L. Ouwehand. ESA SP Vol. 724; ESA SP Vol. 724. ISBN: 978-92-9092-288-9, I. 13., Ed.; Chengdu, 2014.

ASSIMILATION OF REMOTELY SENSED CANOPY VARIABLES INTO CROP MODELS FOR AN ASSESSMENT OF DROUGHT-RELATED YIELD LOSSES: A COMPARISON OF MODELS OF DIFFERENT COMPLEXITY

Casa R.¹, Silvestro P.C.¹, Yang H.^{2,4}, Pignatti S.³, Pascucci S.³, Yang G.^{2,4}

¹Department of Agriculture, Forests, Nature and Energy (DAFNE), Università della Tuscia Viterbo, Via San Camillo de Lellis, Viterbo, Italy (rcasa@unitus.it)

²National Engineering Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

³CNR IMAA, Tito Scalo, Potenza, Italy

⁴Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, China

ABSTRACT

The assimilation of biophysical crop canopy variables retrieved from remotely sensed data into two crop models of differing degree of complexity is assessed in this study, in the context of the development of tools suitable for the estimation of yield losses due to drought. The more complex AQUACROP model, developed by FAO and the simpler SAFY model were employed to estimate wheat grain yield for an area in the Shaanxi Province in China through the assimilation of biophysical variables retrieved from Landsat and HJ1A and HJ1B satellites for three growing seasons (2013 to 2015). Results were validated with ground yield data.

Index Terms— Wheat, AQUACROP, SAFY, Ensemble Kalman Filter

1. INTRODUCTION

Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop water response models is a promising methodology for the assessment of drought impact on crops, suitable for high spatial and temporal resolution applications [1]. In particular such methodology allows to estimate yield losses resulting from the occurrence of drought, provided that the crop models used do take into account the relevant crop physiological processes occurring during water stress. The most suitable models to study the impact of drought on crops must simulate crop growth according to the variables related to the availability of soil water, such as crop evapotranspiration.

For operational assimilation schemes, a balance must be found between the accurate description of processes leading to very complex crop models with hundreds of parameters and input variables and simpler models, with a more coarse or empirical description of the relevant processes [2]. The

former are usually more difficult to manage during assimilation, the latter could be more suitable operationally, with less uncertainty introduced during parameterization, but could be less accurate.

In the context of a Sino-European Dragon-3 program funded by ESA and the Ministry of Science and Technology (MOST) of the P.R. China, the "Farmland Drought" project deals with the development of crop variables assimilation schemes and their validation at the farmland scale for a test site in China. The work developed in this context focused on the application of a methodology of assimilation of biophysical variables into two crop models of different complexity and the comparison of validation results for an area in the Province of Shaanxi.

2. METHODOLOGY

Two models were studied: the more complex AquaCrop [3] and a modified version of the simpler SAFY [4] model.

Aquacrop was chosen because it was shown to reproduce quite well the yield response to water, for several crops and climates, including China [5], and because it employs CC as a main state variable, which is easier to estimate from remote sensing than LAI, not being subject to saturation at high values. Aquacrop uses a set of 71 parameters and the code is not accessible, so the assimilation procedure employed was based on the recalibration, through a numerical optimization technique (simplex method), of a limited set of parameters chosen according to a sensitivity analysis study [6]. The optimization algorithm developed minimizes the difference between CC simulated by the model and observed from remote sensing at each observation date. The final grain yield simulated by the model takes thus into account the observations characterizing the seasonal crop growth trend.

The simpler SAFY model, developed by Duchemin et al. [4], is based on Monteith's concept of conversion of absorbed photosynthetically active radiation into biomass,

then converted to grain yield using empirical partitioning coefficients. The original SAFY code was modified by adding a very simple water balance module according to the FAO 56 [7] equations.

The comparison between the two models was done using three different criteria: accuracy, complexity and plasticity. Accuracy is the ability of the model to fit reference measured data [8], in this case the yield, using root mean square error (RMSE) as sample index. Complexity is the parsimony of the model in representing the biophysical system, it gives information about amount and relevance of model parameters [9]. In this study it is evaluated using as sample index the proportion of relevant parameters (S) over all total number of model parameters (T):

$$R_p = \frac{S}{T}$$

Plasticity is the tendency of the model to change its behaviour when employed in different conditions [10], changes in model behaviour can be quantified using an indicator of model plasticity (L):

$$L = TDCC \cdot e^{\sigma_{SAM}^{-1}}$$

where TDCC is the top-down concordance coefficient [11] and σ_{SAM} is the standard deviation of a normalized agrometeorological indicator [10].

Models for complex biological systems comprise a large array of biophysical and physiological parameters and input variables that require the application of a Sensitivity Analysis (SA) to identify the contribution of each parameter to the system response. This is required since, typically during the assimilation phase, only a limited number of parameters and input variables can be tackled, so it's important to choose the most influential ones.

Therefore, in the first part of the work, the Morris and EFAST global sensitivity analysis methods [12] have been employed, to identify and rank the array of the model parameters utilized in the assimilation [6]. The results from the SA allowed to quantify the complexity and the plasticity of the two models.

The second part of the work concentrated on the retrieval of leaf and canopy scale biophysical variables related to water stress from remote sensing (RS) data. For this task, the methodologies employed make use of techniques based on radiative transfer that is applied to the crop spectral response mechanisms. In particular, algorithms based on the training of Artificial Neural Networks (ANN) with PROSAIL model [13, 14] have been adapted for the use with Chinese multispectral satellite (Landsat8, HJ1A and HJ1B satellites). The retrieved values of LAI and fractional ground canopy cover (CC) were compared with measurements collected in several wheat fields in the Yangling area at the time of HJ1A and HJ1B satellites overpass (Fig. 1).

The last part of the work consisted in setting up the procedure in which biophysical variables (i.e. LAI and CC) are tested for the assimilation into crop water response

models using assimilation methods based on the Ensemble Kalman Filter.

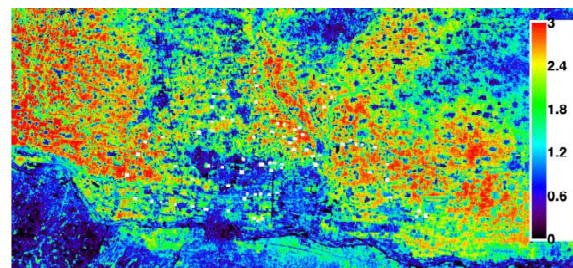


Fig. 1. Map of LAI for the Yangling area derived from a Landsat8 image of 7th April 2014 using ANN and PROSAIL. The white dots are the ground sampling points.

The application of assimilation methods allowed the calculation of models accuracy. For this model, the Ensemble Kalman Filter assimilation method was chosen [1]. Also in this case, a subset of the model parameters was chosen to be updated during the assimilation, according to the outputs of the sensitivity analysis.

These assimilation procedures, using both models, were used with data collected in Yangling (Shaanxi Province) in 2013, 2014 and 2015 and were validated for a number of wheat fields for which ground yield data had been recorded.

3. RESULTS AND DISCUSSION

From the SA results, parameters were considered relevant if the sum of EFAST Sensitivity indices was greater than a value of 0.15. This value was set exclusively to quantify and compare the complexity of Aquacrop and SAFY. It resulted that the R_p for Aquacrop was 0.11, indicating that the relevant parameters were on average 11% of the total, while for SAFY a R_p value of 0.08 was obtained.

SA also showed that the SAFY model had a plasticity index (L) equal to 0.169 while Aquacrop 0.299. This means that SAFY has a higher plasticity than Aquacrop (L ranges from 0 to about 1.51, with the highest plasticity at 0).

Model accuracy was assessed from the RMSE between the yield simulated by the assimilation into crop models and ground data (Fig. 2). Aquacrop had an RMSE of 0.62 t ha^{-1} , for the year 2013, but the inability to access the code forced us to use an assimilation method with a high computational cost. The simulations were made only for 2013, because the processing time for each point was too high. The assimilation method applied to SAFY for the same year had a RMSE of 0.78 t ha^{-1} . The accuracy was lower than for the assimilation method applied to Aquacrop, but the computational cost and the processing are much lower than the previous method. For the other years the estimation accuracy was worse than for 2013, particularly for 2015.

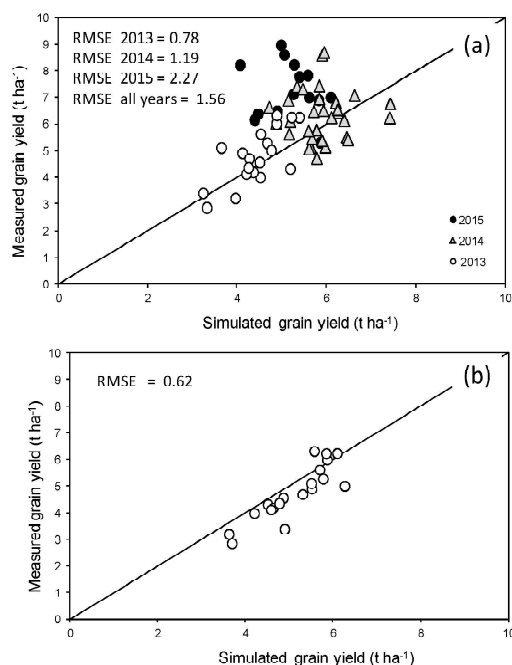


Fig. 2: SAFY (a) and AquaCrop (b) accuracy simulating wheat grain yield for the ground data of Yangling

4. CONCLUSION

Preliminary results with the method used with AquaCrop, have shown its adequacy to provide yield estimation as affected by water availability even though some drawbacks appear for the spatialized applications. Its highest complexity makes the calibration more difficult, the choice of values to assign to the parameters which remained fixed (not re-calibrated) was based on the literature and on a preliminary calibration, but for regional scale applications this could be a strong limitation. Its lower plasticity means that this model is less affected by the boundary conditions, such as weather or soil and management, which makes it less suitable to establish a relationship between drought and yield. Its high accuracy is a strong point but its higher computational cost makes it rather less suitable for a spatialized application.

Preliminary results with the simpler SAFY model seem to indicate that the EnKF assimilation procedure performed better, also given the lower number of parameters, and proved to be quite robust despite the lower complexity.

The validation results were satisfactory despite the problem caused by the small field sizes in Yangling which would have probably required higher spatial resolution data, as compared to the 30 m pixel size which were available.

However, the work carried out highlights the potential of remote sensing data assimilation with simple crop water response models in the assessment of drought-induced crop yield losses.

5. REFERENCES

- [1] Wallach, D. Makowski, D., Jones, J.W. 2006. Working with Dynamic Crop Models, Elsevier, 151-169.
- [2] Casa R., Varella H., Buis S., Guérif M., De Solan B., Baret F., 2012. Forcing a wheat crop model with LAI data to access agronomic variables: evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *European Journal of Agronomy*, 37, 1-10.
- [3] Steduto, P., Hsiao, T.C., Raes, D. & Fereres, E. (2009). AquaCrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agron. J.* 101(3), 426-437.
- [4] Duchemin B., Maisongrande P., Boulet G, Benhadj. I (2008). A simple algorithm for yield estimates: evaluation for semi-arid irrigated winter wheat monitored with ground-based remotely-sensed data. *Environmental Modelling and Software* 23, 876-892
- [5] Jin, X., Feng, H., Zhu, X., Li, Z., Song, S., Song, X., Guo, W., 2014. Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PloS One*, 9(1), e86938.
- [6] Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci S., Castaldi, F., Casa, R., 2016. Comparative global sensitivity analysis for two crop models of different complexity for the assessment of drought related yield losses. *Submitted to Agricultural Water Management*
- [7] Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, Rome, Italy.
- [8] Confalonieri, R., Acutis, M., Bellocchi, G., & Donatelli, M. (2009). Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice, 220, 1395–1410.
- [9] Fox, D. G. (1981). Judging Air Quality Model Performance. *Bulletin of the American Meteorological Society*. [http://doi.org/10.1175/1520-0477\(1981\)062<0599:JAQMP>2.0.CO;2](http://doi.org/10.1175/1520-0477(1981)062<0599:JAQMP>2.0.CO;2)
- [10] Confalonieri, R., Bregaglio, S., & Acutis, M. (2012). Quantifying plasticity in simulation models. *Ecological Modelling*, 225, 159–166.

- [11] Iman, R. L., Conover, W. J., & Conover, W. J. (1987). A Measure of Top-Down Correlation. *Technometrics*, 29, 351–357.
- [12] Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10), 1509–1518.
- [13] Jacquemoud S., Verhoef W., Baret F., Bacour C., Zarco-Tejada P.J., Asner G.P., Christophe F. & Ustin S.L. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization, *Remote Sens. Environ.* 113(1), S56–S66
- [14] Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., ... & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens. Environ.* 110(3), 275–286.

SPATIALIZED APPLICATION OF REMOTELY SENSED DATA ASSIMILATION METHODS FOR FARMLAND DROUGHT MONITORING USING TWO DIFFERENT CROP MODELS

Silvestro Paolo Cosmo ⁽¹⁾, Casa Raffaele ⁽¹⁾, Pignatti Stefano ⁽²⁾, Castaldi Fabio ⁽¹⁾, Yang Hao ⁽³⁾, Yang Guijun ⁽³⁾

⁽¹⁾ Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy), Email: rcasa@unitus.it; p.c.silvestro@unitus.it

⁽²⁾ Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy) Email: stefano.pignatti@cnr.it

⁽³⁾ National Engineering Research Center for Information Technology in Agriculture (NERCITA), 2449-26, Beijing 100097, P.R. China Email: yanggj@nercita.org.cn; yangh@nercita.org.cn

ABSTRACT

The aim of this work was to develop a tool to evaluate the effect of water stress on yield losses at the farmland and regional scale, by assimilating remotely sensed biophysical variables into crop growth models. Biophysical variables were retrieved from HJ1A, HJ1B and Landsat 8 images, using an algorithm based on the training of artificial neural networks on PROSAIL.

For the assimilation, two crop models of differing degree of complexity were used: Aquacrop and SAFY.

For Aquacrop, an optimization procedure to reduce the difference between the remotely sensed and simulated CC was developed. For the modified version of SAFY, the assimilation procedure was based on the Ensemble Kalman Filter.

These procedures were tested in a spatialized application, by using data collected in the rural area of Yangling (Shaanxi Province) between 2013 and 2015. Results were validated by utilizing yield data both from ground measurements and statistical survey.

1. INTRODUCTION

Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop water response models is a promising methodology for the assessment of drought impact on crops, suitable for high spatial and temporal resolution applications [1]. In particular such methodology allows to estimate yield losses resulting from the occurrence of drought, provided that the crop models used do take into account the relevant crop physiological processes occurring during water stress. The models suitable to study the impact of drought on crops should simulate crop growth taking into account factors such as the availability of soil water and its impact on crop evapotranspiration.

