

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Scuola di Ingegneria e Architettura
Corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

UN SISTEMA DI PREVISIONE ADATTIVA
DELL'UMIDITÀ DEL SUOLO

Elaborato in
Data Mining

Relatore

Prof. Matteo Golfarelli

Co-relatore

Dott. Joseph Giovanelli

Presentata da

Marco Meluzzi

Seconda Sessione di Laurea
Anno Accademico 2022 - 2023

PAROLE CHIAVE

Agricoltura di precisione

Umidità del terreno

Previsione adattiva

Machine Learning

Reti Neurali

Indice

Sommario	xi
Introduzione	xiii
1 Definizione del problema	1
1.1 Suolo	4
1.2 Proprietà fisico-idrologiche del suolo	6
1.3 Sensoristica	8
1.4 Struttura del campo	10
2 Stato dell'arte	15
2.1 Modelli fisico-numerici	16
2.2 Modelli a media mobile	18
2.3 Modelli di Machine Learning	18
2.3.1 Approcci classici	19
2.3.2 Approcci legati al Deep Learning	21
2.4 Automated Machine Learning	24
3 Orchard3D-Lab	29
3.1 Modello di simulazione	30
3.2 Modulo di assimilazione dei dati	34
3.3 Modulo di tuning dei parametri	35
4 Approccio	37
4.1 Creazione rolling window	39
4.2 Definizione del piano irriguo	41
4.3 Modello di previsione	42
4.4 Adattamento del modello di previsione nel tempo	43
5 Validazione	45
5.1 Configurazione	45
5.1.1 Analisi dei dati meteorologici	46
5.1.2 Scelta dei piani irrigui	46

5.1.3	Scelta delle tessiture	50
5.1.4	Suddivisione dei dataset di training, validation e testing .	53
5.1.5	Selezione degli algoritmi e degli spazi di ricerca	53
5.1.6	Valutazione dell'errore del modello	56
5.1.7	Adattamento del modello	57
5.2	Esperimenti	58
5.2.1	Test dei modelli su scenari conosciuti	58
5.2.2	Test dei modelli su scenari non conosciuti	60
5.2.3	Adattamento dei modelli su scenari non conosciuti	62
6	Conclusioni e sviluppi futuri	71
	Bibliografia	73

Elenco delle figure

1.1	Volume di produzione di kiwi nel mondo nel 2021, per Paese (in migliaia di tonnellate)	2
1.2	Triangolo per la determinazione della classe granulometrica del suolo	5
1.3	Curva di ritenzione idrica per differenti tessiture, figura da [1] .	7
1.4	Assunzione di simmetria per una griglia 2D. Le parti tratteggiate replicano i valori misurati dai sensori	11
1.5	Modellazione del layout del campo	12
1.6	Layout del frutteto di kiwi situato a Errano (RA)	13
1.7	Layout del sistema di irrigazione e della griglia di sensori	14
1.8	Layout dettagliato della griglia di sensori	14
2.1	Ottimizzazione bayesiana, figura da [2]	26
2.2	Step di SMBO, figura da [2]	26
3.1	Panoramica e principali componenti di Orchard3D-Lab	30
3.2	Proiezioni della densità delle radici sui piani xy (a sinistra) e xz (a destra)	34
3.3	Esempio di funzione di profilazione su una griglia 2D di sensori .	35
4.1	Scenario di applicazione del sistema di previsione dell'umidità del suolo	38
4.2	Scenario di adattamento del sistema di previsione dell'umidità del suolo	39
4.3	Processo generale di creazione delle finestre sovrapposte	40
4.4	Pipeline di ML adottata	42
4.5	Processo di adattamento del modello di previsione nel tempo . .	43
5.1	Dati meteorologici reali rilevati a Martorano nel periodo giugno-settembre 2017	47
5.2	Dati di irrigazione simulati mediante Orchard3D-Lab, in riferimento allo scenario meteo di Martorano, alla tessitura di Errano e all'intervallo di irrigazione di 2 giorni (periodo dal 14 giugno al 16 giugno 2017)	48

5.3	Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 2 giorni	49
5.4	Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 4 giorni	49
5.5	Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 7 giorni	50
5.6	Tre differenti tessiture selezionate	51
5.7	Andamento dell'umidità del suolo SANDY, relativamente allo scenario meteo di Martorano 2017 e all'intervallo di irrigazione di 7 giorni	52
5.8	Andamento dell'umidità del suolo SILTY, relativamente allo scenario meteo di Martorano 2017 e all'intervallo di irrigazione di 7 giorni	52
5.9	Processo di adattamento del modello di previsione da una tessitura e l'altra o da un piano irriguo e l'altro	57
5.10	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore, relativamente alla tessitura franco argillosa, all'intervallo di irrigazione di 2 giorni e all'orizzonte di 4 giorni .	59
5.11	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore con il modello base, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni	65
5.12	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il primo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni	66
5.13	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il secondo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni	67
5.14	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il terzo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni	68
5.15	Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il terzo ciclo di adattamento, passando dalla tessitura franco argillosa a quella franco limosa, con intervallo di irrigazione di 7 giorni e orizzonte di 7 giorni	70

Elenco delle tabelle

1.1	Proprietà del suolo al variare della classe granulometrica	6
1.2	Confronto delle diverse tipologie di sensori in termini di qualità del servizio	10
2.1	Riassunto degli approcci classici di ML	21
2.2	Riassunto degli approcci di Deep Learning	23
3.1	Parametri richiesti per la simulazione numerica	31
5.1	Spazio di ricerca per Linear Regression	54
5.2	Spazio di ricerca per Random Forest	55
5.3	Spazio di ricerca per SVR	55
5.4	Spazio di ricerca per le reti neurali	56
5.5	Risultati dei modelli di base per tutti gli algoritmi, le tessiture, i piani irrigui e gli orizzonti	59
5.6	Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 2 giorni	60
5.7	Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 4 giorni	61
5.8	Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 7 giorni	61
5.9	Risultati da tessitura franco argillosa a franco sabbiosa, con intervallo di irrigazione di 7 giorni	62
5.10	Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il primo ciclo di adattamento . . .	62
5.11	Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il secondo ciclo di adattamento .	63

5.12	Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il terzo ciclo di adattamento . . .	63
5.13	Risultati da intervallo di irrigazione di 7 giorni a 2 giorni, con tessitura franco argillosa	64
5.14	Risultati adattamento reti da intervallo di irrigazione di 7 giorni a 2 giorni, con tessitura franco argillosa	64
5.15	Risultati adattamento reti da tessitura franco argillosa a franco limosa, con intervallo di irrigazione di 7 giorni	69
5.16	Risultati adattamento reti da tessitura franco argillosa con intervallo di irrigazione di 7 giorni a tessitura franco limosa con intervallo di irrigazione di 2 giorni	69

Sommario

L'agricoltura di precisione ha assunto un'importanza centrale in tempi recenti, complici le mutate condizioni climatiche a livello globale che causano periodi di siccità sempre più frequenti. Per questo motivo, conoscere in anticipo quelle che saranno le condizioni di umidità del suolo in un certo istante nel futuro è fondamentale per poter supportare i processi di irrigazione in campo agricolo. Prendendo come riferimento una reale piantagione di kiwi situata ad Errano, nella regione Emilia-Romagna, dotata di un impianto di irrigazione a goccia e di una griglia di sensori bi-dimensionale per la rilevazione dell'umidità del terreno in tempo reale, questo lavoro propone un sistema di previsione dell'umidità del suolo basato su reti neurali, in grado di adattarsi in maniera veloce a diverse tipologie di terreno e a differenti schemi di irrigazione. Servendosi di un apposito simulatore di flussi idrologici, è possibile costruire in tempi più rapidi un modello di previsione di base, specifico per un certo terreno o piano irriguo. Tale modello, può essere in seguito migliorato con l'avanzare della campagna produttiva e con l'arrivo in tempo reale di nuovi dati dal campo, oppure adattato ad un altro schema di irrigazione o ad un terreno con differenti caratteristiche fisiche rispetto quello di partenza.

Introduzione

L'agricoltura è da sempre una delle attività fondamentali dell'essere umano, sia per quanto riguarda il sostentamento in sé sia dal punto di vista economico, facendo parte del settore primario. Come ogni altra attività umana, essa ha tratto benefici e si è evoluta di pari passo con l'avanzamento tecnologico, passando attraverso diverse fasi e rivoluzioni capaci di influenzare lo sviluppo stesso delle società.

Una delle ultime frontiere è la cosiddetta *agricoltura di precisione*, che si pone l'obiettivo di ottimizzare la produzione in modo che sia di qualità e al tempo stesso sostenibile. È quindi possibile descriverla come l'insieme di tecniche e di metodi in grado di applicare "il trattamento giusto, nel luogo giusto e al momento giusto" [3]. Per fare ciò, è necessario raccogliere e saper interpretare correttamente i dati provenienti dai campi, possibilmente in tempo reale, e applicare di conseguenza lo specifico trattamento, che può riguardare ad esempio l'irrigazione o l'applicazione di concimi e fertilizzanti. Vengono quindi sviluppati sistemi *data-driven*, capaci cioè di prendere decisioni sfruttando direttamente i dati raccolti, che possono spaziare da semplici strumenti di monitoraggio e controllo fino a veri e propri strumenti di previsione.

Tutto questo si traduce in un'effettivo vantaggio per le aziende agricole da una parte, che possono aumentare la produttività dei terreni andando a diminuire al contempo le risorse utilizzate con una conseguente diminuzione dei costi, e per l'ambiente dall'altra, riducendo gli sprechi di acqua o di altre sostanze immesse nel terreno.

L'urgenza di applicare e rendere più diffuse ed efficienti queste metodologie e processi è esacerbata dal contesto climatico corrente e del prossimo futuro, caratterizzato da eventi estremi quali la siccità, che sempre più frequentemente affligge anche il nostro territorio. Non da ultimo, l'aumento del fabbisogno alimentare derivante dalla fortissima crescita della popolazione a livello mondiale: secondo il rapporto congiunto dell'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) e della Food and Agriculture Organization of the United Nations (FAO) sulle prospettive agricole 2022-2031 [4] si prevede, infatti, che il consumo alimentare globale aumenterà dell'1.4% annuo nel prossimo decennio, mentre la produzione agricola globale aumenterà dell'1.1% all'anno.

Alle motivazioni appena elencate si aggiungono le pubblicazioni inerenti al tema dell'agricoltura di precisione che nell'ultimo periodo hanno avuto un andamento sempre crescente, con un incremento del 300% negli ultimi sette anni [5].

Il lavoro di seguito presentato parte dal progetto *Agro.Big.Data.Science* [6], nato dall'esigenza di diverse realtà produttive locali di sfruttare i *Big Data*¹ a loro disposizione relativi alla gestione delle filiere produttive del kiwi, del pero e dello spinacio, al fine di ottenere da questi le informazioni necessarie all'efficientamento di vari processi che impattano sulla sostenibilità economica e ambientale. Le tre filiere in questione sono dotate della sensoristica necessaria per l'acquisizione di dati in tempo reale dai campi, che vengono usati congiuntamente a dati e informazioni provenienti da database pubblici (*e.g.*, dati meteo). Ad alto livello, il progetto si articola in 5 fasi distinte:

- **Fase 1:** analisi delle filiere produttive e comprensione dei dati;
- **Fase 2:** sviluppo di modelli di analisi;
- **Fase 3:** analisi dell'evoluzione tecnologica, progettazione e deploy della piattaforma Big Data;
- **Fase 4:** sviluppo di algoritmi e relativi testing per la stima delle rese, della raccolta e dell'irrigazione in diverse piantagioni;
- **Fase 5:** validazione dei risultati.

Facendo riferimento alla filiera del kiwi, il contesto del problema che verrà analizzato è quello relativo alla predizione dell'umidità del suolo finalizzata all'elaborazione di un piano irriguo rilasciato giornalmente per aumentare l'efficienza di consumo dell'acqua in impianti dotati di sensoristica di precisione. Nello specifico, il contributo apportato al problema può essere riassunto nei seguenti punti fondamentali:

- formalizzazione del problema;
- studio dell'effetto di diversi piani irrigui applicati a differenti tipi di terreno (*i.e.*, tessiture) tramite Orchard3D-Lab, un simulatore di flussi idrologici;

¹Con il termine *Big Data* si fa riferimento a una raccolta di dati così estesa in termini di volume, velocità di produzione e di consumo, varietà e attendibilità, tale da non poter essere gestita con i tradizionali sistemi di memorizzazione e di analisi.

- creazione di dataset sintetici sfruttando Orchard3D-Lab e addestramento di modelli basati su reti neurali in grado di predire l'umidità del campo in diversi piani irrigui e tessiture;
- applicazione di tecniche di adattamento dei modelli a nuovi terreni e piani irrigui su un'architettura di acquisizione dati in tempo reale;
- analisi e confronto delle capacità dei modelli nel predire l'umidità del suolo su terreni e piani irrigui conosciuti e, a seguito delle tecniche di adattamento, su nuovi terreni e piani irrigui.

La tesi è quindi organizzata come segue: nel Capitolo 1 viene fornita la definizione generale del problema e viene introdotta la terminologia e i concetti agronomici specifici, nel Capitolo 2 si illustrano i precedenti lavori e risultati in letteratura, nel Capitolo 3 si descrive il simulatore utilizzato per generare i dati sintetici necessari all'addestramento dei modelli, nel Capitolo 4 viene presentato l'approccio adottato, nel Capitolo 5 la validazione dei risultati e infine nel Capitolo 6 le conclusioni e gli sviluppi futuri.

Capitolo 1

Definizione del problema

Il lavoro qui presentato si colloca nel contesto di un progetto più ampio, volto a trarre vantaggio dall'utilizzo di una gran mole di dati provenienti da diverse filiere agroalimentari. I campi che compongono queste filiere produttive sono infatti dotati di diversi sensori che, congiuntamente a fonti esterne (*e.g.*, dati dell'Agenzia Regionale per la Prevenzione, l'Ambiente e l'Energia dell'Emilia-Romagna (ARPAE), immagini satellitari della European Space Agency (ESA)), possono raccogliere informazioni importanti per supportare in maniera più puntuale il lavoro dei contadini, fornendo loro una panoramica dell'andamento della campagna produttiva, la stima delle rese e della data di raccolta, un consiglio irriguo giornaliero e delle dashboard interattive con cui consultare gli output provenienti dai campi in tempo reale. L'obiettivo ultimo di tutto ciò è quello di interpretare correttamente i dati in modo tale da prendere decisioni più consapevoli e quindi rendere più efficiente il processo produttivo secondo diversi parametri. Così facendo, aumenta la quantità e la qualità della produzione da un lato, e si riducono gli sprechi dall'altro, attraverso una gestione più oculata delle risorse.

Questo lavoro si concentra sulla filiera del kiwi e sull'utilizzo dei dati qui raccolti per cercare di ottenere una predizione quanto più accurata dell'umidità del suolo dei campi nel tempo, in modo tale da conoscere con un certo anticipo quello che dovrebbe essere il pattern di asciugatura del terreno e poter quindi programmare in maniera intelligente le irrigazioni. Un processo di questo tipo consente infatti di risparmiare l'acqua immessa nel frutteto, considerando che è tendenza comune irrigare con un quantitativo superiore a quello strettamente necessario, allo scopo di ottenere frutti più grandi. Tale pratica riduce però la massa secca del frutto, pregiudicandone il mantenimento dopo la raccolta.

La pianta da cui ha origine il frutto del kiwi, denominata *Actinidia deliciosa*, è originaria della Cina meridionale, che costituisce il primo produttore a livello mondiale. Tra l'Ottocento e il Novecento si diffuse dapprima in Europa e poi

in Nuova Zelanda, dove sorsero le prime vere e proprie coltivazioni intensive di kiwi. Ad oggi, l'Italia si posiziona al terzo posto, con una produzione annua che si è attestata intorno alle 416.060 tonnellate nel 2021 [7], come mostrato in Figura 1.1.

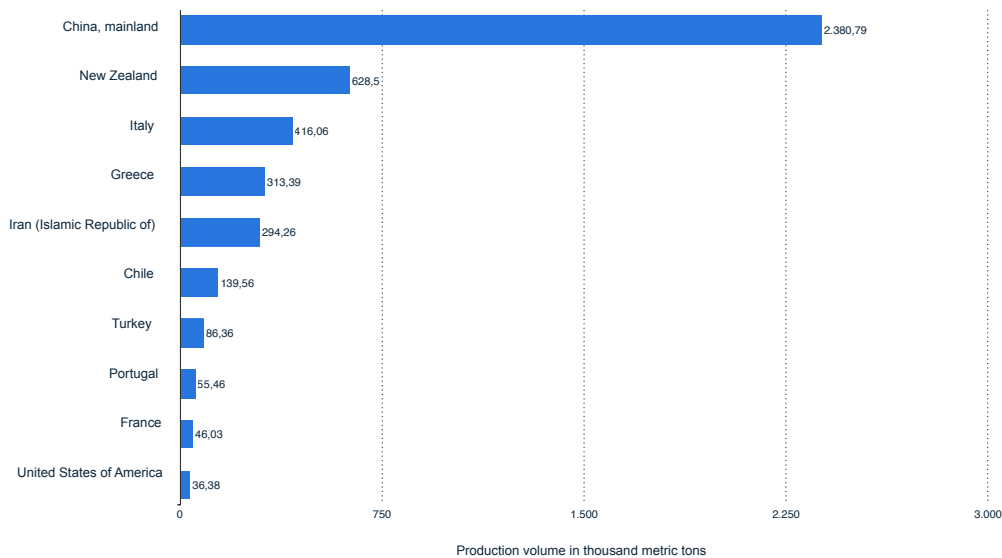


Figura 1.1: Volume di produzione di kiwi nel mondo nel 2021, per Paese (in migliaia di tonnellate)

Per quanto riguarda la produzione nazionale, la regione Emilia-Romagna, dove si trova la piantagione di kiwi di riferimento per questo lavoro, si posiziona al secondo posto a livello di superficie impiegata per tale coltivazione, con una quota del 20% sul totale. Inoltre, nel biennio 2021-2022 la produzione ha riscontrato un incremento di circa il 32% nella regione [8].

Raccolto nel periodo autunnale, il kiwi nasce da una pianta rampicante che può raggiungere i 10 metri di altezza: per questo motivo, necessita di appositi sostegni a supporto di una crescita ordinata. Essa predilige un clima temperato e zone soleggiate, al riparo dal vento forte che rischia di spezzarne le lunghe ramificazioni.

La coltivazione del kiwi si distingue per le elevate esigenze idriche, soprattutto nel periodo che va da maggio ad agosto, caratterizzato da alte temperature e da una bassa probabilità di pioggia, e che corrisponde alle fasi di fioritura, allegagione e sviluppo dei frutti. Allo stesso tempo, la pianta è molto sensibile sia alla carenza idrica che agli eccessi idrici. Questi ultimi determinano una sfavorevole interazione fra suolo e radici, danneggiando il molto superficiale apparato radicale e portando quindi alla moria della pianta. Tale situazione

si verifica in modo particolare nelle aree in cui si ha una piovosità invernale elevata e dove i terreni non drenano bene poiché poveri o del tutto privi di sostanza organica. Su un suolo di questo tipo, l'accumulo delle piogge, assieme ad una errata gestione irrigua, può formare dei ristagni idrici e rendere il terreno asfittico, vale a dire che la circolazione dell'aria nei pori non è sufficiente per l'eccessiva quantità di acqua. In queste condizioni, l'apparato radicale della pianta collassa e perde la capacità di assorbire i nutrienti. Nella situazione opposta di carenza idrica, invece, la pianta può andare incontro al disseccamento parziale o totale del lembo fogliare.

A partire dal mese di agosto fino ad ottobre, l'apporto idrico deve essere progressivamente ridotto, al fine di non compromettere la qualità del frutto. In generale, il kiwi ha quindi bisogno di irrigazioni brevi e frequenti. Gli impianti di irrigazione ideali allo scopo sono quelli a goccia, capaci di mantenere il terreno umido in maniera omogenea, irrigando in modo preciso e localizzato, senza spreco d'acqua. Un'ala gocciolante viene installata lungo il filare su un filo di ferro appeso tramite appositi ganci di sostegno, in modo da non intralciare la lavorazione del terreno. In base alle specifiche caratteristiche del suolo ed esigenze irrigue, è possibile regolare la portata e i passi tra i diversi gocciolatori, affidandosi eventualmente ad una doppia ala gocciolante. È possibile altresì integrare l'impianto a goccia con un sistema di irrigazione a pioggia, che garantisce protezione da eventuali brinate o temperature elevate. Introducendo microirrigatori sopra e sotto la chioma, la pioggia generata crea un microclima capace di salvaguardare gemme e fiori del kiwi ed evitare così di compromettere il futuro raccolto. Lo svantaggio di quest'ultima tipologia di irrigazione, nota anche come microaspirazione, è il fatto che richiede una quantità d'acqua superiore rispetto all'irrigazione a goccia, ma permette d'altra parte di fornire alle radici una risposta idrica più veloce. Altre tipologie di impianti possono essere la microaspirazione tramite microirrigatori a corta gittata posizionati a terra mediante appositi picchetti di supporto, e la subirrigazione, che permette alla pianta di ricevere l'acqua direttamente alle radici grazie all'ala gocciolante interrata [9, 10, 11, 12].