For operational assimilation schemes, a balance must be found between the accurate description of processes leading to very complex crop models with hundreds of parameters and input variables and simpler models, with a more coarse or empirical description of the relevant processes [2]. The former are usually more difficult to manage during assimilation, the latter could be more suitable operationally, with less uncertainty introduced

during parameterization, but could be less accurate.

In the context of Sino-European Dragon-3 program funded by ESA and the Ministry of Science and Technology (MOST) of the P.R. China, the "Farmland Drought" project dealt with the development of crop variables assimilation schemes and their validation at the farmland scale for a test site in China. The work developed in this context focused on the application of a methodology of assimilation of biophysical variables into two crop models of different complexity and the comparison of validation results for a rural area in the Province of Shaanxi.

2. STUDY AREA AND MATERIALS

2.1 In situ data

The study site used in this work is located in the rural area of Yangling, in the Province of Shaanxi, Central China.

The experimental site is characterized by a semi-arid climate with a long term average rainfall of 573 mm per year, a minimum temperature of -4°C in winter and a maximum temperature of 35°C. The climatic data sets included three growth seasons, from October 2012 to June 2014 and consisted of rainfall, temperatures (mean, maxima and minima), relative humidity and wind speed. The crop considered in this study is winter wheat, the main crop of the study area. Field measurements were carried out from March to June for each crop season. The details of field measurements are shown in Tab. 1.

Table 1. Field measurements of LAI (L), biomass (b) and yield (y) conducted in Yangling from 2013 to 2015. n. pts are the number of sample fields.

Year	n. pts	March	April	May	June
2013	49	30 Mar L, b 5 Mar,	27 Apr L, b 22 Apr,	16	1 Jun b,y
2014	35	28 Mar L, b	27 Apr L, b	May L, b	9 Jun b,y
2015	28	27 Mar L, b	25 Apr L,b		5 Jun b,y

They consisted of observations in 28 to 49 points

(different for each season), distributed in different wheat fields. Information about the soil properties, or winter wheat variety, or sowing date was not available.

2.2 Remote sensing data

The remote sensing data set consists of several images acquired by Landsat 8, HJ1A and HJ1B, during the three crop cycles (Tab. 2).

Table 2. Remote Sensing data set for Yangling, 2013-2016 crop cycles

Year	Month	Day	Satellite
2013	March	5	HJ1B
2013	March	28	HJ1B
2013	April	26	HJ1A
2013	May	11	HJ1A
2014	March	15	Landsat8
2014	April	7	HJ1A
2014	April	29	HJ1A
2014	May	20	Landsat8
2015	February	23	Landsat8
2015	March	29	HJ1A
2015	April	28	HJ1B

Images acquired by Landsat 8 were provided by the United States Geological Survey with atmospheric and geographical correction, as reflectance values. Images acquired by HJ were corrected geographically identifying a series of ground control points on the map and warping the images. The HJ data were converted from Digital Number (DN) to radiance using the coefficients provided in the metadata. Once the image data were converted into radiance, atmospheric correction was performed using the FLAASH atmospheric correction module in ENVI (Harris, FL, USA). A winter middle latitude atmospheric model was used for images acquired between January and March, while a summer middle latitude model was used for images acquired between April and June. The selected aerosol model was rural with no aerosol retrieval and an initial visibility between 20 km and 40 km, depending on the image.

3. METHODS

In order to obtain an estimation of yield and to evaluate the effect of drought, two crop growth models were chosen among those able to correctly represent crop response to water stress and suitable for regional scale applications. We compared two models with respect to three different criteria: complexity, plasticity and accuracy. Subsequently the assimilation of biophysical variables retrieved from remote sensing was carried out, validating the simulated yield with point field measurements. Yield values were also compared with those published by the National Bureau of Statistics of the People's Republic of China.

3.1 Crop Models

The two models considered are Aquacrop [3] and SAFY, the Simple Algorithm For Yield [4]. Aquacrop is a "water driven" productivity model that simulates Canopy Cover (CC), biomass and yield of a crop mainly as a function of the water productivity, i.e. the biomass produced per unit of water transpired by the vegetation [3]. A more detailed explanation about how the Aquacrop model works is presented by Raes et al. (2009) [5]. SAFY is based on the Monteith concept [6], which links the production of total dry phytomass to the photosynthetically active portion of solar radiation (PAR) absorbed by the crop. A detailed description of SAFY was presented by [4]. In this work a modified version of SAFY was used, introducing a dependence of biomass yield on crop water stress, as described by the FAO irrigation and drainage paper n. 56 [7].

3.2 Comparison Criteria: Complexity, Plasticity and Accuracy

Complexity is the parsimony of the model in representing the biophysical system, it gives information about amount and relevance of model parameters [8]. In this study it is evaluated using as sample index the proportion of relevant parameters (S) over all total number of model parameters (T):

$$R_p = \frac{S}{T} \quad (1)$$

Plasticity is the tendency of the model to change its behaviour when employed in different conditions [9]. Changes in model behaviour can be quantified using an indicator of model plasticity (L):

$$L = TDCC \cdot e^{\sigma_{SAM}^{-1}} \quad (2)$$

where TDCC is the top-down concordance coefficient and σ_{SAM} is the standard deviation of the normalized agrometeorological indicator.

Accuracy is the ability of the model to fit reference measured data [10], in this case the yield, using root mean square error (RMSE) as a sample index. The accuracy of the model is evaluated by measuring the RMSE between yield measured in the field and that estimated by the models. It has been reported ([11], [4]) that the accuracy of Aquacrop is higher than SAFY, but additional considerations apply for an application at a regional scale.

3.3 Global Sensitivity Analysis

Most assimilation methods update the parameters on the basis of observations of the state variables. For computational requirements the number of parameters should be limited. Sensitivity analysis can help to identify the parameters that most affect the output of the model. We used a combination of two different Global Sensitivity Analysis (GSA) methods: Morris and EFAST ([9]; [12]). The first is a screening method and it is used to identify negligible parameters, taking

advantage of its high computational speed. In this way, negligible parameters were excluded from the SA performed with the second method, in order to reduce its running time. EFAST is a variance based method, its accuracy is higher than Morris and using a reduced set of parameters the computational time is reduced, balancing accuracy and running speed. The SA is also useful to quantify the complexity and the plasticity of the models

3.4 Estimation of LAI and Canopy Cover

LAI and Canopy Cover were calculated using the algorithm developed by Baret et al. (2006) [13] which is based on the training of artificial neural networks (ANN) over a data base simulated using radiative transfer models. A coupling of PROSPECT and SAIL models [14] was adopted to simulate the reflectance spectra of vegetation, in this case winter wheat.

A wide range of possible canopy scenarios are described by the combination of 14 parameters which relate to canopy, leaves, soil and observation geometry. Selecting and sampling the range for each parameters, the model PROSAIL generates a sufficient number of simulation (in this study 41472) to describe all expected scenarios.

The set of simulated reflectance is resampled according to the characteristics of the used sensor, then it is divided in two parts, one of them with the purpose of training ANNs and the other to evaluate the theoretical performance.

To make the simulations more realistic and reduce the ANN training hyper-specialization, a Gaussian noise error is added on the basis of an additive and multiplicative model. After the training of the neural networks, the coefficients and bias are stored to be used later in the estimation of variables LAI, FAPAR and FCOVER from reflectance observed values.

3.5 Ensemble Kalman Filter Assimilation on SAFY

In order to improve the accuracy of SAFY for simulating Yield, an EnKF based assimilation method was adopted [1]. The first step of this method is the individuation of main parameters, as previously done by means of the SA. The ranking of the most important parameters (and relative range of variation) is shown in Tab. 3.

Table 3. List of the most influential parameters of the SAFY model, as detected by the SA, allowed to vary during the assimilation.

Parameter	Min	Max	Description
-----------	-----	-----	-------------

Pgro_R2P	0.7	0.9	Climatic efficiency: ratio of incoming photosynthetically active radiation (PAR) to global radiation
Pgro_Lue	1.75	1.95	Effective light-use efficiency: ratio of energy produced as DAM from APAR (g.MJ ⁻¹)
Pgro_P2G	0.0098	0.012	Partition coefficient to grain
Pfen_MrgD	288	300	Day of the year of emergence
Pfen_PrtA	0.16	0.2	Make vary the origin and slope of the partitioning function
Pfen_SenA	1260	1540	Temperature threshold to start senescence (°C)
Ptfn_Topt	20	22	Optimum temperature for plant development (°C)
Da	1.35	1.55	Soil bulk density (g cm ⁻³)
SMT_sen	2070	2530	GDD to complete senescence

All the model parameters were calibrated using, as a reference, 5 points of the measurement field campaign of Yangling 2013. Even the parameters that vary in the assimilation were initially calibrated in this way. The calibrated values are used as mean values of the range in which the parameters vary, the minimum and maximum of the ranges are chosen arbitrarily in a surrounding of the mean value. It is possible to divide the assimilation algorithm into the following steps:

- 1) The algorithm generates an ensemble of N (in this case N=100) values of parameters P_p

$$P_k^j = P_k + \delta_k^j \quad (3)$$

where:

$j=[1,...,N]$

N= number of ensemble elements (100 in this case)

k= number of parameters δ =error drawn from $N(0,\sigma)$

- 2) Running SAFY in order to obtain N values of LAI for the first observation date with the addition of an error ε
- 3) The algorithm generates an ensemble of N observations from RS data:

$$M_{t_i}^j = M_{t_i} + \tau_{t_i}^j \quad (4)$$

where M is the set of state variables measurements, t_i is the index in the temporal series, τ is the measurements error drawn from $N(0,\sigma)$

- 4) Calculating the Kalman gain (using the covariance matrix)

$$K_{t_i} = \frac{\sum_{t_i} R^T}{R \sum_{t_i} R^T R + \text{var}(\tau_{t_i})}$$

- 5) Application of the Kalman filter to obtain a corrected LAI and parameters values estimated
- 6) Replacement of LAI simulated with the LAI calculated at the previous step, and further running of the model.
- 7) At each observation date repeat from step 3). When the last observation has been assimilated SAFY runs to the end and outputs the yield.

3.6 Optimization method on Aquacrop

Aquacrop is a water-driven crop growth model, for this reason it is linked with water balance and the yield simulated by this model is strongly influenced by water stress. These characteristics make it suitable to describe the links between drought and yield. Another advantage of Aquacrop is the use of CC as a main state variable which turns out to be estimated more accurately than LAI from remote sensing. The problems related to Aquacrop are mainly due to its complexity and the unavailability of the code to the scientific community. For these two reasons it was not possible to perform an updating assimilation of the state variable (as done for SAFY). The assimilation method employed was based on the recalibration, through a numerical optimization technique (simplex method), of a limited set of parameters chosen according to the sensitivity analysis study. The first 4 steps of this algorithm are the same of the previous used with SAFY, the difference is at the fifth step, the simplex optimization method is applied. This method finds the best parameters which minimize the following function:

$$MSE = \sum_{i=1}^n \frac{(CC_s(i) - CC_m(i))^2}{n} \quad (6)$$

where:

MSE = Mean Square Error

n = numbers of parameters

CC_s = simulated Canopy Cover

CC_m = measured Canopy Cover

3.7 Regional Scale Application

The application at the regional scale was done for the winter wheat growth cycle of Yangling of 2013 and 2014. We chose these two years because it was possible to validate the results with official yield data made available by the Shaanxi Provincial Bureau of Statistics and the National Bureau of Statistics of the People's Republic of China.

For each year the biophysical variables were assimilated using the values retrieved from 3 images acquired between March and May. For each image, only the wheat fields included in the rural area of Yangling were considered. To identify the pixels in which wheat is present, 5 images acquired between February and June (one for each month) have been used. The images were

converted into LAI maps through the PROSAIL-ANN algorithm. The pixels of each image have been considered wheat if the following conditions were true:

- LAI of the 1st date smaller than LAI of 2nd date and this latter smaller than LAI of 3rd date
- LAI of 3rd date higher than 4
- LAI of 4th date higher than 3.5
- LAI of 5th date smaller than 0.5 (after harvesting)

After the application of a mask to limit the assimilation only to wheat fields, each image map contained approximately 600000 pixels, which made a pixel-by-pixel assimilation not feasible. To reduce the computational cost of the algorithm, a classification of the pixels based on the LAI of the images that were used in the assimilation was employed. The images values of LAI were rounded at the first decimal place, if the number in that place was odd, it was rounded at the nearest lower even number. In this way the range between the minimum and maximum value of LAI was divided into steps of 0.2. The total number of classes is the product of the number of steps between the minimum feature value of LAI for each image and the maximum. Each class is defined by the combination of 3 numbers, 0.2 multiples, belonging to the range min LAI-max LAI of each image. In this way, after running the algorithm for all the combination of LAI, it is possible to assign a value of yield at each class. Classifying the points of map at each pixel the corresponding yield value was assigned. In this way it was possible to obtain yield maps of the region.

4. RESULTS AND DISCUSSION

4.1 Complexity and Plasticity results

The complexity of the two crop models tested was quantified using the parameter ratio (R_p). The parameters were considered relevant if the sum of EFAST Sensitivity indices was greater than an arbitrary threshold value fixed at 0.15. This value was set exclusively to quantify and compare the complexity of Aquacrop and SAFY. It resulted that the R_p for Aquacrop was 0.11, indicating that the relevant parameters were on average 11% of the total, while for SAFY a R_p value of 0.08 was obtained. The complexity of Aquacrop resulted to be higher than SAFY, i.e. the former uses a higher number of equations and input parameters than the latter, making the calibration more complex and the running time longer.

The plasticity of Aquacrop and SAFY was assessed by calculating the L index. It was found that SAFY has a plasticity index (L) of 0.169 while the L index for Aquacrop is 0.299. This means that SAFY has a higher plasticity than Aquacrop (L ranges from 0 to about 1.51, with the highest plasticity at 0). The plasticity quantifies how much the climatic conditions change the influence of specific groups of parameters on the output. A high plasticity would limit the generalization

capability of a crop model, strengthening the need for site-specific calibrations and sensitivity analyses.

4.2 LAI and Canopy Cover validation

The ANN algorithm which use for training the simulation done by PROSAIL model [12] was applied on several images acquired by HJ1A, HJ1B and Landsat 8 (Tab. 2) in 3 years (2013-2015) from March to May of each years. This procedure allowed to obtain the LAI and Canopy Cover maps for the area of interest, which were compared to field measurements for validation.