Di seguito, la sezione 1.1 descrive le caratteristiche fisiche del suolo essenziali per poter classificare le varie tipologie di terreno, la sezione 1.2 introduce diversi modi di modellare il flusso idrologico del suolo e quantificarne il contenuto di umidità, la sezione 1.3 offre una breve panoramica sulle differenti tipologie di sensori impiegabili per misurare l'umidità del terreno, infine la sezione 1.4 illustra il layout e le caratteristiche specifiche del campo reale di riferimento.

1.1 Suolo

Conoscere a fondo le specifiche caratteristiche e proprietà del suolo è di fondamentale importanza per una sua corretta gestione e per garantire una buona resa del raccolto e il mantenimento della piantagione nel tempo.

Il suolo in quanto tale, risulta essere la combinazione di sostanza organica e sostanza minerale che permette la vita di piante e animali, ed è pertanto una caratteristica esclusiva della Terra. Basti pensare che in esso è contenuta un'ampia varietà di organismi viventi (microbiota), per la maggior parte ancora ignoti, tale da rappresentare l'habitat con più biodiversità sul pianeta [13]. Le sue funzioni principali comprendono la produzione di biomassa, la mitigazione degli effetti delle sostanze inquinanti fungendo da barriera ed ostacolandone il passaggio nella catena alimentare e nelle falde sotterranee, e infine la capacità di preservare il proprio habitat biologico e la rispettiva riserva genetica che ospita.

Dal punto di vista di una pianta, il suolo costituisce il mezzo attraverso il quale essa assorbe le sostanze nutrienti mediante le sue radici. La dinamica che tali sostanze seguono, in modo particolare l'acqua, è direttamente influenzata dalla tessitura, la proprietà fisica attraverso la quale è possibile classificare il suolo in base alla composizione percentuale delle particelle solide che lo compongono. La parte inorganica del suolo è infatti costituita da particelle di dimensioni diverse che determinano tre distinte classi granulometriche:

- **sabbia** (*sand*): diametro compreso tra 2 e 0.05 millimetri;
- **limo** (*silt*): diametro compreso tra 0.05 e 0.002 millimetri;
- **argilla** (*clay*): diametro inferiore a 0.002 millimetri.

La specifica tessitura indica quindi la distribuzione percentuale delle particelle di sabbia, limo e argilla in un suolo, determinandone le proprietà fisiche, chimiche e meccaniche del terreno stesso, come la plasticità¹, la permeabilità all'aria e all'acqua, la capacità di trattenere l'acqua e la disponibilità di elementi nutritivi.

La tassonomia ufficiale prevede 12 classi granulometriche distinte, che possono essere facilmente determinate servendosi di un apposito triangolo², raffigurato in Figura 1.2. Conoscendo le percentuali delle tre componenti sopra menzionate, è sufficiente seguire le linee corrispondenti: l'area in cui ricade il punto di intersezione delle tre linee determina la specifica tipologia di suolo.

¹La plasticità è la proprietà del terreno di subire una deformazione e di mantenerla quando cessa l'azione della forza che l'ha generata.

²Il triangolo della tessitura del suolo fu sviluppato per la prima volta nel 1911 da Milton Whitney e fu revisionato nel 1927 da Davis e Bennett.

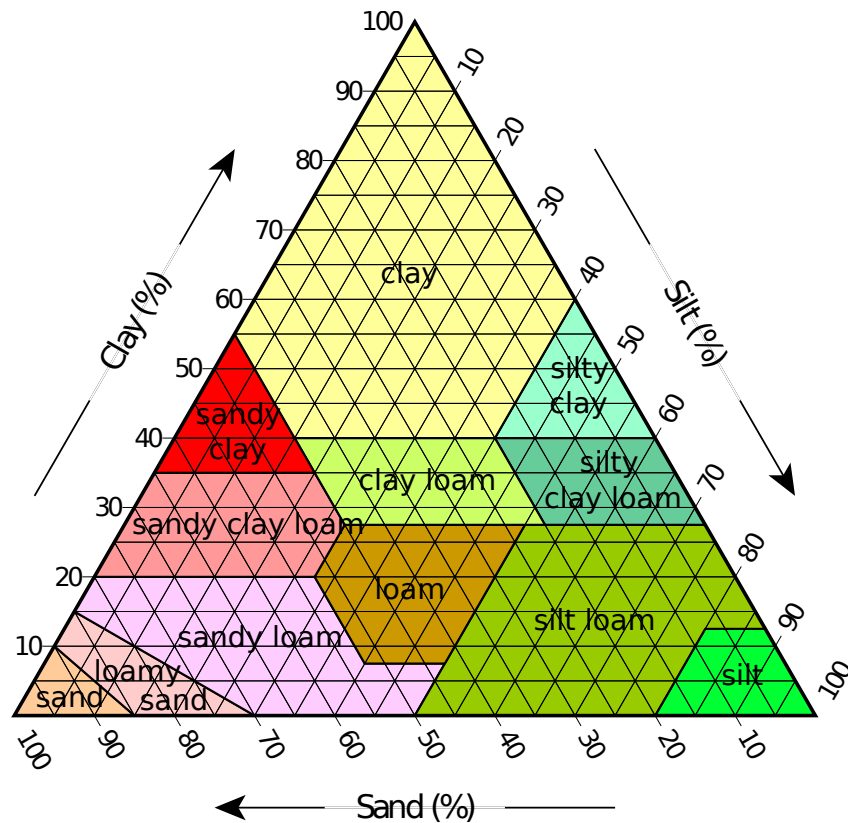


Figura 1.2: Triangolo per la determinazione della classe granulometrica del suolo

Le caratteristiche fisiche della sabbia, del limo e dell'argilla differiscono molto tra loro, pertanto la predominanza di una frazione sull'altra influenza fortemente le caratteristiche fisiche e agronomiche del terreno. Quando la componente sabbiosa predomina, il terreno risulta essere ben areato e dotato di buon drenaggio, di conseguenza la sua capacità di trattenere l'acqua e gli elementi nutritivi è molto scarsa e tenderà ad asciugarsi in fretta. Al contrario, i terreni con prevalenza di argilla sono caratterizzati da scarsa areazione e difficile drenaggio, quindi rimangono più a lungo umidi. Infine, i terreni limosi hanno caratteristiche simili a quelli argillosi, con una capacità di trattenere gli elementi nutritivi che si posiziona a metà tra i suoli sabbiosi e quelli argillosi. La Tabella 1.1 riassume come le diverse proprietà del suolo vengono influenzate in base alla tessitura [14].

Quando le percentuali delle tre componenti sono abbastanza equilibrate, si parla di terreni "franchi" (*loam*), e sono quelli solitamente di maggior interesse agricolo.

Proprietà del suolo	Tipologia del suolo		
	Sabbioso	Limoso	Argilloso
Areazione	Eccellente	Buona	Scarsa
Drenaggio	Eccellente	Buono	Scarso
Capacità di trattenere i nutrienti	Bassa	Media	Alta
Capacità di trattenere l'acqua	Bassa	Media	Alta

Tabella 1.1: Proprietà del suolo al variare della classe granulometrica

1.2 Proprietà fisico-idrologiche del suolo

L'umidità di un terreno, o *soil moisture*, (SM) è il risultato combinato di diversi processi che avvengono nel suolo, tra cui l'infiltrazione di acqua, la redistribuzione, la percolazione, l'evaporazione e la traspirazione delle piante. Si definisce *field capacity* (FC) la quantità massima di SM che il terreno può trattenere dopo che l'acqua in eccesso è stata drenata e la velocità del movimento verso il basso diminuisce, cosa che di solito, in terreni permeabili con struttura e consistenza uniformi, avviene entro 2-3 giorni dopo una pioggia o un'irrigazione [15].

Il punto di avvizzimento, o *wilting point*, rappresenta invece la quantità minima di acqua trattenuta dal suolo prima che una pianta inizi ad appassire definitivamente. Se il contenuto di acqua del terreno scende ulteriormente al di sotto di questo valore, la pianta non è in grado di riacquistare il turgore cellulare.

Per stimare il fabbisogno idrico delle colture e gestire la programmazione dell'irrigazione viene spesso introdotto il concetto di *plant available water*. Definito come la differenza tra la *field capacity* e il *wilting point*, esso rappresenta la quantità massima di acqua immagazzinata in un profilo di suolo che può essere effettivamente utilizzata dalla pianta.

L'umidità del suolo può essere espressa in diversi modi. La misura più comunemente utilizzata è il *volumetric soil water content* (SWC), che rappresenta la quantità di acqua percentuale contenuta nel suolo.

Un'altra misura spesso sfruttata per la SM è il *soil water potential* (SWP), che indica l'energia potenziale dell'acqua nel suolo. Ogni componente di un sistema possiede, infatti, dell'energia libera capace di compiere lavoro in condizioni di temperatura costante. Nello specifico, il potenziale idrico quantifica il lavoro che le piante devono compiere nell'assorbimento radicale. Il SWP viene quantificato prendendo come riferimento l'acqua pura (*i.e.*, senza soluti) libera da forze esterne (esclusa la gravità), ad una pressione, temperatura e altitudine di riferimento. L'acqua pura assume pertanto un potenziale pari a zero. La presenza di particelle di soluto riduce l'energia

libera dell'acqua e di conseguenza il potenziale idrico. Per questo motivo, il SWP assume sempre valori inferiori a zero. Il range ammissibile per il SWP è solitamente compreso tra -33 e -1500 cbar: ciò significa che l'estremo superiore (-33 cbar) è il potenziale associato alla *field capacity*, mentre l'estremo inferiore (-1500 cbar) è il potenziale associato al *wilting point*.

La quantità volumetrica di acqua nel suolo e il potenziale idrico possono essere messi in relazione attraverso la curva di ritenzione idrica. Come si vede in Figura 1.3, la curva cambia a seconda della tessitura del suolo: più grandi sono le particelle (*e.g.*, nei terreni sabbiosi), più l'acqua viene rilasciata a potenziali maggiori. Al contrario, quando le particelle che costituiscono il suolo sono più piccole (*e.g.*, nei terreni argillosi), l'acqua viene rilasciata a potenziali minori. Ciò trova riscontro con quanto illustrato nella sezione precedente: i terreni con preponderanza di sabbia, rispetto a quelli in prevalenza argillosi, hanno una minore capacità di trattenere l'acqua e di conseguenza si asciugano più in fretta.