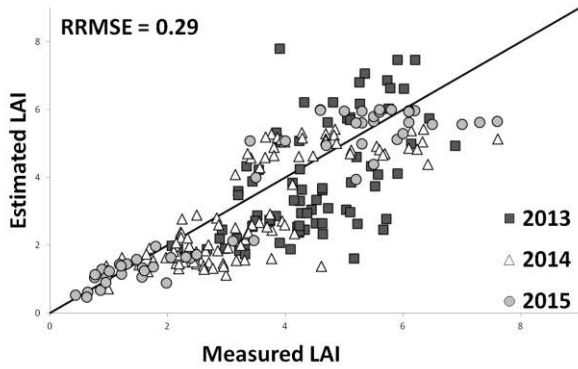


Figure 1. Validation results for the retrieval of LAI from HJ1A, HJ1B and Landsat 8 images, using field measurements of 3 years in Yangling rural area.

In Fig. 1 the comparison between LAI retrieved from the satellite images, by the PROSAIL-ANN algorithm, and the LAI measured in field campaigns is shown. The results indicate that the algorithm tends to underestimate the value of LAI compared to the field measurements, especially for the 2013 images. This depends on the quality of the images. The region of interest is strongly characterized by the presence of clouds and fog, so the probability for acquiring clear image is very low. The atmospheric correction for these images is fundamental and it strongly influences the results.

The estimation error is greater for higher values of LAI. In fact for LAI between 0 and 2 the difference between estimates and measures is low (Fig.1). Between 3 to 4 there is the highest error, for LAI values higher than 4 the error between simulations and measurements is lower again. The total relative root mean squared error (RRMSE) was of 0.29 %.

Fig. 2 shows the results obtained for Canopy Cover. The field measured LAI values were converted into Canopy Cover using the relationship between CC and LAI for wheat proposed by by Nielsen et al. (2011) [15]:

$$CC = 94 \cdot [1 - \exp(-0.43 \cdot LAI)]^{0.52} \quad (7)$$

In this case the RRMSE was lower than for LAI, i.e. 10% against 29% for LAI. Saturation issues affects less the estimation of CC. So the use of CC would be

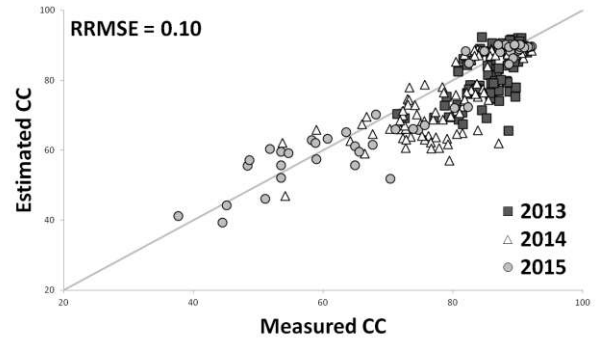


Figure 2. Validation results for the retrieval of canopy cover (CC) from HJ1A, HJ1B and Landsat 8 images, using field measurements of 3 years in Yangling rural area. The field measured LAI was converted into CC using the equation proposed by Nielsen et al. (2011)

preferable as a biophysical variable to be used in the assimilation.

These validation results provided information on the estimation error of the observed biophysical variables, which was considered during the assimilation.

4.3 Point application of EnKF Assimilation: comparison of two different models

As described in the paragraphs 3.5 and 3.6, the inaccessibility of the Aquacrop code did not permit the application of an updating assimilation method, as for SAFY, but it obliged to use an optimization method. The assimilation has been applied on the points where the field measurements were carried out, in this way it was possible to test the two different methods, comparing their accuracy after the assimilation and their performances.

In Fig. 3 the accuracy of the yield estimates obtained after the assimilation are shown for SAFY and Aquacrop.

There was a considerable difference in the computational time. In fact, using a PC with an Intel Core i7-4770 CPU at 3.4 GHz, for the optimization method with Aquacrop it was of around 40 minutes for each simulated data point, against the 30 seconds of the EnKF with SAFY.

For these reasons, whilst the EnKF-SAFY method was applied at every field measurements points (for 3 different years), the Aquacrop optimization algorithm was applied only to a limited sample of points (Fig. 3b). Considering only the points analyzed with the Aquacrop optimization method, the RRMSE for the EnKF-SAFY method was lower to 15.3%. Therefore the accuracy of EnKF-SAFY method for a limited sample of points is higher, albeit only slightly, than that of Aquacrop optimization method.

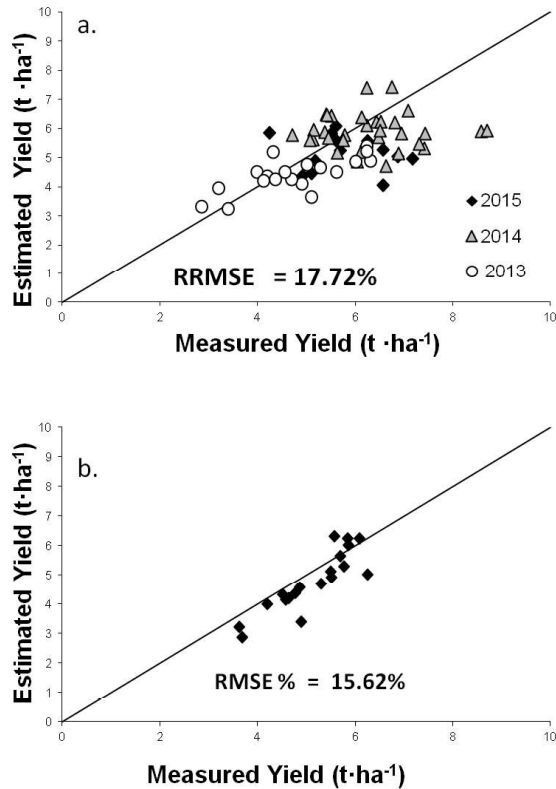


Figure 2. Point application: a) Comparison between Measured and SAFY_EnKF Simulated Yield. b) Comparison between Measured and Aquacrop Optimization Simulated Yield.

4.4 Regional scale application results

The spatialised application of the assimilation was carried out only for SAFY, because of the above mentioned computational constraints with Aquacrop.

Fig. 4 shows the Yangling winter wheat yield maps estimated for 2013 and 2014.

The mean yield calculated for 2013 was 5.395 t ha⁻¹, whereas the valued provided by the Shaanxi Provincial Bureau of Statistics was 5.495 t·ha⁻¹.

For 2014, the mean yield estimated was 6.2244 t·ha⁻¹, while the value published by National Bureau of Statistics of the People's Republic of China was 6.13 t·ha⁻¹. The estimates obtained with SAFY highlight lower yields in 2013 as compared to 2014, in line with official statistics. The yield estimation error of the EnKF with SAFY for the punctual application was around 17% and it is confirmed by the regional application, where for the 2013 it was about 18% and for 2014 about 16%.

The 2013 growth season was characterized by low rainfall between October (the month of sowing) and May (senescence period) and by low temperature in winter, particularly between December and January, when temperatures were always below zero. The yield

in 2013 was around the 13% lower than 2014. This yield loss could be thus explained by low winter temperatures and water stress.

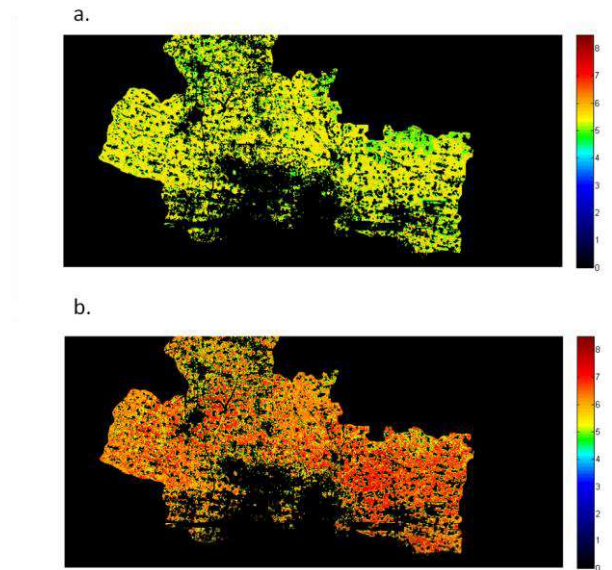


Figure 3. Wheat yield map for Yangling estimated using the EnKF assimilation with SAFY for 2013(a) and 2014 (b).

The winter wheat crop was exposed to thermal stress only between December and January, while temperatures were higher than the seasonal average for the remaining part of 2013. The period of low temperatures, even below zero, concerned a rather short period, therefore this does not appear to be the main cause of loss of yield. The Yangling rural area is classified as semi-arid, rainfall occur in the summer, while winters are almost totally lacking rain. Winter wheat crops have to be irrigated, but irrigation is not always sufficient to compensate for the lack of rain, especially in the fields far from rivers and lakes in the Yangling area. These observations allow to assume that the decrease in yield between 2013 and 2014 is mainly due to drought in the winter of 2013 which affected crop stand establishment.

5. CONCLUSIONS

Results obtained with the assimilation of CC into AquaCrop, have shown the adequacy of this method to provide yield estimation as affected by water availability, even though some drawbacks appear for the spatialized applications. Its highest complexity makes the calibration more difficult, the choice of values to assign to the parameters which remained fixed (not re-calibrated) was based on the literature and on a preliminary calibration, but for regional scale applications this could be a strong limitation. Its lower plasticity means that this model is less affected by the

boundary conditions, such as weather conditions or scenario characteristics, which makes it less suitable to establish a relationship between drought and yield. Its high accuracy is its strong point, but its high computational cost makes it less suitable for a spatialized application.

Results obtained from the assimilation of LAI into the SAFY model seem to indicate that the EnKF assimilation procedure performed better. Given the lower number of parameters of the model, it proved to be quite robust despite the lower complexity of SAFY. The accuracy of SAFY is lower than that of Aquacrop, but using the EnKF assimilation method it is possible to increase it, reaching acceptable accuracy, albeit slightly less than that of Aquacrop.

The application on regional scale of EnKF with the SAFY model proved to be a good method, both for the accuracy (confirmed by official statistics) and for computational requirements. To further validate this approach it would be appropriate to repeat this study on different regions, with different climatic characteristics from each other, for which it is possible to have accurate statistical information about the yield.

6. REFERENCES

- [1] Wallach, D. J. W. J., Makowski, D., (2006). *Working with Dynamic Crop Models. Evaluation Analysis, Parameterization and Applications*. (J. W. J. D. Wallach, D. Makowski, Ed.). Elsevier Science.
- [2] Casa, R., Varella, H., Buis, S., Guérif, M., Solan, B. De, & Baret, F. (2012). Forcing a wheat crop model with LAI data to access agronomic variables : Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *European Journal of Agronomy*, 37(1), 1–10. <http://doi.org/10.1016/j.eja.2011.09.004>
- [3] Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, 101(3), 426. <http://doi.org/10.2134/agronj2008.0139s>
- [4] Duchemin, B., Maisongrande, P., Boulet, G., & Benhadj, I. (2008). A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. *Environmental Modelling & Software*, 23(7), 876–892. <http://doi.org/10.1016/j.envsoft.2007.10.003>
- [5] Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description. *Agronomy Journal*, 101(3), 438–447. <http://doi.org/10.2134/agronj2008.0140s>
- [6] Monteith, J. L., & Moss, C. J. (1977). Climate and the Efficiency of Crop Production in Britain [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*.
- [7] Allen, R. G., Pereira, L. S., Raes, D., Smith, M., & W, A. B. (1998). Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. *Irrigation and Drainage*, 1–15.
- [8] Confalonieri, R., Acutis, M., Bellocchi, G., & Donatelli, M. (2009). Multi-metric evaluation of the models WARM , CropSyst , and WOFOST for rice, *Ecological Modelling* 220, 1395–1410. <http://doi.org/10.1016/j.ecolmodel.2009.02.017>
- [9] Confalonieri, R., Bregaglio, S., & Acutis, M. (2012). Quantifying plasticity in simulation models. *Ecological Modelling*, 225, 159–166. <http://doi.org/10.1016/j.ecolmodel.2011.11.02>
- [10] Fox, D. G. (1981). Judging Air Quality Model Performance. *Bulletin of the American Meteorological Society*. [http://doi.org/10.1175/1520-0477\(1981\)062<0599:JAQMP>2.0.CO;2](http://doi.org/10.1175/1520-0477(1981)062<0599:JAQMP>2.0.CO;2)
- [11] Karunatatne, A. S., Azam-Ali, S. N., Izzi, G., & Steduto, P. (2011). Calibration and validation of fao - aquacrop model for irrigated and water deficient. *Expl Agric*, 47, 509–527. <http://doi.org/10.1017/S0014479711000111>
- [12] Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software*, 51, 323–332. <http://doi.org/10.1016/j.envsoft.2013.10.017>
- [13] Baret, F., Pavageau, K., Béal, D., Weiss, M., Berthelot, B., & Regner, P. (2006). Algorithm Theoretical Basis Document for MERIS Top of Atmosphere Land Products (TOA _ VEG), *Contract ESA A0/I-4233/02/I-LG* (March 2006), 1–37.
- [14] Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., ... Ustin, S. L. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113, S56–S66. <http://doi.org/10.1016/j.rse.2008.01.026>
- [15] Nielsen, D. C., Miceli-Garcia, J. J., & Lyon, D. J. (2012). Canopy Cover and Leaf Area Index Relationships for Wheat, Triticale, and Corn. *Agronomy Journal*, 104(6), 1569. <http://doi.org/10.2134/agronj2012.0107n>

FARMLAND DROUGHT EVALUATION BASED ON THE ASSIMILATION OF MULTI-TEMPORAL MULTI-SOURCE REMOTE SENSING DATA INTO AQUACROP MODEL

Yang Guijun⁽¹⁾, Yang Hao⁽¹⁾, Jin Xiuliang⁽¹⁾, Pignatti Stefano⁽²⁾, Casa Raffaele⁽³⁾, Silvestro Paolo Cosmo⁽³⁾

⁽¹⁾ National Engineering Research Center for Information Technology in Agriculture (NERCITA), 2449-26, Beijing 100097, P.R. China, Email: yanggj@nercita.org.cn; yangh@nercita.org.cn

⁽²⁾ Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy) Email: stefano.pignatti@cnr.it

⁽³⁾ Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy), Email: rcasa@unitus.it; p.c.silvestro@unitus.it

ABSTRACT

Drought is the most costly natural disasters in China and all over the world. It is very important to evaluate the drought-induced crop yield losses and further improve water use efficiency at regional scale. Firstly, crop biomass was estimated by the combined use of Synthetic Aperture Radar (SAR) and optical remote sensing data. Then the estimated biophysical variable was assimilated into crop growth model (FAO AquaCrop) by the Particle Swarm Optimization (PSO) method from farmland scale to regional scale.