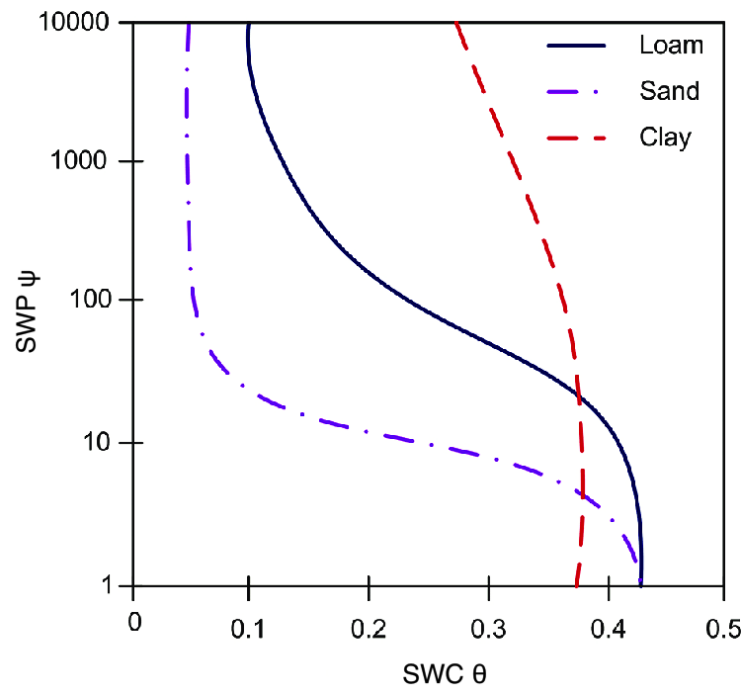


Figura 1.3: Curva di ritenzione idrica per differenti tessiture, figura da [1]

La curva di ritenzione idrica può essere determinata in maniera sperimentale oppure sfruttando le cosiddette funzioni di pedotrasferimento. Tali funzioni derivano da diverse forme di analisi di regressione che consentono di ottenere le proprietà idrauliche del suolo dalla conoscenza delle proprietà fisiche di base

come la tessitura, la densità apparente e il contenuto di sostanza organica del terreno [16].

La stima di SWC o SWP può avvenire secondo diverse modalità, che vanno dal rilevamento diretto mediante appositi sensori situati nel terreno, alla simulazione usando modelli numerici. Questi ultimi, hanno il vantaggio di essere più economici rispetto all'installazione di sensori fisici e permettono di ottenere una grande quantità di dati in tempi più rapidi.

Nella sezione successiva vengono illustrate le varie tipologie di sensori che possono essere impiegate per la misurazione diretta dell'umidità del suolo.

1.3 Sensoristica

Ad alto livello, i sensori che possono essere adottati per il monitoraggio dell'umidità del terreno ricadono fondamentalmente in due categorie distinte:

- **i sensori remoti**, come suggerisce il nome, sono quelli che riescono a monitorare il terreno da una certa distanza e sono pertanto indicati per grandi appezzamenti di terreno. Non riescono quindi a fornire una misurazione in profondità dell'umidità, oltre ad essere ostacolati dall'eventuale presenza di chiome voluminose delle piante. Per questi motivi, la loro applicabilità in un contesto di irrigazione di precisione è limitata. In riferimento alla filiera del kiwi, non possono essere impiegati a causa dell'ingombro generato dalle piante stesse e dell'impossibilità di effettuare misurazioni in profondità, direttamente nel terreno. In questa categoria ricadono a loro volta diverse tipologie di sensori:
 - **Unmanned Aerial Vehicle (UAV)**, rappresentano i veicoli aerei che non richiedono un intervento umano per poter volare e quindi possono essere gestiti da remoto. Ricadono in questa categoria i droni agricoli, impiegati principalmente per la mappatura dei terreni, ma adatti anche alla rilevazione dell'umidità del suolo.
 - **satellitari**, permettono di controllare in tempo reale lo stato di salute dei campi tramite un'analisi spettrale di immagini satellitari ad alta risoluzione. Infatti, i satelliti sono in grado di rilevare le emissioni elettromagnetiche di alcune lunghezze d'onda della banda del visibile e vicine all'infrarosso: in questo modo, riescono a catturare la luce solare riflessa dalle piante e dal suolo e produrre una sorta di radiografia del campo. Le informazioni che possono fornire spaziano dallo sviluppo vegetativo delle piante, tramite la rilevazione di parametri come il Normalized Difference Vegetation Index (NDVI), che aiuta a individuare le piante fotosinteticamente attive

e quindi in salute, alle informazioni sulla temperatura e umidità del suolo, attraverso il Normalized Difference Moisture Index (NDMI), che grazie alla rilevazione del tasso di umidità permette di individuare lo stress idrico delle colture [17]. I sensori satellitari hanno il vantaggio di non richiedere complesse installazioni che necessitano di manodopera, oltre al fatto di non interferire con le attività nel campo.

- i **sensori in-situ**, sono quelli che vengono installati direttamente nel sottosuolo, nella zona radicale della pianta, e hanno quindi il vantaggio di poter essere utilizzati anche in spazi chiusi come serre e vivai. La loro vicinanza alla piantagione li rende particolarmente precisi ed affidabili. In base alla specifica tecnologia che usano, si dividono in:

- **volumetrici**, che misurano il contenuto volumetrico percentuale di acqua nel terreno in maniera indiretta, sfruttando altre proprietà del suolo, dal momento che una misurazione diretta richiederebbe di estrarre un campione di suolo e procedere con l'asciugatura e la pesatura. Esistono diverse tecnologie specifiche di sensori volumetrici, che si differenziano per la particolare caratteristica del suolo che esaminano per stimarne l'umidità. Sono in grado di fornire misurazioni accurate, ma tipicamente necessitano di una calibrazione specifica per la tipologia di suolo in cui sono installati.
- **basati sul potenziale**, che misurano il potenziale del terreno. Rispetto ai sensori volumetrici, sono più economici e più semplici da usare, anche se necessitano di una manutenzione più attenta, essendo sensibili alle temperature più fredde.

In conclusione, la scelta di una tipologia di sensore piuttosto che di un'altra deve essere guidata dai requisiti specifici e dalle garanzie che si vogliono avere. La Tabella 1.2 confronta i diversi tipi di sensori appena descritti in termini di qualità del servizio [18].

A seconda del layout specifico che si vuole implementare nel campo, è possibile individuare due diverse installazioni di sensori *in-situ*:

- **singolo-sensore (0D)**, è l'installazione più semplice ed economica che si limita a fornire una misurazione puntuale dell'umidità del suolo, essendo formata da un solo sensore posizionato in un certo punto in profondità. In questo modo, non è quindi possibile monitorare l'umidità a profondità diverse.
- **multi-sensore**, che può essere a sua volta mono-dimensionale (1D), bi-dimensionale (2D) o tri-dimensionale (3D). L'installazione 1D permette

di raccogliere i dati di umidità a diverse profondità, ed è quella più comunemente sfruttata. Le installazioni 2D e 3D, invece, permettono di raccogliere le informazioni anche sugli altri assi, benché non siano spesso impiegate per via dell'alto numero di sensori richiesto. Nel caso 2D, infatti, più sensori vengono posizionati a parità di profondità, formando una matrice in grado di tenere traccia dell'intera fetta di terreno. Lo scenario 3D può essere visto come un ampliamento del precedente, in cui più matrici 2D vengono affiancate consentendo di monitorare direttamente una porzione cubica di suolo.

Qualità del servizio	Tipologia di sensore		
	UAV	Satellitare	In-situ
Flessibilità	Alta	Bassa	Bassa
Adattabilità	Alta	Bassa	Bassa
Costo	Basso	Alto	Basso
Consumo di tempo	Basso	Basso	Alto
Rischio	Basso	Medio	Basso
Accuratezza	Alta	Bassa	Moderata
Distribuzione	Facile	Difficile	Moderata
Fattibilità	Sì	No	Sì
Disponibilità	Sì	No	No
Operabilità	Facile	Complessa	Facile

Tabella 1.2: Confronto delle diverse tipologie di sensori in termini di qualità del servizio

1.4 Struttura del campo

Modellare in maniera formale la struttura del campo aiuta a comprenderne le dinamiche e di conseguenza a prendere decisioni più consapevoli. A tal fine, è possibile definire innanzitutto il concetto di *volume di suolo* come segue:

Definizione 1. *Dato un albero, il suo volume di suolo è un parallelepipedo di suolo che contiene la maggior parte delle sue radici. Il volume in questione è centrato nella posizione dell'albero.*

Tale volume è in genere soggetto a una forte variabilità orizzontale e verticale causata da molti fattori correlati, tra cui l'aspirazione irregolare delle radici, la copertura limitata del sistema di irrigazione e le differenze negli strati del suolo in termini di composizione ed esposizione agli agenti atmosferici.

Per quanto riguarda la gestione dell'irrigazione, è poi possibile definire il concetto di *griglia di sensori*:

Definizione 2. Una griglia di sensori $S = s^1, \dots, s^{|S|}$ è un layout n -dimensionale di $|S|$ sensori installati in un volume di suolo. Ciascun sensore s^i è definito da uno spostamento tridimensionale $(s^i.x_1, s^i.x_2, s^i.x_3)$ rispetto al centro del volume di suolo e da un valore di umidità del suolo $s^i.v$.

In base alla dimensionalità n scelta, la griglia assume le sembianze di una linea ($n = 1$), di un rettangolo ($n = 2$) o di un parallelepipedo ($n = 3$).

Installati i sensori che vanno a costituire la griglia appena definita, è quindi possibile monitorare l'andamento dell'umidità nel volume di suolo. Tuttavia, considerando che i sensori di per sé forniscono delle misurazioni puntuali dell'umidità, mentre invece questa varia in maniera continua nel volume, solitamente si introduce la nozione di *profilo di umidità*:

Definizione 3. Data una griglia di sensori n -dimensionale S , il profilo di umidità è una griglia n -dimensionale $P = \{p^1, \dots, p^{|P|}\}$ che approssima, in ogni p^i , l'umidità del suolo misurata da S . P è a grana fine rispetto a S poiché $|P| > |S|$.

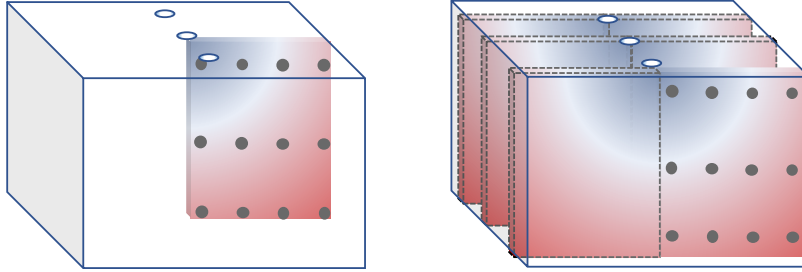


Figura 1.4: Assunzione di simmetria per una griglia 2D. Le parti tratteggiate replicano i valori misurati dai sensori

Partendo dalle misurazioni dei sensori, è quindi possibile stimare l'umidità del terreno dell'intero volume di suolo con una risoluzione a grana fine. Dal momento che il profilo fa riferimento alla regione coperta dai sensori, che è più piccola rispetto all'intero volume di suolo, tale stima è significativa solamente facendo un'assunzione di simmetria, secondo la quale le porzioni del volume di suolo che non sono coperte dalla griglia di sensori seguono lo stesso comportamento della zona monitorata. Il profilo delle porzioni non coperte dalla griglia di sensori può quindi essere ottenuto mediante rotazione o traslazione del profilo. La Figura 1.4 mostra l'effetto dell'applicazione dell'assunzione di

simmetria su griglie 2D, valida quando i gocciolatori (*dripper*) che forniscono l'irrigazione sono sufficientemente vicini da garantire un'umidità omogenea del terreno [19].

La Figura 1.5 illustra il layout complessivo di un campo utilizzando i concetti sopra riportati. Supponendo che i gocciolatori siano posizionati a distanza ravvicinata, ciò garantisce un'umidità omogenea lungo la riga, d'altra parte l'effetto dell'irrigazione cala tra una fila di piante e l'altra. Questo fa sì che il volume effettivamente irrigato (colorato in blu nella figura) sia solo una porzione del volume di suolo complessivo, che presenterà quindi delle regioni non coinvolte nell'irrigazione (colorate in arancione nella figura).

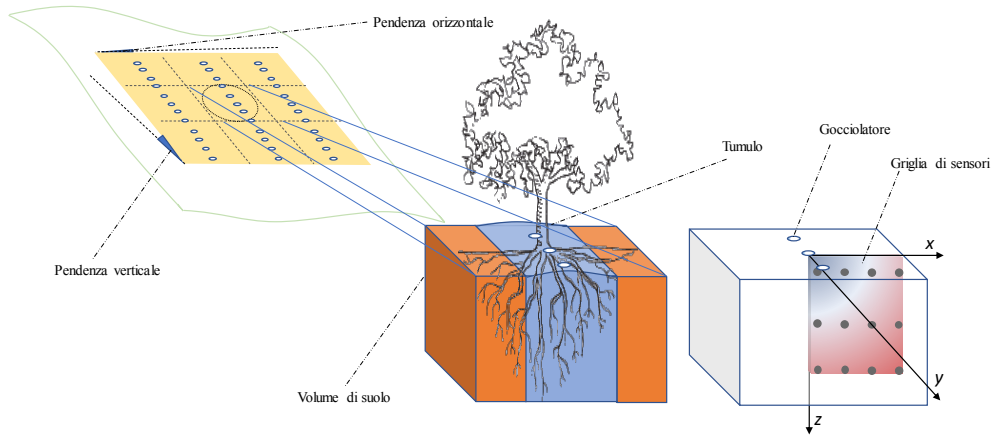


Figura 1.5: Modellazione del layout del campo

Il campo reale preso come riferimento per questo lavoro è una piantagione di kiwi situata nella località di Errano, in provincia di Ravenna. La situazione dettagliata è raffigurata in Figura 1.6: gli alberi di kiwi nel frutteto sono organizzati in file lunghe 170 metri e distanti 4.5 metri l'una dall'altra, mentre all'interno della fila la distanza tra le piante è pari a 2 metri. Il campo nel suo insieme presenta una pendenza dell'1% lungo l'asse x e del 2.5% lungo l'asse y . Ciascuna pianta, inoltre, è posizionata su un tumulo di terra per favorire il drenaggio, con pendenza del 20% sull'asse x e del 2.5% sull'asse y .

La tessitura del suolo è così composta: 30% di sabbia, 30% di limo e 40% di argilla ed è pertanto classificabile come terreno franco argilloso (*clay loam*).

Il sistema di irrigazione è costituito da un impianto a goccia a singola ala gocciolante, con una portata pari a 4 L/h, vale a dire che il singolo *dripper* è capace di erogare fino a 4 litri di acqua in un'ora.

Nel campo è installata una griglia *multi-sensore* bi-dimensionale (2D). Dal momento che i gocciolatori sono vicini tra di loro, il volume irrigato può essere considerato simmetrico lungo la riga, pertanto una griglia di questo

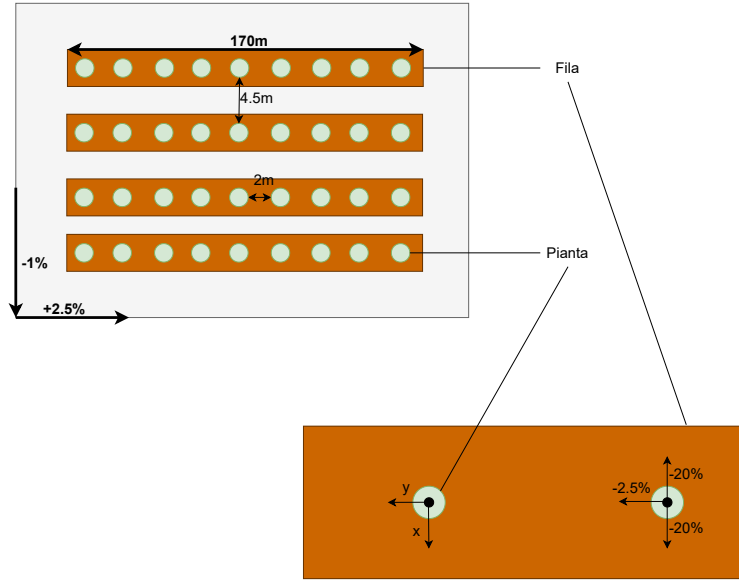


Figura 1.6: Layout del frutteto di kiwi situato a Errano (RA)

tipo è sufficiente a rappresentare l'intero volume di suolo. Il layout generale è raffigurato in Figura 1.7: la griglia è allineata lungo l'asse x del campo, con il primo sensore centrato in C_1 , subito al di sotto del *dripper* D_2 . La misurazione in C_i è in realtà frutto della combinazione di tre diversi sensori Y_i , B_i e R_i che assieme forniscono il valore di umidità in C_i . La Figura 1.8 mostra invece il dettaglio della disposizione dei vari sensori. Un totale di 12 sensori sono posizionati a diverse profondità (lungo l'asse z in figura), rispettivamente a 20, 40 e 60 centimetri nel sottosuolo. Ogni livello di profondità ospita quattro sensori (lungo l'asse x in figura), rispettivamente a 0, 25, 50 e 80 centimetri di distanza rispetto al gocciolatore.

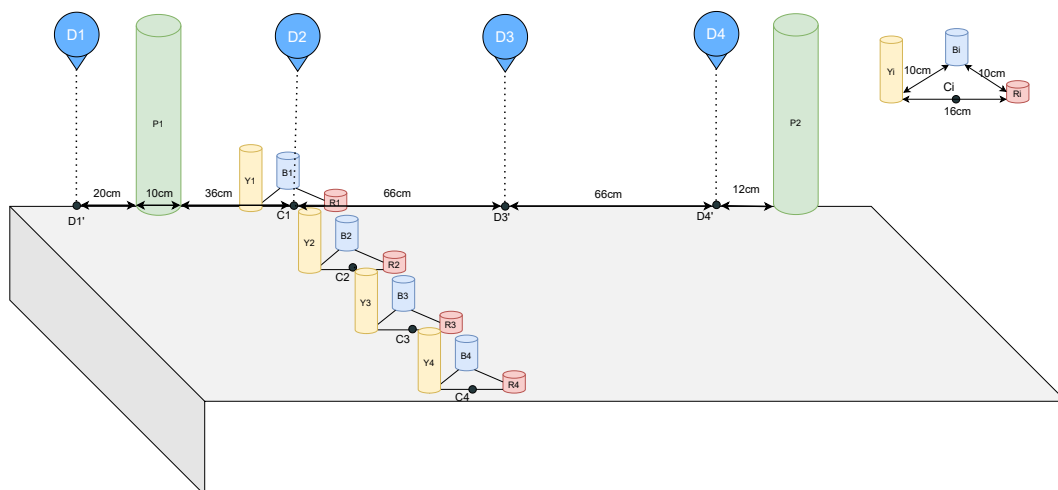


Figura 1.7: Layout del sistema di irrigazione e della griglia di sensori

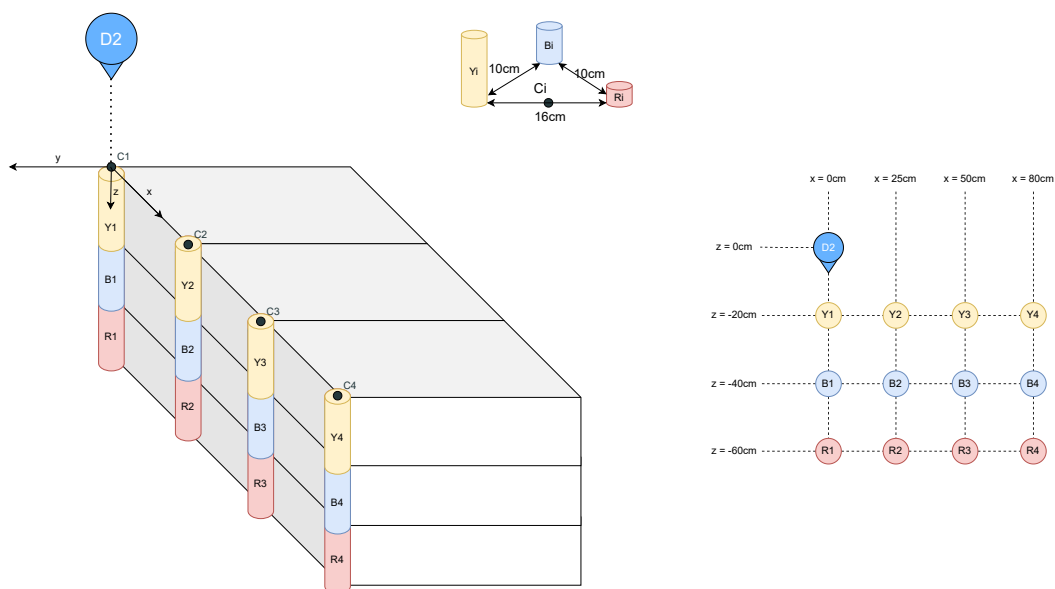


Figura 1.8: Layout dettagliato della griglia di sensori

Capitolo 2

Stato dell'arte

Nell'ambito dell'agricoltura di precisione, gli obiettivi che si possono perseguire e che hanno come oggetto l'umidità del terreno, possono essere riassunti nelle seguenti due categorie:

- **Simulazione:** stima la dinamica del terreno nel tempo con l'obiettivo di comprenderne il comportamento in circostanze specifiche. Scopo ultimo della simulazione non è tanto vincolarsi agli effettivi valori di umidità del suolo e ai dati meteorologici, bensì quello di testare diversi scenari al fine di comprendere il comportamento generale del suolo e creare quindi un modello. Combinando una descrizione grossolana dell'umidità del terreno con ipotesi statistiche e la conoscenza delle caratteristiche del suolo è possibile ottenere una rappresentazione a grana fine del terreno.
- **Previsione:** partendo dai valori attuali di umidità in un determinato campo, stima i suoi valori futuri per consentire all'utente di prendere le decisioni opportune (*e.g.*, irrigare).