At farmland scale, the most important crop parameters of AquaCrop model were determined to reduce the used parameters in assimilation procedure. The Extended Fourier Amplitude Sensitivity Test (EFAST) method was used for assessing the contribution of different crop parameters to model output. Moreover, the AquaCrop model was calibrated using the experiment data in Xiaotangshan, Beijing.

At regional scale, spatial application of our methods were carried out and validated in the rural area of Yangling, Shaanxi Province, in 2014. This study will provide guideline to make irrigation decision of balancing of water consumption and yield loss.

1. INTRODUCTION

Drought is the most costly natural disasters in China and all over the world. The frequent occurrence of drought, coupled with the impact of global warming, poses an increasingly severe threat to the Chinese agricultural production, especially in semi-arid regions. The grain loss caused by drought accounted for 60% of all grain losses caused by meteorological disasters, resulting in 58% or more of economic losses in China. However, since its definition is spatially variant and context dependent, the accurate and timely drought monitoring and prediction has always been a challenge.

Therefore, in the context of Dragon 3 cooperation project, the main objectives of our project are to develop new methods to monitor and predict farmland drought accurately and timely in agricultural vegetated area basing on multi-source remote sensing data. Our main research work include: 1) the monitoring of key biophysical variables of crop and soil in farmland

drought by optical and radar remote sensing data, 2) the risk assessment of farmland drought by time series remote sensing and meteorological data, and 3) the crop loss evaluation under farmland drought mainly based on AquaCrop crop model [1]. In the last 2 years, our cooperation research focused on developing methods for the assimilation of biophysical variables into crop growth model, in order to assess drought-induced crop yield losses at regional scale. The brief description of the work was introduced in the following sections.

2. STUDY AREA AND MATERIALS

2.1 Study Area

The field measurements were conducted in the Yangling district of Shaanxi, China (Figure 1). It is characterized by a typical continental climate and belongs to the semi-arid region of China. The maximum temperature is 26.1°C in the summer, and the minimum temperature is -1.2 °C in the winter. In all the seasons, these climates experience extensive and rapid daily temperature changes, and the temperature difference between day and night is significant. The average annual precipitation is 635.1 mm and the frost-free period is 211 d on average.

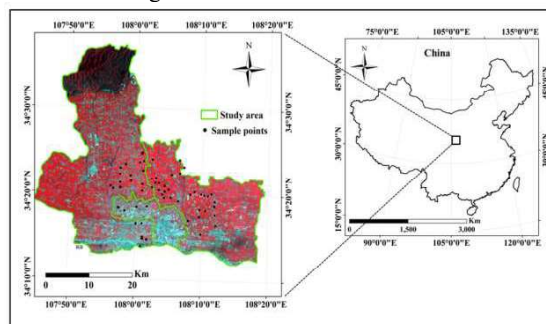


Figure 1. The location map of the study area in Shaanxi, China

2.2 In-situ data and remote sensing data

Field measurements were carried out from March to June for each crop season. The biophysical parameters of crop and soil, including LAI, biomass and soil moisture etc., were measured. Detailed information was presented in [2, 3].

The remote sensing data set consists of several images

acquired by Radarsat-2 and HJ 1-A/B during the three crop cycles (Tab. 1).

Table 1. Remote Sensing data set for Yangling

Radarsat-2	HJ 1-A/B
05/03/2014	04/03/2014
29/03/2014	07/04/2014
22/04/2014	29/04/2014
16/05/2014	20/05/2014
09/06/2014	11/06/2014

3. METHODOLOGY

3.1 Biomass estimation from Full Polarization SAR

Two categories of methods were tested: eigenvalue-based and model-based polarimetric scattering decomposition. The Cloude-Pottier method decomposes the coherency matrix into different eigenvectors and eigenvalues which classify and describe the primary scattering mechanisms. Physical information provided by the decomposition is characterized by four variables: entropy (H), alpha angle (α), and anisotropy (A). In contrast, the Freeman-Durden method decomposes the backscatter response into three categories of basic scattering components: volume scattering (Vol), double-bounce scattering (Dbl), and odd scattering (Odd). Typical polarimetric features were utilized for exploring the polarimetric scattering mechanism of oilseed rape. In this research, polarimetric ratios were also constructed based on polarimetric scattering components: Vol/(Odd+Dbl), or Vol/Odd.

The temporal evolution of scattering intensity features and polarimetric features during the entire growing season was analysed as a function of DAS based on all 88 oilseed rape fields. These features will be evaluated to find potential indicators for crop biomass. The evaluation criterion for these indicators is whether the temporal evolution curve is smooth. If the curve is smooth, simple methods can be used to estimate biomass, which will meet the requirements of operational applications.

When just single polarization channel data were available, the three polarization channels, HH, HV, and VV were investigated, respectively. When dual polarization channel data were available, we selected 4 indicators HH/VV, HH/HV, VH/VV ($HV \approx VH$) and Polarization Discrimination Ratio (PDR). When quad polarization channel data were available, in addition to the decomposition methods described above, we also investigated Radar Vegetation Indices (RVI) as a contrast.

3.2 Biomass estimation by SAR and Optical data

Based on published literature and the sensitivity of optical and polarization features to LAI and biomass, radar polarimetric parameters and optical spectral vegetation indices were used to estimate LAI and biomass of winter wheat in China. In this sub-package,

radar polarimetric parameters (HH, HV, VV, HH/VV, HH/HV, VV/HV, Vol/Span, Dbl/Span, Odd/Span, and radar vegetation index (RVI)) and optical vegetation indices (ratio vegetation index (RVI), normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), optimized soil adjusted vegetation index (OSAVI), enhanced vegetation index (EVI), and modified triangular vegetation index (MTVI2)) were selected to analyse the relationships between vegetation indices and LAI, biomass.

Firstly, the relationships of parameters with LAI and biomass were examined using linear and nonlinear regression analysis. The parameters of the satellite images included the optical spectral vegetation indices (OSVIs) of HJ-1A/B and the radar polarimetric parameters (RPPs) of RADARSAT-2. Secondly, LAI and biomass were estimated with combined OSVIs and RPPs (the product of OSVIs and RPPs (COSVI-RPPs)). Finally, the COSVI-RPPs were used to estimate the LAI and biomass with multiple stepwise regression (MSR) and partial least squares regression (PLSR) methods. The MSR combines a forward selection and backward elimination. The PLSR is an extension of the multiple linear regression model (e.g., multiple regression or general stepwise regression).

3.3 AquaCrop model sensitivity analysis

Aquacrop is a "water driven" productivity model that simulates Canopy Cover (CC), biomass and yield of a crop mainly as a function of the water productivity, i.e. the biomass produced per unit of water transpired by the vegetation (Steduto et al. 2009). A more detailed explanation about how the Aquacrop model works is presented by Raes et al. (2009).

The most important crop parameters of AquaCrop model were determined to reduce the used parameters in assimilation procedure. The Extended Fourier Amplitude Sensitivity Test (EFAST) method was used for assessing the contribution of different crop parameters to model output. 40 crop parameters in AquaCrop were selected with fluctuations of $\pm 10\%$, $\pm 30\%$, $\pm 50\%$ to evaluate the model output. And in the EFAST method, 2 indicators, including First order sensitivity index (FOSI) and total sensitivity index (TSI), were used to assess the contribution of different crop parameters to 4 model output, including static variables (yield and Maximum dry biomass) and dynamic variables (canopy cover and dry biomass).

3.4 PSO assimilation of AquaCrop model

Particle Sparm Optimization (PSO) method was used in the AquaCrop model assimilation. Firstly, 8 most sensitive model parameters as a particle were selected. Then, we selected the canopy cover or biomass as the data state variables. Run the AquaCrop model in order to obtain an ensemble of N estimated CCs or Biomass. PSO method will update the AquaCrop model to minimum the difference between simulated biomass and remote sensed biomass. Finally we will obtain the yield output. The flowchart of this procedure is showed in

Figure 2.

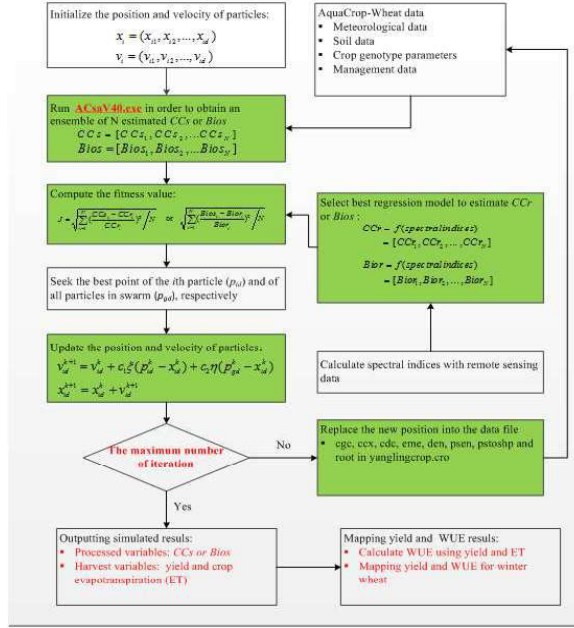


Figure 2. The flowchart of PSO method

4. RESULTS AND DISCUSSION

4.1 Biomass estimation from Full Polarization SAR

The results of the temporal evolution of HH, HV, and VV scattering intensity (in power unit) is presented in Figure 3, which demonstrates their respective scattering response behaviour during the entire growing season [6].

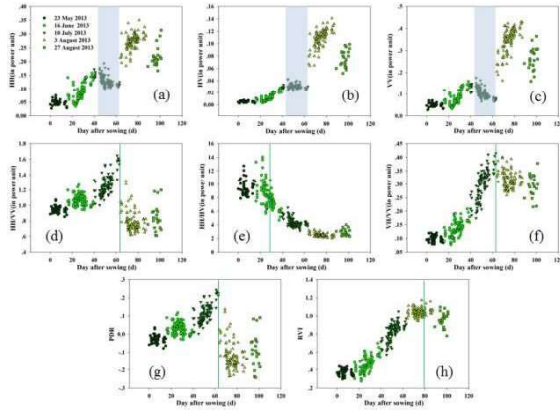


Figure 3. Temporal evolution of different scattering intensity features with DAS. (a)(b)(c), HH, HV, and VV results for single polarization; (d)(e)(f)(g), HH/VV, HH/HV, HH/HV, HH/HV, HH/HV, HH/HV, and PDR results for dual polarization; (h) RVI results for quad polarization.

In addition to scattering intensity, unique polarimetric features can be extracted from fully polarimetric SAR data by target polarimetric decomposition methods. In

Figure 4, the temporal evolution of Entropy, Alpha, and Anisotropy, 3 classic parameters calculated from the Cloude-Pottier decomposition method are presented.

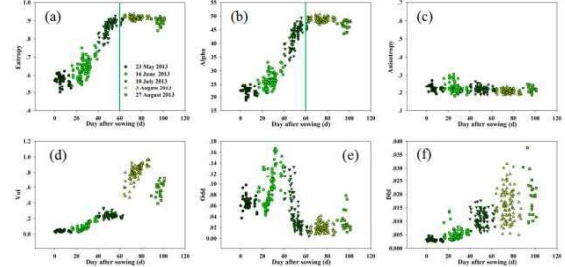


Figure 4 Temporal evolution of different polarimetric features with DAS. (a)(b)(c), Entropy, Alpha, and Anisotropy results by eigenvalue-based decomposition; (d)(e)(f), Volume, Odd, and Double scattering component results by model-based decomposition.

From the Figure 3 and 4, it was revealed that full polarization data have more capability in characterising crop growth, especially when aiming to utilise simple linear models. When we observe the temporal evolution of field observations of biomass and polarimetric ratio, it was found that there is a highly similar pattern between Vol/Odd or Vol/(Odd+Dbl) and biomass parameters when DAS<80. Therefore, it is likely that Vol/Odd or Vol/(Odd+Dbl) have great potential in estimating crop biomass parameters in operational methods.

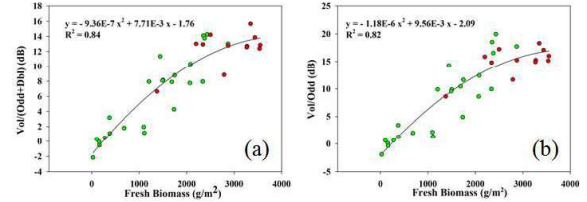


Figure 5 The relationship between fresh biomass and polarimetric indicators, (a) Vol/(Odd+Dbl), (b) Vol/Odd. Red points indicate DAS>75.

4.2 Biomass estimation by SAR and Optical data

Based on the correlations (R^2) between the RPPs and the LAI and biomass from the field experiment, the RPPs that were best correlated with LAI and biomass were RVI ($R^2 = 0.63$) and DERD ($R^2 = 0.71$) [3]. Thus, these indices were used to establish the combined optical spectral vegetation indices and radar polarimetric parameters (COSVI-RPPs) by multiplying each with optical spectral vegetation indices ($RVI \times OSVIs$ and $DERD \times OSVIs$). In this study, six OSVIs were individually combined with RVI and with DERD to analyze the relationships of these COSVI-RPPs with LAI and biomass. Further, we used these correlations to establish LAI and biomass regression equations for winter wheat. The results revealed the highest R^2 value was the $MTVI2 \times RVI$ ($R^2 = 0.68$) for winter wheat LAI. In contrast, the highest R^2 was the $EVI \times RVI$ (R^2

= 0.80) for winter wheat biomass. The results demonstrated that the COSVI-RPPs were highly significantly related to LAI and biomass. They could be used to estimate LAI and biomass in winter wheat.

To validate the estimation accuracy of the regression equations for LAI and biomass, the predicted values and the measured values were compared based on the RMSE and nRMSE. The results were shown in Table 5.6. There was good consistency between the predicted values and the measured values. The RMSE and nRMSE values for the regressions between the MTVI2 $\times$ RVI and LAI were 0.65 and 19.52%, and those for the regressions between the EVI $\times$ RVI and biomass were 146.33 g/m² and 20.21%, respectively. The results showed that MTVI2 $\times$ RVI and EVI $\times$ RVI were better than RVI, MTVI2 and EVI alone for estimation of LAI and biomass in winter wheat, respectively (Figure 6). The results suggested that the MTVI2 $\times$ RVI and EVI $\times$ RVI could be used to improve the estimation accuracy of LAI and biomass, respectively.