Il lavoro qui presentato rientra quindi nella sfera della previsione. L'obiettivo è quello di andare oltre alla predizione dell'umidità del terreno di un singolo campo/tessitura, e di cercare di adattare il modello nel tempo, man mano che nuovi dati provenienti dai campi si rendono disponibili, in modo che sia in grado di operare anche su terreni diversi da quello iniziale.

In letteratura, i lavori riguardanti l'agricoltura di precisione sono molteplici e spaziano dal rilevamento di malattie e pesticidi al miglioramento della resa del raccolto, passando per la determinazione di un piano di *smart irrigation* capace di ridurre gli sprechi di acqua ma che sia allo stesso tempo efficace. Nonostante ciò, quelli che si focalizzano sulla previsione dell'umidità del suolo nei campi e che fanno uso di un'adeguata griglia di sensori per la raccolta dei dati in loco, in grado di supportare metodologie *data-driven* impiegando approcci di Machine Learning (ML), solo recentemente hanno iniziato a diffondersi.

Di seguito si presentano i lavori noti in letteratura che hanno avuto come oggetto la predizione dell'umidità del terreno in ambito agricolo, considerando prevalentemente quelli che hanno fatto uso di sensori *in-situ* per la raccolta dei dati direttamente dal campo.

Una prima distinzione può essere fatta in base alla tecnica specifica utilizzata:

- **modelli fisico-numerici:** si tratta di modelli analitici che *codificano* in maniera diretta tutte le regole fisiche che governano le dinamiche idrologiche. Essendo stati per molto tempo l'unico approccio disponibile, sono considerati ancora oggi strumenti solidi e ben accettati.
- **modelli a media mobile:** sfruttano dei tradizionali metodi matematici lineari per compiere deduzioni su serie storiche di dati. A partire da un piccolo campione di esempi, sono in grado di fare assunzioni sulla base di regole conosciute a priori, facendo leva sull'inferenza dai dati, sulla ricerca delle relazioni tra variabili e sul significato di tali relazioni.
- **modelli di Machine Learning:** sfruttano le moderne tecniche di apprendimento automatico supervisionato per imparare le dinamiche idrologiche da un'ampia collezione di esempi, ciascuno corredato dal rispettivo valore da predire. Rispetto ai modelli numerici, la loro applicazione in ambito di agricoltura di precisione è relativamente recente. Oltre a mostrare benefici in termini di efficacia, questi approcci possono contare su un basso consumo di risorse computazionali (una volta ottenuto il modello) e sulla rapidità con cui i risultati sono disponibili. A differenza dei modelli a media mobile, necessitano di una fase di addestramento mediante la quale acquisiscono la capacità di generalizzare e quindi di fornire predizioni accurate su dati mai visti prima. Una collezione separata di dati viene pertanto fornita in seguito per testare la bontà del modello creato.

2.1 Modelli fisico-numerici

Storicamente, la maggior parte dei lavori in letteratura ha affrontato il problema della modellazione delle dinamiche dell'umidità del suolo utilizzando approcci matematici. Tra questi, particolare rilievo hanno assunto i modelli numerici basati sulla fisica, capaci di sfruttare come predittore la descrizione quantitativa delle proprietà idro e termo-fisiche basata su parametri fisici fondamentali del suolo. A questo scopo, la suite HYDRUS [20, 21, 22] è stata ampiamente utilizzata per l'analisi e la modellazione dei flussi d'acqua, sia in una dimensione (HYDRUS-1D) [23] che in due o tre dimensioni (HYDRUS-2D/3D) [24]. Questo programma simula la dinamica del suolo utilizzando

diverse equazioni note, come l'equazione di Richards [25, 26] per il movimento dell'acqua in terreni insaturi e l'equazione di Penman-Monteith [27] che approssima l'evapotraspirazione netta dai dati meteorologici o dalle funzioni di pedotrasferimento. Un altro modello talvolta usato è AQUACROP [28]: sviluppato dalla FAO, simula le interazioni tra terreno, piante e atmosfera e la risposta del rendimento delle colture all'acqua. Bittelli *et al.* hanno proposto il modello Orchard3D-Lab, che estende le potenzialità di un precedente sistema sviluppato dagli stessi autori, ovvero CRITERIA [29, 30, 31]. Orchard3D-Lab è stato appositamente sviluppato per calcolare il bilancio idrico del suolo di un frutteto, tenendo conto della distribuzione non uniforme dell'apparato radicale, della crescita e dello sviluppo delle piante con un calcolo completo del bilancio idrico del suolo e delle previsioni meteorologiche a breve termine [19].

In relazione alla predizione di *soil moisture*, sono stati condotti dei lavori che hanno sfruttato questi modelli fisici. Shein *et al.* [32] hanno confrontato i valori di SM misurati a diverse profondità in una piantagione in Russia con quelli calcolati da HYDRUS-1D, trovando piccole differenze per le profondità minori (fino a 40 cm). Ultimamente è tendenza comune affidarsi prima di tutto ad una fase di tuning dei parametri del suolo e/o della pianta per poi procedere con la previsione vera e propria. Basak *et al.* [33] hanno quindi proposto un approccio basato su due modelli "ibridi" (Naive Accumulative Representation (NAR) e Additive Exponential Accumulative Representation (AEAR)) che combinano la visione *data-driven*, dal momento che i parametri vengono stimati dai dati, con quella basata sulla fisica, visto che sono progettati per assomigliare ai modelli idrologici relativi ai processi del suolo. Il modello AEAR, che è stato convalidato anche dagli studiosi del suolo, ha fornito previsioni migliori rispetto ai modelli esistenti, per orizzonti temporali da 10 a 24 ore. Gli autori di Orchard3D-Lab fanno lo stesso, ma rispetto agli altri modelli menzionati la parametrizzazione relativa alla pianta e al suolo avviene mediante l'utilizzo di più avanzate tecniche basate su modelli probabilistici e una procedura di assimilazione permette l'allineamento del modello durante la sua esecuzione. Queste caratteristiche hanno permesso di predire in maniera precisa l'umidità del terreno in un campo di kiwi su un orizzonte temporale fino a 7 giorni.

Gli approcci basati sulla fisica hanno il vantaggio di offrire una rappresentazione precisa e a grana fine dell'umidità del suolo, e a seguito di un tuning che sfrutta i dati reali raccolti sono in grado di simulare quello che effettivamente accade nel suolo tramite equazioni note in letteratura. Tuttavia, non catturano fenomeni come crepacciamenti ed isteresi e tendono a generare modelli *site-specific*, la cui fruibilità è limitata all'area in cui sono stati addestrati, o ad aree con le stesse caratteristiche fisiche. Al variare delle caratteristiche del terreno, è quindi necessario procedere con tuning progressivi.

Karandish *et al.* [34] hanno infine confrontato il modello HYDRUS-2D

con alcuni modelli di ML (*i.e.*, Multiple Linear Regression (MLR), Adaptive Neuro Fuzzy Inference System (ANFIS) e Support Vector Machine (SVM)) in un'indagine di due anni su un campo di mais per simulare serie temporali di dati SWC durante le stagioni di crescita del 2011 e 2012. HYDRUS-2D si è classificato al primo posto per la stima di SWC, essendo più affidabile rispetto ai modelli di ML. Questi ultimi, potrebbero però essere favorevolmente utilizzati soprattutto quando sono disponibili pochi dati.

2.2 Modelli a media mobile

Una piccola parte di lavori si è concentrata sull'applicazione di alcuni modelli autoregressivi a media mobile (ARMA) finalizzati alla predizione di valori futuri, primo fra tutti il modello AutoRegressive Integrated Moving Average (ARIMA). Si tratta di un modello statistico tradizionalmente usato per la predizione di dati di serie temporali lineari. Come suggerisce il nome, questo approccio è in grado di predire una particolare variabile considerando i suoi valori passati differenziati, calcolando l'errore di regressione come una combinazione lineare di termini di errore verificatisi contemporaneamente e in vari momenti nel passato. L'efficiacia di questo tipo di modelli si manifesta soprattutto con serie temporali non stazionarie, caratterizzate dalla presenza di trend o di stagionalità.

El Mezouari *et al.* [35] hanno applicato diversi algoritmi, tra cui ARIMA, per predire l'umidità del suolo a 5 centimetri di profondità al fine di aiutare nella programmazione dell'irrigazione. Le serie temporali utilizzate hanno fatto riferimento ai dati reali di umidità e temperatura del suolo raccolti con cadenza oraria nell'arco di 5 anni (2010-2014) negli Stati Uniti, a cinque profondità diverse (5, 20, 35, 50, 75 cm). I risultati hanno dimostrato che ARIMA è riuscito a fornire prestazioni allineabili agli altri algoritmi di ML utilizzati nel lavoro (Random Forest e XGBoost), ma che ha superato le prestazioni degli altri due nel caso di previsione eseguita usando solo l'umidità del suolo.

Anche Khanna *et al.* [36] si sono avvalsi di ARIMA per la previsione dell'umidità del suolo in un raccolto di grano a una profondità di 10, 45 e 80 centimetri. Rispetto alla soluzione basata su Long Short-Term Memory (LSTM), il loro modello ha ottenuto risultati migliori per le previsioni a breve termine, mentre quest'ultimo ha offerto prestazioni migliori su orizzonti temporali più lunghi.

2.3 Modelli di Machine Learning

In tempi recenti, i lavori in materia di predizione dell'umidità del suolo in ambito agricolo hanno virato sempre di più verso l'adozione di approcci

data-driven, grazie allo slancio dato dai diversi modelli di ML esistenti. Di seguito si illustrano i principali lavori distinguendo tra quelli che hanno fatto uso di metodi considerati più classici e quelli che hanno utilizzato approcci più avanzati.

2.3.1 Approcci classici

Partendo dagli approcci più classici, diverse pubblicazioni hanno impiegato approcci basati su SVM, un metodo di ML originariamente pensato per task di classificazione. Dato un insieme di esempi di addestramento, ciascuno contrassegnato come appartenente a una tra due categorie, SVM trasforma ciascun esempio in un punto nello spazio in modo da massimizzare l'ampiezza del divario tra le due categorie. I nuovi esempi forniti al modello vengono quindi mappati nello stesso spazio e fatti ricadere nella categoria relativa al lato del divario in cui si trovano. L'algoritmo può anche eseguire in modo efficiente una classificazione non lineare utilizzando il cosiddetto *kernel trick*, mappando implicitamente i propri input in spazi di feature ad elevata dimensionalità. Support Vector Regression (SVR) è la versione modificata di SVM, in cui la variabile dipendente è numerica anziché categorica. È una tecnica non parametrica e consente la creazione di modelli non lineari utilizzando le *kernel function*.