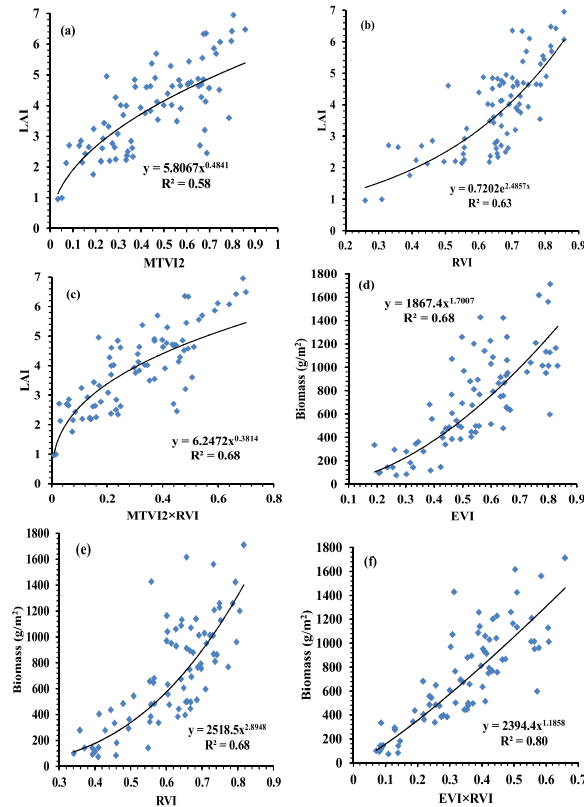


Figure 6. Comparison of MTVI2 (a) RVI (b) and MTVI2 $\times$ RVI (c) for estimation of LAI and EVI (d), RVI (e) and EVI $\times$ RVI (f) for estimation of biomass, respectively.

The MSR and PLSR methods were used to estimate LAI and biomass for winter wheat based on the combined, optical and radar indices. The LAI values

were estimated from MSR and PLSR regression equations obtained based on all the combined optical and polarimetric radar indices, which had R² values of 0.73 and 0.76, respectively. Similarly, the biomass values were estimated from MSR and PLSR regression equations and the R² values were 0.81 and 0.85, respectively (Table 2). All the COSVI-RPPs, OSVIs, and RPPs were selected as variables to estimate LAI using MSR and PLSR, and the R² values of the regression equation were 0.74 and 0.78, respectively. The R² of the MSR and PLSR biomass models based on all the COSVI-RPPs, OSVIs, and RPPs were 0.83 and 0.87, respectively.

Table 2 Comparison of multiple stepwise regression and partial least squares regression methods for estimating LAI and biomass of winter wheat based on OSVIs, RPPs, and COSVI-RPPs (n = 80)

Methods	Variables	R ²	RMSE (g/m ²)	nRMSE (%)
Multiple Stepwise regression	SAVI $\times$ RVI, OSAVI $\times$ DERD, MTVI2 $\times$ DERD	0.81**	142.63	19.71
Partial least squares regression	12 COSVI-RPPs	0.85**	137.21	18.96
Multiple stepwise regression	MTVI2, DERD, SAVI $\times$ RVI, OSAVI $\times$ DERD, MTVI2 $\times$ DERD,	0.83**	140.34	19.39
Partial least squares regression	12 COSVI-RPPs, 6 OSVIs, 15 RPPs	0.87**	134.68	18.61

4.3 AquaCrop model sensitivity analysis

Figure 7 showed the sensitivity analysis results of AquaCrop model by EFAST method. The Figure 7(a) was the first order sensitivity index (FOSI) result for yield. And Figure 7 (b) is the total sensitivity index (TSI) result for yield. The Figure 8 is the FOSI results for time series dry biomass under different parameter variation ranges.

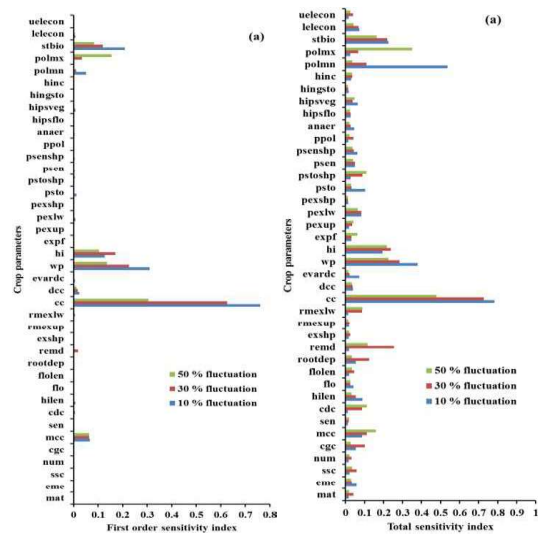


Figure 7. Sensitivity analysis results (a) FOSI result for

measured yield for winter wheat

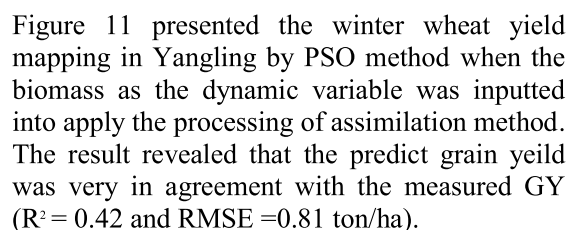
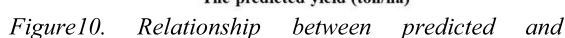
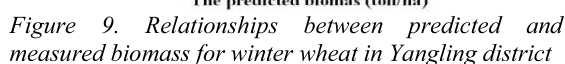


Figure 11. Yield mapping in Yangling by PSO method

5. CONCLUSION

- PSO method which assimilates the remote sensing observation and AquaCrop model can be used to estimate the yield and WUE at regional scale;
- Biomass is more suitable than CC as a state variable in PSO assimilation; and the estimation result of yield based on Biomass ($R^2=0.42$, RMSE = 0.81ton/ha, nRMSE = 17.05%) is better than that based on CC ($R^2 = 0.31$, RMSE = 0.94ton/ha and nRMSE = 19.79%);
- This study provide a method for monitoring the yield and WUE in regional scale by combining the AquaCrop model and RS observation;
- This study provide a guideline for improving the irrigation management of winter wheat in arid regions.



1. Yang, G., Yang, H., Jin, X., Pignatti S., Casa R., Pascucci S., Silverstro P.C. (2014) Research progress of farmland drought monitoring and prediction based on multi-source remote sensing data. ESA SP-724, ESA Publications Division, European Space Agency.

2. Silvestro, P.C., Casa, R., Pignatti, S., Castaldi, F., Yang, H. & Yang, G. (2016). Assimilation of remotely sensed data into crop models for farmland drought assessment: a comparison of models of differing complexity. In '2016 Dragon 3 final results Symposium, Wuhan, China', ESA Publications Division, European Space Agency.
3. Jin, X., Yang, G., Xu, X., Yang, H., Feng, H., Li, Z., Shen, J., Zhao, C., & Lan, Y. (2015). Combined Multi-Temporal Optical and Radar Parameters for Estimating LAI and Biomass in Winter Wheat Using HJ and RADARSAR-2 Data. *Remote Sensing*, 7(10), 13251-13272.
4. Yang, H., Li, Z., Chen, E., Zhao, C., Yang, G., Casa, R., Pignatti, S., & Feng, Q. (2014). Temporal Polarimetric Behavior of Oilseed Rape (*Brassica napus* L.) at C-Band for Early Season Sowing Date Monitoring. *Remote sensing*, 6(11):10375-10394.
5. Yang, H., Chen, E., Li, Z., Zhao, C., Yang, G., Pignatti, S., Casa, R., & Zhao, L. (2015). Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 34, 157-166.
6. Yang, H., Yang, G., Gaulton, R., Zhao, C., Li, Z., Taylor, J., Wicks, D., Minchella, A., Chen, E., Yang, X. In-season Biomass Estimation of Oilseed Rape (*Brassica napus* L.) Using Fully Polarimetric SAR Imagery. *Precision Agriculture*. (in review)
7. Jin, X., Feng, H., Zhu, X., Li, Z., Song, S., Song, X., Yang, G., Xu, X. & Guo, W. (2014). Assessment of the AquaCrop Model for Use in Simulation of Irrigated Winter Wheat Canopy Cover, Biomass, and Grain Yield in the North China Plain. *PLoS ONE*, 9(1): 1–11.
8. Silvestro, P.C., Casa, R., Pignatti, S., Castaldi, F., Yang, H. & Yang, G. (2014). Development of an Assimilation Scheme for the Estimation of Drought-Induced Yield Losses Based on Multi-Source Remote Sensing and the Acquacrop Model. In '2014 Dragon 3 mid-term results Symposium, Chengdu, China', ESA SP-724, ESA Publications Division, European Space Agency.

DEVELOPMENT OF FARMLAND DROUGHT ASSESSMENT TOOLS BASED ON THE ASSIMILATION OF REMOTELY SENSED CANOPY BIOPHYSICAL VARIABLES INTO CROP WATER RESPONSE MODELS

Casa R.¹, Silvestro P.C.¹, Yang H.^{2,4}, Pignatti S.³, Pascucci S.³, Yang G.^{2,4}

¹Department of Agriculture, Forests, Nature and Energy (DAFNE), Università della Tuscia, Via San Camillo de Lellis, Viterbo, Italy (rcasa@unitus.it)

²National Engineering Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

³CNR IMAA, Tito Scalo, Potenza, Italy

⁴Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, China

ABSTRACT

The aim of this work is the development of methods for the assimilation of biophysical variables, estimated from multi-source remote sensing data, into crop growth models, in order to estimate the yield losses due to drought both at the farm and at the regional scale. A methodology to obtain maps of leaf area index (LAI), and fractional canopy cover (CC), from HJ1A and HJ1B Chinese satellite optical data was established, using an algorithm based on the training of artificial neural networks (ANN) on PROSAIL model simulations. Retrieved values of biophysical variables, such as LAI or CC, will be assimilated into crop growth models in order to estimate wheat yield. The present work focused on testing two different approaches using a common dataset gathered in Xiaotangshan (China) with two crop models of different complexity, in order to compare the procedures and analyse the responses of the models, before the subsequent application at a regional scale in Yangling, Shaanxi, Central China.

Index Terms— Crop model, Aquacrop, SAFY, sensitivity analysis, Ensemble Kalman Filter (EnKF)

1. INTRODUCTION

Drought is one of the climatic phenomena that most affect crop production of agro-ecosystems causing huge yield losses worldwide. For example, it is estimated that in China drought-affected crop areas reach 20 million hectares, with a loss of 10-15 million tons of grain due to drought every year. Therefore, methods for quantifying these losses rapidly from remote sensing are extremely valuable. The early assessment of drought from remote sensing is usually based on the detection of anomalies in vegetation indices as compared to long term averages. This is appropriate for regional scale

applications on natural or permanent vegetation, whereas at the farmland scale, the inter-annual changes of vegetation cover, typically related to crop rotations, hinders the use of such methodologies. Therefore our objective is the development of alternative techniques for the assessment of drought impact on crops, suitable for high spatial and temporal resolution applications. Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop water response models is a promising methodology in this respect.

In the context of the Sino-European Dragon-3 program funded by the European Space Agency (ESA) and the Ministry of Science and Technology (MOST) of the P.R. of China, the "Farmland Drought" project deals with the development of assimilation schemes that could allow to estimate the impact of drought on crop yields at the farm or regional scale. The objective of the present work is the development of methods for the assimilation of biophysical variables, retrievable from remote sensing, into simple crop growth models and their test on field data.

2. STUDY SITES AND MATERIALS

Two study sites located in different part of China were used respectively for a local scale and regional scale application. The local scale field site is in Xiaotangshan Experimental Farm (44.17°N, 116.43°E), near Beijing. The farm is representative of the overall soil and crop management practices in this region, characterized by a typical continental climate, with maximum temperatures of 26.1°C in the summer, minimum temperatures of -4.7°C in the winter and a average annual precipitation of around 680 mm. Field experiments were conducted by NERCITA (National Engineering Research Center for Information Technology in Agriculture) during 3 years, from 2008 to 2011, using 3 different winter wheat cultivars sown at 3 different dates

each year. The dataset of these experiments includes measurements of leaf area index (LAI) and above-ground biomass collected in 4-5 different dates per year. Biomass was determined from a 0.25 m² area from each plot. All plant samples were oven dried at 70°C to a constant weight, and final dry weight recorded. LAI was estimated using The LAI-2000 Plant Canopy Analyzer (LI-COR Inc., Lincoln, NE, USA) and subsequently it was converted into Canopy Cover using the following expression [2]:

$$CC = 1 - \exp^{-0.65LAI}$$

Further details on these experiments were published by [1]. The regional scale study site is in a rural area near Yangling (34.30°N, 108.00E) in the Province of Shaanxi. The site is characterized by a semi-arid climate, with a minimum temperature of -4°C in winter, a maximum temperature of 35°C in summer and an average precipitation of around 500 mm per year. Field data were collected by NERCITA during the 2013 wheat growing season. They consisted in the observation of about 50 points, randomly distributed in different winter wheat fields spread across an area of about 750 km². LAI and biomass measurements were collected for each point in 2 different dates (30 March and 27 April) and the yield was recorded at the end of the season (4 June). LAI and Biomass were collected in the same way of Xiaotangshan experiments.

In the period corresponding to this measurement campaign a set of 6 images were acquired by the Chinese satellites HJ1A and HJ1B and one by Landsat 8 (Tab 1).

Table 2: Satellite data set acquired over the Yangling site.

Satellite	Sensor name	Sensor Characteristics	Date of Acquisition (mm/dd/yyyy)
HJ1B	CCD 2	multispectral, 4 bands, VNIR	05/03/2013
HJ1B	CCD 1	multispectral, 4 bands, VNIR	20/03/2013
HJ1B	CCD 2	multispectral, 4 bands, VNIR	28/03/2013
HJ1A	CCD 1	multispectral, 4 bands, VNIR	30/03/2013
HJ1A	CCD 1	multispectral, 4 bands, VNIR	07/04/2013
HJ1A	CCD 1	multispectral, 4 bands, VNIR	26/04/2013
HJ1A	CCD 1	multispectral, 4 bands, VNIR	11/05/2013
Landsat 8	OLI	multispectral, 9 bands, (Visible, NIR, SWIR)	03/04/2013

3. METHODS

A part of this work consisted in setting up the procedure in which biophysical variables, in particular LAI and Canopy Cover, could be used for the assimilation into crop water response models. Two models with different degrees of complexity were studied: the more complex AquaCrop [3] and a modified version of the simpler SAFY model [4].

AquaCrop is a water productivity model, developed by FAO [3], which simulates canopy biomass production and yield in response to water transpired by the crop. The model simulates canopy cover (CC) along the growing season as a function of three main sets of inputs and parameters

concerning climate, crop and soil characteristics. The choice of this crop growth model was suggested by its relative simplicity as compared to more complex models, by its emphasis on water limitation and by the explicit use of CC, as main canopy state variable. This variable is usually easier to estimate from remote sensing than LAI, not being subject to saturation at high values.