Gill *et al.* [37] sfruttano un modello basato su SVM, che prende in input i dati meteorologici e di SM al tempo $t - 1$ e t per predire l'umidità del terreno al tempo $t + 4$ o $t + 7$ (giorni). Hong *et al.* [38] oltre a SVM sfruttano la tecnica di Relevance Vector Machine (RVM), in grado di fornire classificazioni probabilistiche per predire la SM di 2 settimane avanti grazie ad una propria infrastruttura IoT avanzata che tiene conto delle condizioni meteorologiche. Per questioni di efficienza, la frequenza di campionamento è differente da sensore a sensore. Ad esempio, il campionamento a profondità minori sarà più frequente perché qui si è visto che la SM varia più velocemente rispetto a profondità maggiori. Al contrario, nei giorni in cui non piove, il campionamento può rallentare poiché la SM cala lentamente. Infine, Goap *et al.* [39] utilizzano SVR e il clustering k-means per predire la SM a 6 giorni di distanza per agganciarci poi un modulo per la gestione dell'irrigazione automatica, basato su soglie di limite inferiore e superiore e che tiene anche conto delle previsioni meteo future.

Un altro approccio sicuramente molto sfruttato è quello basato su reti neurali *feed-forward* (FFNN), diffusamente impiegate per risolvere problemi di classificazione e regressione. Il loro funzionamento si ispira al modello biologico della corteccia cerebrale dell'uomo, formata da una complessa rete di neuroni (*i.e.*, le cellule del tessuto nervoso) che comunicano tra di loro per mezzo delle sinapsi. Sebbene siano ancora oggetto di approfondite ricerche, dalle evidenze

attuali risulta che i neuroni si dispongano in livelli successivi che abilitano operazioni complesse a partire dalle singole unità, di per sé elementari. Sulla base di ciò, e a seconda della complessità del problema, le reti artificiali possono avere, oltre ai livelli di input e di output, nessuno o più hidden layer. Le connessioni tra i neuroni della rete, come suggerisce il nome della stessa, si muovono esclusivamente in avanti, quindi i neuroni di un livello possono essere collegati solo ai neuroni del livello successivo. Per addestrare una rete ad eseguire un particolare problema, si fa tipicamente affidamento all'algoritmo di retropropagazione dell'errore [40], attraverso il quale i pesi associati ad ogni neurone vengono progressivamente aggiornati e migliorati in modo che la predizione si avvicini a quella desiderata e che deriva dai dati offerti in input alla rete.

Dubois *et al.* [41] hanno sfruttato le FFNN nella predizione della SM a 7 giorni di distanza in una piantagione di patate nel nord della Francia. Impiegando una griglia con sensori low-cost a 3 diverse profondità, per fronteggiarne la limitata accuratezza sono stati installati tre punti di misurazione per ciascuna profondità, di cui poi è stata considerata la media. Infine, tramite tecniche di *feature selection* sono state selezionate le variabili migliori per la predizione.

Liu *et al.* [42] si sono invece affidati a tecniche di Extreme Learning Machine (ELM), che consistono in delle semplici reti FFNN con un singolo (o comunque pochi) hidden layer, per le quali non è richiesto il tuning dei parametri degli hidden node, che possono essere impostati in maniera random e mai aggiornati. I risultati migliori sono stati ottenuti considerando le variabili meteo e i valori passati di SM, tuttavia le predizioni si limitano in questo caso all'ora successiva.

Il lavoro di Kang *et al.* [43] sfrutta le FFNN per realizzare un sistema di supporto alle decisioni per regolare il piano di irrigazione a goccia in un vigneto per la settimana successiva. Il sistema è composto da un primo modulo che predice l'umidità del terreno nella settimana a venire, usando come input l'umidità iniziale del suolo, le variabili meteorologiche, il coefficiente colturale e la quantità di irrigazione. Il risultato funge poi da input per un sistema che aiuta a risparmiare acqua per l'irrigazione durante alcune fasi di crescita delle vigne e allo stesso tempo a raggiungere l'obiettivo produttivo di una resa equilibrata e di una qualità ottimizzata dei frutti.

Risheh *et al.* [44] hanno adottato un approccio di *transfer learning* addestrando due modelli di FFNN su due suoli con caratteristiche differenti, dopodiché hanno usato la rete addestrata su un suolo per addestrare una rete sull'altro suolo. Sono stati condotti due esperimenti che hanno evidenziato come il modello pre-addestrato converga più velocemente, performi meglio nelle epoche iniziali e raggiunga prima il punto di minimo locale rispetto al modello addestrato da zero. Un tale approccio consente da una parte un minore utilizzo di dati e dall'altra di adattarsi in modo flessibile all'aggiunta di nuovi sensori

nel tempo.

In Tabella 2.1 vengono catalogati i lavori appena presentati in base all'algoritmo utilizzato, all'orizzonte di output della predizione, al tipo e al layout dei sensori e alla quantità e provenienza dei dati.

Paper	Metodo	Orizzonte di output	Tipo di sensore	Layout	Periodo dei dati	Provenienza dei dati
[37]	SVM	4, 7 giorni	remoto + <i>in-situ</i>	0D	18/06/1997 - 16/07/1997	LWREW, Oklahoma
[38]	SVM, RVM	2 settimane	volumetrico (5, 10 cm)	1D	01/01/2004 - 31/12/2012	Illinois Climate Network + propria infrastruttura IoT
[39]	SVR, k-means	6 giorni	volumetrico	0D	3 settimane	propria infrastruttura IoT
[41]	FFNN	7 giorni	potenziale (20, 30, 40 cm)	1D	3 anni	propria infrastruttura IoT low-cost in una coltivazione di patate nel nord della Francia
[42]	ELM	15 - 60 minuti	volumetrico (20, 40, 60 cm)	1D	01/01/2011 - 02/06/2013	propria infrastruttura IoT in un meleto in Australia
[43]	FFNN	1 - 7 giorni	volumetrico (30, 60, 90 cm)	1D	2017 - 2021	vigneto negli Stati Uniti
[44]	FFNN (+ TL)	6 - 9 minuti	volumetrico	1D	50 giorni	propria infrastruttura IoT

Tabella 2.1: Riassunto degli approcci classici di ML

2.3.2 Approcci legati al Deep Learning

Diversi tentativi di successo sono stati portati avanti anche utilizzando approcci legati al Deep Learning, un insieme di tecniche basate su reti neurali organizzate in più livelli, in grado di creare una gerarchia di rappresentazioni dei dati sempre più astratte e complesse. Le tecniche utilizzate sono incentrate prevalentemente su reti LSTM, un particolare tipo di reti ricorrenti introdotte allo scopo di risolvere il problema della scomparsa del gradiente che affligge le Recurrent Neural Network (RNN) e che emerge quando si lavora con lunghe sequenze di dati [45]. Questa famiglia di reti sono caratterizzate da un flusso bi-direzionale dell'informazione tale per cui l'output di un certo neurone di un

livello può essere direttamente collegato ad altri neuroni dello stesso livello, ma anche a neuroni del livello precedente. Questa caratteristica le dota di un peculiare "effetto memoria" a breve termine e quindi sono particolarmente adatte ai problemi di previsione di serie temporali, per i quali la semantica di ordinamento degli eventi è significativa. Le reti LSTM riescono a migliorare il processo di memorizzazione, capendo quali informazioni possono essere "dimenticate", e quindi tralasciate, e quali invece sono importanti per predire gli step futuri.

Gao *et al.* [46] hanno implementato una rete LSTM a 5 livelli per predire la SM del giorno successivo in una piantagione di litchi in Cina. Adeyemi *et al.* [47] hanno in aggiunta costruito un modulo per la gestione dell'irrigazione, formato da un sistema di *irrigation planner*, che determina la distribuzione spazio-temporale dell'acqua che serve sul campo, e da un sistema di *irrigation scheduler*, che decide di irrigare il campo nel caso in cui la SM predetta ricada all'interno di un certo intervallo. Kashyap *et al.* [48] hanno confrontato il sistema di irrigazione predittivo realizzato con lo scheduling rule-based di AQUACROP. Il lavoro di Jiménez *et al.* [49], invece, sfrutta le LSTM per predire direttamente la quantità di acqua raccomandata per l'irrigazione dell'ora successiva.

In un recente lavoro, Datta *et al.* [50] propongono un approccio multihead LSTM in grado di superare il principale limite dei modelli basati su LSTM, ovvero il fatto che l'accuratezza nella predizione della SM decresce sensibilmente all'aumentare dell'orizzonte temporale di output. Nello specifico, si tratta di una tecnica di *ensemble* che utilizza quattro modelli diversi di LSTM per ottenere la predizione finale di SM.

Una serie di lavori si sono avvalsi invece di approcci più sofisticati, andando a combinare assieme modelli differenti. Yu *et al.* [51] hanno creato un modello composto da una Residual Neural Network (ResNet)¹, che prende in input i dati di SWC, e da una Bidirectional LSTM (BiLSTM)², che prende in input i dati meteo. La combinazione di questi dati porta alla definizione di una matrice 2D, che può essere interpretata come un'immagine bi-dimensionale in scala di grigi a canale singolo, da qui la possibilità di usare una rete convoluzionale (ResNet) per estrarre le feature dalla "immagine" e una rete ricorrente (BiLSTM) per estrarre le feature serie-temporali. I risultati hanno evidenziato una migliore

¹La ResNet è un modello di Deep Learning che sfrutta la tecnica di *residual learning* che prevede l'aggiunta di connessioni che trasferiscono i dati da un layer all'altro senza che questi subiscano alcuna trasformazione. Così facendo, viene a crearsi un "residuo" che rappresenta la differenza tra l'input e l'output del layer e che aiuta a migliorare le performance del modello.

²La BiLSTM è una rete che aggiunge un ulteriore layer LSTM che inverte la direzione del flusso di informazioni.

Paper	Metodo	Orizzonte di output	Tipo di sensore	Layout	Periodo dei dati	Provenienza dei dati
[46]	Deep-LSTM	1 giorno	volumetrico (40 cm)	0D	09/2019 - 09/2021	propria infrastruttura IoT in un campo di litchi in Cina
[47]	FFNN, LSTM	1 giorno	cosmic-ray + volumetrico (fino a 60 m)	0D	2014 - 2017	dataset COSMOS
[48]	LSTM	1 giorno	cosmic-ray + volumetrico (fino a 20 m)	0D	01/06/2017 - 31/12/2018	dataset India-WRIS per il training + propria infrastruttura IoT per il testing
[49]	LSTM	1 ora	potenziale (15, 30, 60 cm)	1D	2017 - 2019	coltivazione di mais in Alabama
[50]	multihead LSTM (tecnica ensemble)	1 giorno, 1 settimana, 1 mese	volumetrico (10, 30 cm)	2D	2015 - 2021	area forestale del Monte San Antonio (California)
[51]	ResBiLSTM	6 giorni	volumetrico (10, 20, 30, 40, 50 cm)	1D	2016 - 2018	dataset China Meteorological Data Network (piantagioni di mais)
[52]	EDR-LSTM	1, 3, 5, 7, 10 giorni	volumetrico (5 cm)	0D	1996 - 2014	dataset FLUXNET2015 (subset tie1 & tie2)
[53]	PCA-LSTM	1 - 24 ore	volumetrico	0D	2015 - 2018	piantagione di tabacco in Cina
[54]	attention-aware-LSTM	1, 7 giorni	volumetrico (5 cm)	0D	periodi diversi (fino a 10 anni)	dataset FLUXNET2015 (subset tie1 & tie2)
[55]	LSTM-Transformer	1 giorno	volumetrico (10, 40, 100, 200 cm)	1D	2012 - 2022	prateria di Xilin Gol in Mongolia

Tabella 2.2: Riassunto degli approcci di Deep Learning

accuratezza per le profondità maggiori e un'accuratezza peggiore all'aumentare dell'orizzonte di predizione. Partendo dalla considerazione che nella predizione della SM tramite LSTM i dati serie-temporali tra lo step di input e lo step predetto vengono ignorati, Li *et al.* [52] hanno sviluppato un modello composto da un layer LSTM di tipo encoder-decoder e da un successivo layer LSTM completamente connesso. Il primo layer è in grado di considerare le serie temporali intermedie, mentre il secondo prende in input sia l'output generato dal layer precedente che i dati di SM originali. Tale meccanismo, ispirandosi al *residual learning* [56], consente di migliorare l'accuratezza predittiva del layer encoder-decoder e di ridurre l'overfitting.