SAFY [4] is based on Monteith's concept of conversion of absorbed photosynthetically active radiation into biomass, then converted to grain yield using empirical partitioning coefficients. The original SAFY code was modified by adding a very simple water balance module according to the FAO 56 equations [9].

The decision to use two different models was taken after a preliminary test [5] which had shown some drawbacks in the use of AquaCrop for assimilation. The test had shown the adequacy to provide yield estimation as affected by water availability, but there were problems for a spatialized application because of the high computational time. Additionally given the impossibility of modifying the code of AquaCrop an optimization, rather than a proper assimilation, was carried out. In the present work we therefore compare AquaCrop model optimization to the assimilation with the SAFY model, the latter carried out applying the Ensemble Kalman Filter (EnKF) method.

A preliminary sensitivity analysis was performed both for AquaCrop and SAFY, using the input values for the Xiaotangshan study site, in order to detect the most influential parameters in the two models, thereby reducing the number of parameters to be varied during the assimilation. Two different global sensitivity analysis (GSA) methods were used: Morris [7] and Extended Fourier Amplitude Sensitivity Test (EFAST) [8]. Based on the sensitivity analysis we could reduce the number of AquaCrop parameters to optimize from 70 to 9, whereas for SAFY the parameters to vary during the assimilation could be reduced from 50 to 6.

Both assimilation and optimization procedures, used respectively for SAFY and AquaCrop, were tested with field scale point data acquired at Xiaotangshan between 2008 and 2011 [1]. Subsequently they will be applied for a regional scale test using the dataset collected in Yangling in 2013 and the biophysical information retrieved from satellite data, using an algorithm based on the training of artificial neural networks using PROSAIL model simulation [11].

The assimilation methodology used for SAFY follows the EnKF procedure reported by [10]. At first we varied, during the assimilation, the 20 parameters identified by Morris GSA, but then these were reduced to 6 parameters identified as the most influential by EFAST GSA. The other parameters were fixed to values obtained during a preliminary calibration for a single date in the Xiaotangshan 2008 experiment (sowing date 27 September). We tested EnKF using a number of ensembles ranging from 50 to 500, proving that 50 ensembles are sufficient to ensure the

convergence of the method. For AquaCrop, the assimilation algorithm employed minimizes the difference between simulated and observed CC, using an optimization method based on the simplex [13]. We used a number of ensembles of 30, performing the minimisation for each ensemble and finally selecting the parameter set for the ensemble best fitting observed CC values.

The regional scale application dealt with the set-up and validation of methods for the retrieval biophysical canopy variables from remote sensing (RS) data. The algorithm developed by [11] was adapted for the use with HJ1A and HJ1B Chinese multispectral satellite and validated with data acquired in 2013 in the Yangling test site. The retrieved values of LAI were compared to measurements collected in several wheat fields in the Yangling area at the time of HJ1A and HJ1B satellites overpass.

4. RESULTS

The target variable chosen to describe the performances of these two different methods, the parameters optimization applied on AquaCrop and the EnKF applied on SAFY, is the grain yield. In Fig. 1 a comparison is made between the yield simulated by the two models and the yield measured during the Xiaotangshan experiments [1], before and after the application of optimization and assimilation.

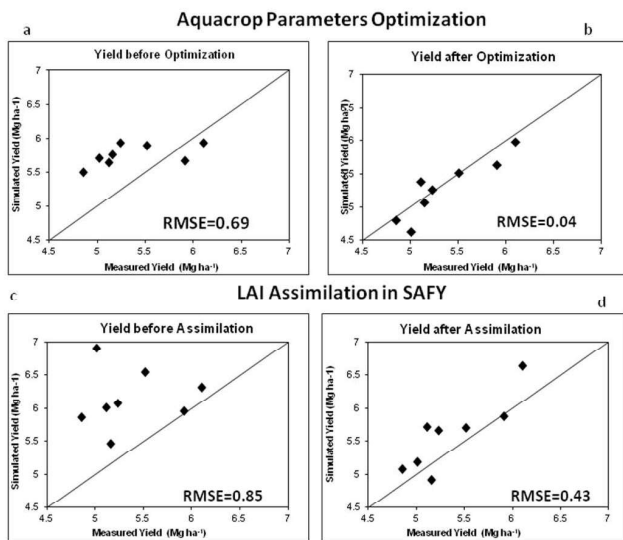


Figure 1. Comparison between simulated and measured yield for : a) AquaCrop ; b) AquaCrop after Parameter Optimization; c) SAFY; d) SAFY after LAI Assimilation with Ensemble Kalman Filter Method

Grain yield was better approximated by the AquaCrop model than by SAFY, in fact the former attained a RMSE of 0.69 whereas the latter a RMSE of 0.85. After the optimization of parameters based on CC measurements, AquaCrop greatly improved the simulation, reaching an RMSE of 0.04. The

Assimilation of LAI in SAFY using the method of Ensemble Kalman Filter improves the simulation, in fact the RMSE is halved (Fig. 1c). Therefore both the optimization of AquaCrop Parameters and LAI Assimilation improve the performances of models. Even if in the first case the RMSE is reduced by about 90%, against a reduction about 50% in the second case, there are some drawbacks. The optimization of AquaCrop parameters attains a good approximation of yield, but with a computational time of around 40 minutes on a PC with an Intel Core i7-4770 CPU at 3.4 GHz for each simulated data point. This would make it unpractical for a spatialized application at the regional scale on hundreds of pixels. The results obtained with the LAI EnKF assimilation method in SAFY show a lower improvement than the first method. This method reduces the RMSE about 50% as compared to the use of SAFY without assimilation, but the computational time is around 20 seconds. Furthermore it can be noted that the estimation of yield simulated by the SAFY assimilation is also better than the estimation of yield by AquaCrop before optimization (the RMSE is reduced from 0.69 to 0.43).

Concerning the regional scale application, the algorithm employed by [11] was successfully adapted to the multispectral data of the Chinese satellites HJ1 and HJ2, providing a realistic time trend of biophysical variables, such as wheat LAI (Fig. 2).

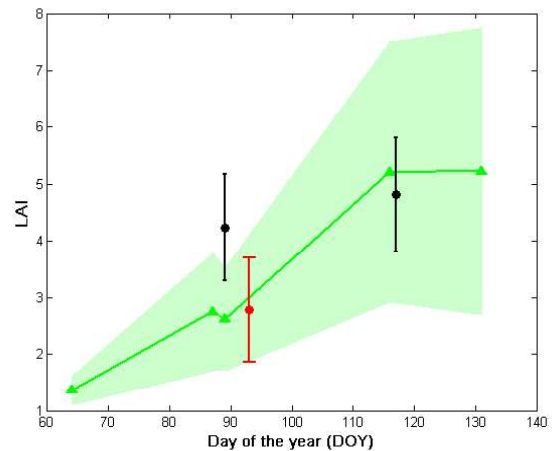


Figure 2. Average values of LAI retrieved by the algorithm of [11] from HJ1 and HJ2 satellite data at the different dates for the Yangling ground measurement points (green triangles) and their standard deviation (green shaded area). Average LAI value (red circle) and standard deviation (bar) retrieved from Landsat 8 data. Average (black circles) and standard deviation of ground measured LAI.

The values of LAI estimated at the different dates of the satellite overpass show, on average, increasing values until the end of April, but the later part of the growth season, with decreasing LAI is missed since the last cloud free satellite image was available for May 11. The increasing spread observed in LAI values along the growth season is also often

observed in reality. LAI ground measurement values were on average higher than satellite retrievals for the first sampling date (end of March), whereas they were more similar to estimates at the second date (end of April). The RMSE values were respectively of 1.45 and 2.13, the latter higher value could be explained by the larger scatter of the data values at the later date. The spatial resolution of the satellite data (30 m) as well as the ground sampling protocol (single points often close to roads) negatively affected these validation results. Additionally the atmospheric correction of the data, performed in this study using the FLAASH module in ENVI, should be improved, since for several ground measurement points (wheat fields) pixel spectra were not those typically expected for vegetation.

The next step in our work will concentrate on the application of the assimilation algorithms at the regional scale on the basis of these LAI estimates, therefore their uncertainty will be taken into account.

5. CONCLUSIONS

Two different methods for the assimilation of biophysical variables retrievable from remote sensing were tested employing field scale data, revealing the pros and cons of using crop models with different degrees of complexity. The more complex Aquacrop model provided better results in the estimation of grain yield, but the assimilation was computationally demanding. SAFY was less accurate, but it was more suitable for a proper assimilation algorithm such as the Ensemble Kalman Filter, making it a good candidate for a spatialized application at the regional scale. For such applications the accurate retrieval of biophysical variables from satellite is quite important. The application of the artificial neural network algorithm of [11] to HJ1 and HJ2 data provided reasonable results, reproducing a realistic time trend of wheat LAI.

6. REFERENCES

- [1] Jin, X., Feng, H., Zhu, X., Li, Z., Song, S., Song, X., ... Guo, W. (2014). Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PloS One*, 9(1), e86938. doi:10.1371/journal.pone.0086938
- [2] Hsiao TC, Heng L, Steduto P, Roja-Lara B, Raes D, et al. (2009) AquaCrop-The FAO model to simulate yield response to water: parametrization and testing for maize. *Agronomy Journal* 101: 448–459.
- [3] Steduto, P., Hsiao, T.C., Raes, D. & Fereres, E. (2009). AquaCrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal* 101(3), 426-437
- [4] Duchemin B., Maisongrande P., Boulet G, Benhadj. I (2008). A simple algorithm for yield estimates: evaluation for semi-arid irrigated winter wheat monitored with ground-based remotely-sensed data. *Environmental Modelling and Software* 23, 876-892
- [5] Silvestro, P. C., Casa, R., Pignatti, S., Castaldi, F., Yang, H., & Yang, G. 2014. Development of an assimilation scheme for the estimation of drought-induced yield losses based on multi-source remote sensing and the AquaCrop model. 2014 Dragon 3 Mid-Term Results International Symposium, 26-29 May, 2014, Chengdu P.R. China.
- [6] Vanuytrecht, E., Raes, D. & Willems, P (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling and Software* 51, 323-332.
- [7] Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22(10), 1509–1518. doi:10.1016/j.envsoft.2006.10.004
- [8] D. Wallach, D. Makowski, J.W. Jones,(2006) “Working with Dynamic Crop Models”, ELSEVIER,pp. 151-169.
- [9] Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, Rome, Italy.
- [10] Wallach, D. Makowski, D., Jones, J.W. 2006. Working with Dynamic Crop Models, Elsevier, 151-169.
- [11] Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., ... & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens. Environ.* 110(3), 275–286.
- [12] Jacquemoud S., Verhoef W., Baret F., Bacour C., Zarco-Tejada P.J., Asner G.P., Christophe F. & Ustin S.L. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization, *Remote Sens. Environ.* 113(1), S56–S66
- [13] Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*, 9, 112-147

DEVELOPMENT OF AN ASSIMILATION SCHEME FOR THE ESTIMATION OF DROUGHT-INDUCED YIELD LOSSES BASED ON MULTI-SOURCE REMOTE SENSING AND THE ACQUACROP MODEL

Silvestro Paolo Cosmo ⁽¹⁾, Casa Raffaele ⁽¹⁾, Pignatti Stefano ⁽²⁾, Castaldi Fabio ⁽¹⁾, Yang Hao ⁽³⁾, Yang Guijun ⁽³⁾

⁽¹⁾ Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy), Email: rcasa@unitus.it; p.c.silvestro@unitus.it

⁽²⁾ Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy) Email: stefano.pignatti@cnr.it

⁽³⁾ National Engineering Research Center for Information Technology in Agriculture (NERCITA), 2449-26, Beijing 100097, P.R. China Email: yanggj@nercita.org.cn; yangh@nercita.org.cn

ABSTRACT

In the context of the Dragon-3 Farmland Drought project, our research deals with the development of methods for the assimilation of biophysical variables, estimated from multi-source remote sensing, into the AquaCrop model, in order to estimate the yield losses due to drought both at the farm and at the regional scale. The first part of this project was employed to refine a methodology to obtain maps of leaf area index (LAI), canopy cover (CC), fraction of adsorbed photosynthetically active radiation (FAPAR) and chlorophyll (Cab) from satellite optical data, using algorithms based on the training of artificial neural networks (ANN) on PROSAIL model simulations.

In the second part, retrieved values of CC were assimilated into the AquaCrop model using the assimilation method of the Ensemble Kalman Filter to estimate grain wheat yield at the field scale.

1. INTRODUCTION

Drought is one of the variables that most affect crop production of agro-ecosystems causing huge yield losses worldwide. It is estimated that in China drought-affected crops area reach 20 million hectares, with a loss of 10-15 million tons grain due to drought every year. Therefore methods for quantifying these losses rapidly from remote sensing are extremely valuable. Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop models is a promising methodology in this respect.

AquaCrop is a water productivity model, which simulates canopy biomass production and yield in response to water transpired by the crop [1]. The model simulates canopy cover (CC) along the growing season as a function of three main sets of inputs and parameters concerning climate, crop and soil characteristics. The choice of this crop growth model for our aims was suggested by its relative simplicity as compared to more complex models, by its emphasis on water limitation and by the explicit use, as main canopy state variable, of CC. This variable is usually easier to estimate from remote sensing than LAI, not being subject to saturation at high values.

Our efforts were focused on the retrieval of CC from optical remote sensing data, through a methodology involving the training of artificial neural networks (ANN) on the PROSAIL model [2]. Successively the assimilation of remotely sensed CC into the AquaCrop model was attempted, through the development of an algorithm based on the general principles of the Ensemble Kalman Filter (EnKF) [3]. This method is suitable for assimilation into complex non-linear models and it has been previously used for the same purpose, e.g. with the WOFOST model [4].

This paper presents a preliminary test of the assimilation of CC retrieved from SPOT satellite data, through the use of EnKF into the AquaCrop model, for the estimation of grain yield at the farm scale.

2. MATERIALS AND METHODS

Wheat yield data are collected on a field within the Maccarese farm, about 30 km West of Rome (Fig. 1), coded as B030 (17.33 ha). Yield data were recorded using a combine harvester (New Holland CX860) equipped with a yield mapping system (Precision Land Management System, New Holland) consisting of global position system (GPS) sensor with differential correction (based on the EGNOS signal), grain flow sensor, a capacitance grain moisture sensor and software that allows to record data and display thematic maps. The raw data provided by the yield mapping system (about 1400 points ha⁻¹) were corrected, deleting points that showed uncorrected and unrealistic values.

SPOT 4 and SPOT 5 satellite images were acquired over the studied field during the growing season of wheat in 2012 on six dates: 16th February (SPOT 4), 2nd March (SPOT 4), 25th March (SPOT 5), 10th May (SPOT 4), 27th May (SPOT 5) and 16th June (SPOT 5). We used the four bands (B1: 0.50 – 0.59 µm; B2: 0.61 – 0.68 µm; B3: 0.78 – 0.89 µm; B4: 1.58 – 1.75 µm) of the multispectral sensors (HRVIR for SPOT 4 and HRG for SPOT 5) of SPOT satellite [5]. SPOT images were processed carrying out radiometric correction obtaining the top of atmosphere (TOA) reflectance. TOA values were then atmospherically corrected into top of the canopy (TOC) reflectance using the SMAC model [6]. The geometric correction of SPOT images was

subsequently carried out using ground control points (GCPs) and nearest neighbour resampling by means of the ENVI software (Exelis VIS, Boulder, USA).



Figure 1. Maccarese farm's field (B030) used for the measure and estimation of wheat grain yield. The background is a SPOT 4 image acquired on 2nd March 2012.

2.1. ANN + PROSAIL algorithm

Several studies conducted in the last decade have shown the possibility of use ANN to estimate products such as LAI, FAPAR, CC and Cab from multispectral or hyperspectral satellite data [7] [8]. In the present study, we have adopted an algorithm relying on simulations with the PROSPECT and SAIL models [9] to train ANN [10] [11]. This method has shown good results reported in literature [7] [2] [8] for application using middle resolution satellite sensors (e.g. MERIS, SPOT Vegetation, Polder). The main steps of the ANN + PROSAIL algorithm are shown in Figure 2.

The first step consists of the parameterization of PROSAIL model in order to generate a high number of simulations of vegetation reflectance (at around 50-90000). This parameter identifies 14 variables that describe the biophysical state of vegetation. For each variable it is necessary to define a statistical distribution (Gaussian or uniform), specifying minimum and maximum values, average, standard deviation and a number of classes in which the range is expected to be divided. The input set is generated by combining these parameters. Then this input set (at around 50-90000 combination of parameters) is used to generate PROSAIL simulation of reflectance spectra.

At a later stage, the set of simulated reflectance is resampled according to the characteristics of the used sensor, and then it is divided in two parts, one of them with the purpose of training ANN and the other one to evaluate the estimate theoretical performance.

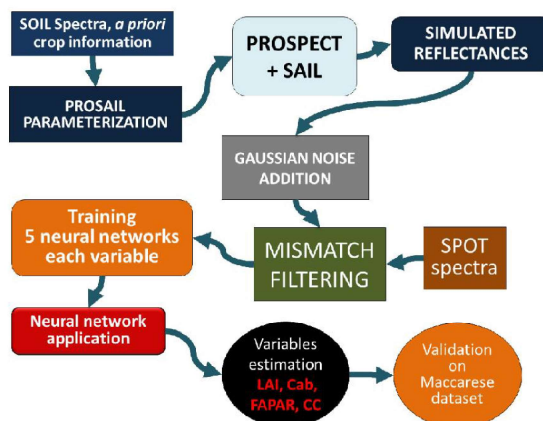


Figure 1. Steps employed in the tests of the ANN algorithm of Baret et al. 2007 for the estimation of biophysical vegetation variables

To make the simulations more realistic and reduce the training hyper-specialization, a Gaussian noise reflectance is added, and calculated on the basis of an additive and multiplicative model error. This one can be considered the same for all bands and can vary for each band of the sensor.

Once the training of the neural networks (which architecture needs to be specified: number of neurons and layers), the coefficients and bias are stored to be used later in the estimation of variables LAI, FAPAR and CC from reflectance observed values (Fig. 3).

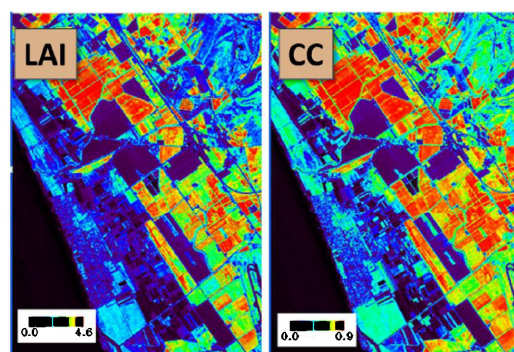


Figure 3. Maccarese area, maps of LAI (on the left) and CC (on the right) obtained from SPOT4 image acquired on 25 March 2012 using the ANN-PROSAIL algorithm.

2.2. AquaCrop model and Ensemble Kalman Filter assimilation data method

Aquacrop is a water productivity model which simulates aboveground biomass production in exchange for water transpired by the crop [1]. The conceptual equation at the core of the AquaCrop growth engine is:

$$B_n = WP^* \sum_{i=1}^n \frac{Tr_i}{ETo_i} \quad (1)$$

where B_n is the cumulative aboveground biomass production after n days; Tr_i is the daily crop transpiration (mm day^{-1}), ETo_i is the daily reference evapotranspiration, and i represent the sequential days of the period in which B_n is produced; WP^* is the normalized crop water productivity (g m^{-2}). Transpiration (Tr) is calculated as a function of the Canopy Cover (CC_i), for this reason CC is a key variable for the yield estimation. AquaCrop model calculates CC as a function of four sets of main parameters and driving variables depending on: weather data, crop type, soil and initial conditions of soil water content.

We tested the possibility of implementing an assimilation scheme in which CC data estimated from remote sensing are used for re-calibrating AquaCrop at each observation date. We tested the Ensemble Kalman Filter (EnKF) method to assimilate CC data into AquaCrop. The EnKF is a recursive Bayesian estimation method, which takes into account the true probability density function of the model states conditioned on observations. The dynamic model is defined by:

$$\varphi_t = F(\varphi_{t-1}) + \varepsilon_{t-1} \quad (2)$$

where φ_t is the vector including the state variable (in our case CC) and a set of parameters that most affect the model and ε_{t-1} is the vector of errors.

AquaCrop model has a substantial number of parameters, for this reason it was necessary to select those which most influence the CC . In this regard we selected a set of 13 parameters (Tab.1) to be recalibrated during the assimilation, whereas all other parameters were kept fixed. The 13 parameters were selected according to previous AquaCrop sensitivity analysis studies [12] and the calibration of the model on data from the Maccarese area.

The EnKF takes into account both observation and model errors, typically through the re-initialisation of an ensemble of non-linear model realisations at each observation date. The choice of the size of the ensemble has been shown to be very important. In our case, since we applied the algorithm on a pixel-by-pixel basis, computation cost forced us to limit the size of the ensemble to 10. Additionally, the model executable code did not allow us to re-initialise it at the observation dates, so we instead recalibrated it, by using sequentially the CC estimates from the EnKF. This considerably increased the computation time. The observation error on CC influences the EnKF correction. As a proxy of this error we used the variance

of ground measurements of CC within ESUs (Elementary Sampling Units) of squares with sides of 20 m. This variance was sometimes rather high and this, combined with the model variance of the ensemble, determined the fact that the CC values obtained at each date from the EnKF could be sometimes rather different from those measured (Fig. 4).

Parameter	Min.	Max	Mean	St. Dev.
CGC (%/GDD)	0.03	0.07	0.035	0.0036
CCX (%)	0.65	0.99	0.7	0.07
CDC (%/GDD)	0.04	0.1	0.10	0.018
GDD_eme (GDD)	40	150	132	15
N° of plants/ha	2×10^6	5.5×10^6	4.1×10^6	0.8×10^6
p - sen	0.55	0.75	0.65	0.13
Sh_stom	1.5	3.5	3	0.3
GDD_root (GDD)	323	1500	1067	104.6
PWP (%)	14	39	21.28	8.66
CN	65	85	75	7.07
FC (%)	22	54	33.9	8.62
WC_ic (%)	14	54	26	11.17
WPo (g/m^2)	15	19	17	1.79

Table 1. Crop and soil parameters of AquaCrop model used for the Ensemble Kalman Filter assimilation method. GDD: growing degree days; CGC: canopy growth coefficient; CCX: maximum canopy cover; CDC: canopy decline coefficient; GDD from sowing to emergence; p-sen: soil water depletion factor for canopy senescence (upper threshold); Sh_stom: shape factor for water stress coefficient for stomatal control; GDD_root: GDD from sowing to maximum rooting depth; PWP: permanent wilting point; CN: curve number; FC: field capacity; WC_ic: initial soil water content; WPo: water productivity.

Data assimilation was carried out using an algorithm designed according to the general principles of the EnKF approach, through the following steps:

- Generate an ensemble of N (in this study $N=10$) values of parameters $P_1 \dots P_{13}$ from a normal distribution using the data of Tab.1.
- Run AquaCrop in order to obtain an ensemble of N CC values for the first observation date.
- Compute the variance-covariance matrix of model error from the ensemble.
- Calculate the Kalman gain from the variance-covariance matrix and an ensemble of N observations of CC (from SPOT data), computed from the sum of measured value and

- an error drawn from a normal distribution with 0 mean and the measurement variance.
- v. Application of the Kalman filter to obtain a corrected CC and parameters values estimated taking into account measured and modelled values and their respective errors.
- vi. Run AquaCrop and find the best ensemble fitting the CC of the previous step by recalibration of the parameters of Table 1, by using the simplex method of optimization [13].
- vii. At each observation date repeat the steps from iii
- viii. Finally run the model until the end of the growth season and obtain a grain yield estimate.

This procedure was applied for each pixel (~450) of the SPOT images acquired on B030 field and the grain yield values were compared with measured values retrieved by the combine harvester data. The estimation error was evaluated by root means square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (3)$$

where y and $\hat{y}$ are respectively the measured and estimated values and N the number of data point.

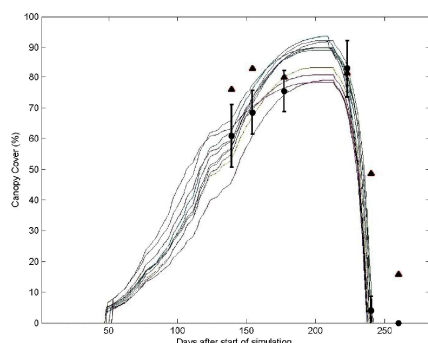


Figure 4. Time trend of canopy cover (CC) simulated by the ensemble of 10 Aquacrop model runs and the observed CC from SPOT4 data (red triangles) and the mean (black dot) and range of variation (black bar) of CC values estimated from the EnKF.

3. RESULTS

The algorithm based on the training of ANN on PROSAIL model simulations proved to be very effective to obtain maps of LAI, CC, FAPAR and Cab (Fig. 5). However, for a more complete assessment of the error and uncertainties in the estimates of the biophysical variables, a more thorough validation would be required. The grain yield map obtained by the

assimilation of CC in AquaCrop model by the EnKF algorithm showed a good correspondence with the map of measured values (Fig. 6). The comparison between estimated and measured grain yield values on all pixel of B030 provided a RMSE of 1.55 t ha⁻¹.

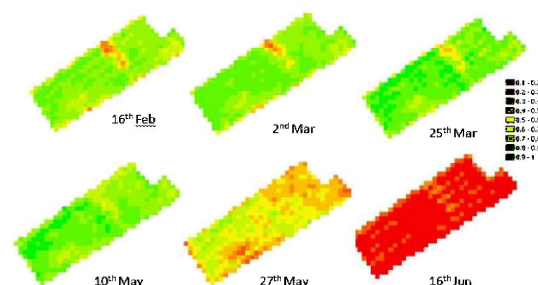


Figure 5. Maccarese area, field B030, maps of green canopy cover (CC) time trend obtained from SPOT4 images during crop cycle. These are the observed (measured) values of CC assimilated into the AquaCrop model.

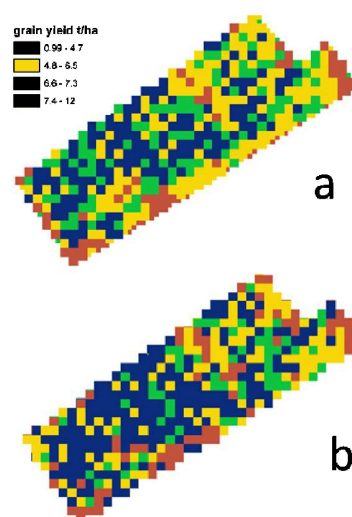


Figure 6. Wheat grain yield on B030 field (a) measured by combine harvester and (b) estimated by from assimilation of canopy cover into AquaCrop model.

4. DISCUSSION AND CONCLUSIONS

The preliminary test of the application of the Ensemble Kalman Filter method to AquaCrop, has shown its adequacy to provide yield estimation as affected by water availability. It has highlighted, however, aspects that need further work and improvements, and in particular:

- i. The choice of parameters to recalibrate. In AquaCrop there are about 100 parameters related to the characteristics of the crop and

the soil. With the aid of a preliminary calibration and sensitivity studies taken from the literature [12], we chose 13 parameters subject to recalibration during the EnKF algorithm application. The choice was rather arbitrary, based on our assessment of which parameters we considered to be the most influential (Tab.1). For a more objective choice, a specific sensitivity analysis is required.

- ii. Fixed parameters values. The choice of values to assign to the other parameters which remained fixed was based on the literature and on a preliminary calibration. For applications outside our test site a proper calibration of the fixed parameters is needed.
- iii. Variable parameters values. The values of variable parameters are chosen from a normal distribution with an assigned mean and standard deviation. The choice of these mean and standard deviation are based on literature and a priori hypothesis. They have a large influence on the spread of the values simulated in the ensemble (Fig. 4)
- iv. Method of optimization. The impossibility in the executable of the AquaCrop model to implement a re-initialisation at the observation dates, lead us to recalibrate it at each date by using sequentially the CC estimates from the Kalman filter in an optimization method. The high computation time required by the optimization, imposed a small ensemble size ($N = 10$).
- v. Measurement and model Error estimates. This preliminary test showed that the error of the model (which was assumed only dependent on the variance of the parameters) was smaller than the error of the observations. Therefore, it is necessary to introduce in an appropriate manner a total error on the model and to have more realistic estimates of CC observation error.

Further developments foreseen in the Dragon-3 Farmland Drought project are the use of a similar approach for the assimilation of multi-sensor data (i.e. optical and radar) into Aquacrop for both field-scale and regional applications in China.

5. ACKNOWLEDGEMENTS

The authors are grateful to ESA for the support to P.C. Silvestro PhD through a Dragon3 Young Scientist Grant (ESA contract No. 4000110311/14/I-BG). We also thank the Centre National d'Études Spatiales (CNES) for funding the programming and acquisition of SPOT scenes (ISIS Project 600). We are grateful to Fred Baret and Marie Weiss for the provision of the ANN

algorithm and for their kind assistance. The authors would also like to thank the staff of the Maccarese SpA farm for their kind assistance.

6. REFERENCES

1. Steduto, P., Hsiao, T.C., Raes, D. & Fereres, E. (2009). AquaCrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agron. J.* **101**(3), 426-437.
2. Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., ... & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens. Environ.* **110**(3), 275-286.
3. Evensen, G. (2007). *Data assimilation : The ensemble Kalman filter*. Springer, Berlin.
4. de Wit A., Duveiller G. & Defourny P. (2012) Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agr. Forest. Meteorol.* **164**, 39-52.
5. Gleyzes, J.P., Meygret, A., Fratter, C., Panem, C., Baillarin, S., & Valorge, C. (2003). *Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. IEEE international, vol. 1, pp 300-302.
6. Rahman, H., & Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *Int. J. Remote Sens.* **15**(1), 123-143
7. Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens. Environ.* **105**(4), 313-325.
8. Verger A., Baret F. & Camacho F. (2011). Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote Sens. Environ.* **115**, 415 - 426.
9. Jacquemoud S., Verhoef W., Baret F., Bacour C., Zarco-Tejada P.J., Asner G.P., Christophe F. & Ustin S.L. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **113**(1), S56-S66.

10. Baret, F., Pavageau, K., Béal, D., Weiss, M., Berthelot, B., & Regner, P. (2006). Algorithm Theoretical Basis Document for MERIS Top of Atmosphere Land Products (TOA _ VEG), (March), 1–37.
11. Weiss M., Baret F., Garrigues S., & Lacaze R. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products, *Remote Sens. Environ.* **110**(3), 317-331.
12. Vanuytrecht, E., Raes, D. & Willems, P (2014). Global sensitivity analysis of yield output from the water productivity model. *Environ. Modell. & Softw.* **51**, 323-332.
13. Lagarias, J.C., Reeds, J. A., Wright, M. H. & Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM J. Optimiz.* **9**(1), 112-147.

Development of an Assimilation Scheme for the Estimation of Drought-Induced Yield Losses Based on Multi-Source Remote Sensing and the Aquacrop Model

Silvestro Paolo Cosmo⁽¹⁾, Casa Raffaele⁽¹⁾, Pignatti Stefano⁽²⁾, Yang Hao⁽³⁾, Yang Guijun⁽³⁾

⁽¹⁾ Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy), Email: p.c.silvestro@unitus.it

⁽²⁾ Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy)

⁽³⁾ National Engineering Research Center for Information Technology in Agriculture (NERCITA), 2449-26, Beijing 100097, P.R. China

Abstract:

In the context of the Dragon-3 Farmland Drought project, our research deals with the development of methods for the assimilation of biophysical variables, estimated from multi-source remote sensing, into the AquaCrop model, in order to estimate the yield losses due to drought both at the farm and at the regional scale.

The first part of this project was employed to refine a methodology to obtain maps of LAI, Canopy Cover (CC), FAPAR and Cab from satellite optical data, using algorithms based on the training of ANN on PROSAIL model simulations.

In the second part, retrieved values of Canopy Cover were assimilated into the AquaCrop model using the assimilation method of the Ensemble Kalman Filter.

Introduction:

Drought is one of the variables that most affect crop production of agro-ecosystems causing huge yield losses worldwide. It is estimated that in China drought-affected crops area reaches $20 \cdot 10^6$ hectares, drought disaster area of $8 \cdot 10^6$ hectares, $10 \cdot 15 \cdot 10^6$ ton grain reduction due to drought every year. Therefore methods for quantifying these losses, rapidly from remote sensing are extremely valuable. Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop models is a promising methodology for this aim.

Aquacrop is a water productivity model, which simulates canopy biomass production and yield in response to water transpired by the crop (Steduto et al., 2009). The model simulates canopy cover (CC) along the growing season as a function of three main sets of inputs and parameters concerning climate, crop and soil characteristics. The choice of this crop growth model for our aims was suggested by its relative simplicity as compared to more complex models, by its emphasis on water limitation and by the explicit use, as main canopy state variable, of CC. This variable is usually easier to estimate from remote sensing than LAI, not being subject to saturation at high values.

Our efforts were focused on the retrieval of CC from optical remote sensing data, through a methodology involving the training of artificial neural networks (ANN) on the PROSAIL model (Baret et al., 2007). Successively the assimilation of remotely sensed CC into the Aquacrop model was attempted, through the development of an algorithm based on the general principles of the Ensemble Kalman Filter (EnKF). This method is suitable for assimilation into complex non-linear models and it has been previously used for the same purpose, e.g. with the WOFOST model (De Wit et al., 2012).

This poster presents a preliminary test of the assimilation of CC retrieved from SPOT satellite data, through the use of EnKF into the AQUACROP model, for the estimation of grain yield at the farm scale.

Objectives and study area

The objective of this work is the development of an algorithm for the application of the Ensemble Kalman Filter (EnKF) assimilation method to the AquaCrop model.

The poster presents the preliminary results with an application example based on a case study using farm-scale data collected in Maccarese (Central Italy).

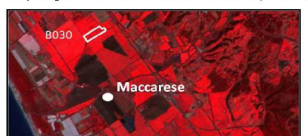


Figure 3: Maccarese farm's field (B030) used for validation of wheat grain yield estimates. The background image is a SPOT 4 image acquired on 2nd March 2012.

Methods:

This study is divided in two parts. The first dealt with the refinement of a methodology (Fig. 2) for retrieving maps of biophysical variables (LAI, CC, FAPAR and Cab) from satellite multispectral data (Baret et al., 2007). A set of preliminary tests was used to calibrate and validate this algorithm using CHRIS Proba and SPOT satellite data (Tab.1).

This algorithm was then employed in order to obtain maps of biophysical crop canopy variables from SPOT4 data acquired in Maccarese during the 2012 wheat growing season.

Canopy Cover (CC) maps obtained in this way (Fig. 5) were then used for the assimilation into the AQUACROP model.

The data assimilation was carried out developing an algorithm based on the Ensemble Kalman Filter (EnKF) method (Makowski et al., 2006). In this method the dynamic model is defined by:

$$\phi_t = F(\phi_{t-1}) + \varepsilon_{t-1}$$

Where:

ϕ_t is the vector including the state variable (in our case CC) and a set of parameters that most affect the model (Tab.2)

ε_{t-1} is the vector of errors.

We selected a set of 13 parameters (Tab.2) to be recalibrated during the assimilation, whereas all other parameters were kept fixed. The 13 parameters were selected according to previous Aquacrop sensitivity analysis studies (Vanuytrec et al., 2014), whereas the fixed parameter values were obtained from the literature and the calibration of the model on data from the Maccarese area.

Data assimilation with the method of Ensemble Kalman Filter were applied step by step as shown in Figure 4.

Figure 2: Steps employed in the tests of the ANN algorithm of Baret et al. 2007 for the estimation of biophysical vegetation variables

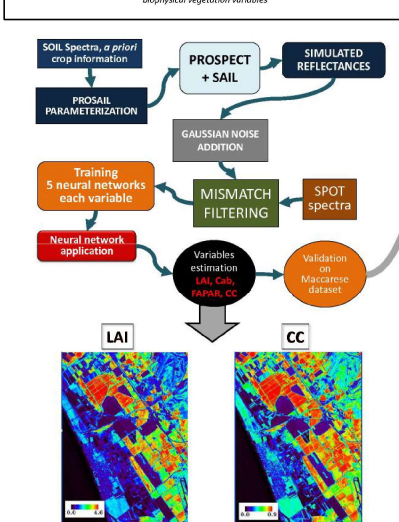
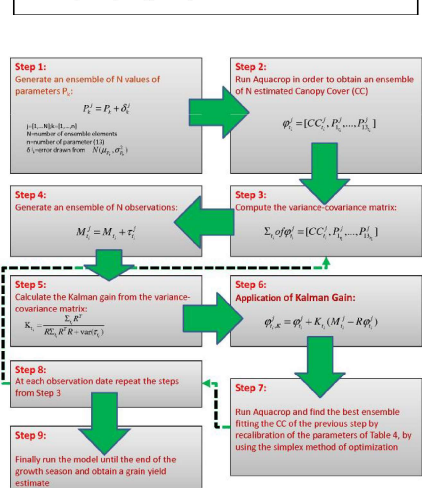


Figure 3: Maccarese area, maps of LAI (on the left) and CC (on the right) obtained from SPOT4 image acquired on 25 March 2012 using the ANN-PROSAIL algorithm

Parameter	Min.	Max.	Mean	St. Dev.
CGC (kg/GDD)	0.03	0.07	0.035	0.0036
CCX (%)	0.65	0.99	0.7	0.07
CDC (kg/GDD)	0.04	0.1	0.10	0.018
GDD_max (GDD)	40	150	132	15
N° of plants/ha	2×10^6	5.5×10^6	4.1×10^6	0.8×10^6
p_sen	0.55	0.75	0.65	0.13
Sh_sen	1.5	3.5	3	0.3
GDD_max (GDD)	323	1500	1067	104.6
PWP (%)	14	39	21.28	8.66
CN	65	85	75	7.07
WC_i (%)	22	54	33.3	8.62
WC_i (%)	14	54	26	11.17
WPa (g/m²)	15	19	17	1.79

Table 1. The crop and soil parameters of Aquacrop model recalibrated during the Ensemble Kalman Filter assimilation application. GDD: growing degree days; CGC: canopy growth coefficient; CCX: maximum canopy cover; CDC: canopy decline coefficient; GDD_max: GDD from sowing to emergence; p_sen: soil water depletion factor for canopy senescence (upper threshold); Sh_sen: shape factor for water stress coefficient for stomatal control; GDD_sow: GDD from sowing to maximum rooting depth; PWP: permanent wilting point; CN: curve number; FC: field capacity; WC_i: initial soil water content; WPa: water productivity.

Figure 4: Ensemble Kalman Filter Assimilation Procedure step by step. EnKF assimilation method is a recursive Bayesian estimation method, which takes into account the true probability density function of the model states conditioned on observations



Dataset used for preliminary test and validation of ANN-PROSAIL algorithm

CROP	n.ESU	LAI			Fcover		
		min	max	mean	min	max	mean
Winter wheat	17	0.12	1	1.45	0.12	0.61	0.29
Durum wheat	7	1.35	2	1.67	0.37	0.48	0.42
Maize	17	0.42	3.12	1.59	0.26	0.80	0.52
Total	3	3.82	5.29	4.41	0.73	0.89	0.69

Table 2: Data ranges and means of LAI and Fcover of the validation set from Maccarese (Central Italy)

Results:

The algorithm based on the training of ANN on PROSAIL model simulations proved to be very effective to obtain maps of LAI, CC, FAPAR and Cab. However, for a more complete assessment of the error and uncertainties in the estimates of the biophysical variables, a more thorough validation would be required. This information is essential in the application of the EnKF algorithm.

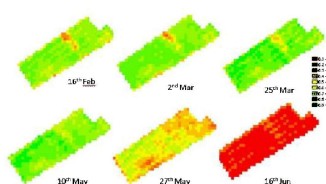


Figure 5: Maccarese area, field B030: maps of green canopy cover (CC) time trend obtained from SPOT4 images during crop cycle. These are the observed (measured) values of CC assimilated into the Aquacrop model

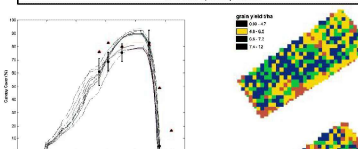


Figure 6: Time trend of canopy cover (i.e. CC) simulated by the ensemble of 10 Aquacrop model runs and the observed CC from SPOT4 data (red triangles) and the mean (black dot) and range of variation (black bar) of CC values estimated from the EnKF.

Figure 7: Wheat grain yield on B030 field (a) measured by combine harvester and (b) estimated by from assimilation of canopy cover into Aquacrop

Discussion

This preliminary test of the application of the Ensemble Kalman Filter method to Aquacrop, has shown its adequacy to provide yield estimation (Fig. 7) as affected by water availability. It has highlighted, however, aspects that need further work and improvements:

- Choice of parameters to recalibrate.** In Aquacrop there are about 100 parameters related to the characteristics of the crop and the soil. With the aid of a preliminary calibration and sensitivity studies taken from the literature (Vanuytrec et al., 2014), we chose 13 parameters subject to recalibration during the EnKF algorithm application. The choice was rather arbitrary, based on our assessment of which parameters we considered to be the most influential (Tab.4). For a more objective choice, a specific sensitivity analysis is required.
- Fixed Parameters Values.** The choice of values to assign to the other parameters which remained fixed, was based on the literature and on a preliminary calibration. For applications outside our test site a proper calibration of the fixed parameters is needed.
- Variable Parameters Values.** The values of variable parameters are chosen from a Normal distribution with an assigned mean and standard deviation. The choice of these mean and st. deviation are based on literature and a priori hypothesis. They have a large influence on the spread of the values simulated in the ensemble (Fig. 7).
- Method of Optimization.** The impossibility in the executable of the Aquacrop model to implement a re-initialisation at the observation dates, lead us to recalibrate it at each date by using sequentially the CC estimates from the Kalman filter in an optimization method. The high computation time required by the optimization, imposed a small ensemble size (N = 10).
- Measurement and Model Error estimates.** This preliminary test showed that the error of the model (which was assumed only dependent on the variance of the parameters) was smaller than the error of the observations (Fig.7). Therefore, it is necessary: 1) to introduce in an appropriate manner a total error on the model; 2) to have more realistic estimates of CC observation error.

Major References:

- Baret, F., Hagolle, C., Gelber, B., Bicheron, P., Miras, B., Huc, M., ... & Leroy, M. (2007). LAI, FAPAR and Fcover CYCLOPS global products derived from VEGETATION, Remote Sens. Environ. 103(3), 279-286.
- de Wit, A., G. Duvellet, and P. Defourny. 2012. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. Ag. For. Meteorol. 164: 39-52.
- Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal of Optimization, 8, 1117-1147.
- Makowski, D., Gueff, M., Jones, J.W., Graham, W. (2006). Data assimilation with crop models. In: "Working with Dynamic Crop Models", Wallach, D., Makowski, D. and Jones, J. W. (eds.), Elsevier, 151-172.
- Steduto, P., Iqbal, T., Rötter, R., & Fereres, E. (2009). Aquacrop: the FAO crop model to simulate yield response to water I: concepts and underlying principles. Agric. J. 120(1), 426-432.
- Vanuytrec, E., Rötter, R., & Williams, P. (2014). Global sensitivity analysis of yield output from the water productivity model. Environmental Modelling & Software, 51, 33-50. doi:10.1016/j.envsoft.2013.10.017