Nel contesto della predizione di SM in una piantagione di tabacco in Cina, Jiang *et al.* [53] si sono avvalsi della tecnica di Principal Component Analysis

(PCA) per ridurre la dimensionalità degli input senza per questo eliminare le varie correlazioni tra le variabili. Li *et al.* [54] hanno fatto qualcosa di simile utilizzando meccanismi di *attention-aware*³ per individuare le cinque migliori variabili nella predizione dell'umidità del suolo e trovare la correlazione tra ciascun predittore e la SM. Zhou *et al.* [55] propongono un approccio che combina i meccanismi di *attention* basati su modelli Transformer [57], adatti a predire serie temporali multivariate, con le reti LSTM per aggiungere funzionalità di "memoria" altrimenti assenti nei primi modelli.

La Tabella 2.2 schematizza i lavori appena citati che hanno sfruttato modelli avanzati di ML, evidenziando per ognuno le informazioni rilevanti quali il tipo di sensori utilizzati e il layout della griglia, la quantità e l'origine dei dati e l'orizzonte di output della predizione.

2.4 Automated Machine Learning

Il veloce avanzamento avvenuto in tempi recenti nell'adozione delle tecniche di Machine Learning e di Deep Learning ha portato ad un aumento esponenziale dei possibili algoritmi che si possono impiegare nella risoluzione dei problemi. Non esistendo un algoritmo che a priori sia indicato per ogni scenario o tipologia di dataset a disposizione, indirizzare la scelta verso una soluzione che si adatti al caso specifico risulta essere un obiettivo spesso irrealizzabile. Inoltre, nella scelta dei diversi iper-parametri che compongono un particolare algoritmo, è opportuno seguire un approccio quanto più metodologico e scientifico, anziché procedere per tentativi. Per questi motivi, è nato il campo dell'Automated Machine Learning (AutoML), un processo in grado di automatizzare l'applicazione del Machine Learning a problemi reali. Uno strumento di AutoML è quindi in grado di esplorare in maniera efficiente grandi spazi di ricerca, altrimenti non esplorabili nella loro interezza viste le dimensioni. L'individuazione della migliore combinazione di iper-parametri è un problema ben noto in letteratura, che prende il nome di Hyper-Parameter Optimization (HPO) [58], che può essere formulato nel modo seguente:

³Il meccanismo di *attention* è una tecnica applicata alle reti neurali che permette di associare un peso, e quindi una importanza, diverso per ciascuna parte dell'input, in base all'effettiva rilevanza per il problema che si sta affrontando.

dato

- un dataset D suddiviso in D_{train} e D_{val}
- un algoritmo A con i relativi iper-parametri $\lambda_1, \dots, \lambda_n$
- gli spazi (domini) di ricerca $\Lambda_1, \dots, \Lambda_n$ associati ad ogni iper-parametro
- lo spazio di ricerca complessivo $\Lambda = \Lambda_1 \times \dots \times \Lambda_n$, dato dal prodotto cartesiano dei precedenti domini definiti
- una configurazione $\lambda \in \Lambda$, data da un campionamento dello spazio
- l'istanza A_λ dell'algoritmo A con assegnati gli specifici valori degli iper-parametri λ
- una funzione di loss $\mathcal{L}$, che quantifica l'errore di valutazione⁴ $\mathcal{L}(A_\lambda, D_{train}, D_{val})$

si vuole trovare

$$\lambda^* \in \operatorname{argmin}_{\lambda \in \Lambda} \frac{1}{k} \sum_{i=1}^k \mathcal{L}(A_\lambda, D_{train}, D_{val}) \quad (2.1)$$

che è la configurazione di iper-parametri che minimizza la funzione di loss, chiamata anche funzione obiettivo.

L'approccio risolutivo comunemente utilizzato è quello basato sull'ottimizzazione bayesiana, ma è anche disponibile una serie di metodi *model-free*, tra cui si possono annoverare varie tecniche come *grid search*, *random search*, le euristiche, gli algoritmi genetici e le ottimizzazioni *multi-resolution* [59].

Data la funzione obiettivo da minimizzare e l'insieme dei dati osservati, rappresentati dagli iper-parametri testati in precedenza, l'ottimizzazione bayesiana può essere vista come un algoritmo iterativo composto da un modello surrogato e da una funzione di acquisizione. Il modello surrogato è costituito a sua volta dalla funzione prevista, ovvero una stima della funzione obiettivo, e dall'intervallo di confidenza, che indica la possibile varianza. Lo scopo della funzione di acquisizione è invece quello di suggerire la prossima configurazione da visitare, cercando di calibrare l'esplorazione di nuovi punti nello spazio con lo sfruttamento di quelli già visitati e che si sanno essere promettenti (Figura 2.1).

Una possibile formalizzazione dell'ottimizzazione bayesiana è la Sequential Model-Based Optimization (SMBO), che può essere riassunta come segue:

⁴L'errore di valutazione indica l'errore commesso dagli iper-parametri dell'algoritmo allenato su D_{train} e validato su D_{val} , il quale non viene mai visto durante l'addestramento.

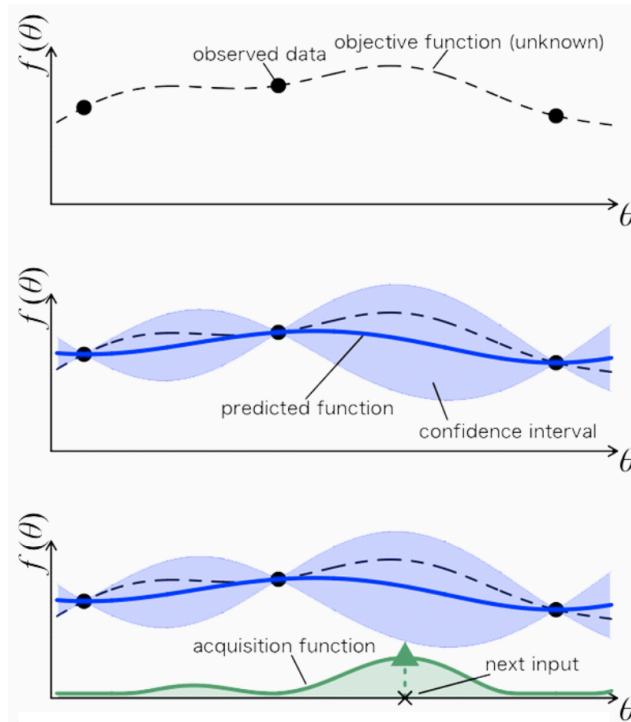


Figura 2.1: Ottimizzazione bayesiana, figura da [2]

1. Valutazione di alcune configurazioni casuali di iper-parametri.
2. Costruzione di un modello surrogato.
3. Uso del modello e della funzione di acquisizione per trovare la prossima configurazione di iper-parametri da valutare.
4. Valutazione della configurazione di iper-parametri.
5. Aggiornamento del modello surrogato incorporando i nuovi risultati.
6. Ripetizione degli step 3-5 fino al superamento del budget.

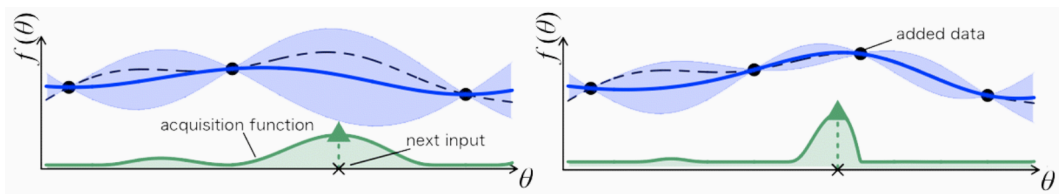


Figura 2.2: Step di SMBO, figura da [2]

La figura Figura 2.2 esemplifica la dinamica dell'ottimizzazione appena descritta. Sono possibili più implementazioni, che si differenziano in base a come SMBO costruisce il modello surrogato. Un'opzione prevede di usare Random Forest (SMAC) [60], in modo tale che la funzione prevista venga modellata come la media sulle previsioni dei suoi singoli alberi per λ , mentre l'intervallo di confidenza come la varianza sulle previsioni degli stessi. Altre implementazioni possono essere Gaussian Process (GP) [61] e Tree-Parzen Estimator (TPE) [62].

Gli approcci bayesiani hanno il vantaggio di riuscire a convergere con poco budget e di fornire informazioni dettagliate. D'altra parte, nel caso di spazi di iper-parametri molto grandi, soffrono del problema di *cold-start* che rallenta la convergenza iniziale, oltre al fatto che non forniscono alcuna ottimizzazione per ridurre i costi di valutazione, considerando che la costruzione del modello surrogato è un processo oneroso. Questo problema può essere mitigato con FLAML [63, 64], uno strumento di AutoML sviluppato da Microsoft e rilasciato sotto forma di libreria Python che è specializzato nella ricerca automatica e a basso costo computazionale di modelli accurati di Machine Learning. Il contributo che apporta rispetto ad altre soluzioni concorrenti è quello di condurre la ricerca del miglior modello considerando sia il costo di configurazione, ossia il costo computazionale associato all'addestramento e alla valutazione dell'errore del modello, sia l'errore di configurazione, ovvero l'errore del modello valutato. Saper bilanciare questi due elementi è importante, dal momento che una soluzione di AutoML orientata unicamente alla ricerca di un basso errore di configurazione può portare all'esplorazione di configurazioni inutilmente costose, viceversa, una soluzione ottimizzata per la ricerca di un basso costo di configurazione consentirebbe di esplorare configurazioni economiche ma errate. Entrambe le componenti sono condizionate da diversi fattori, quali la scelta dell'algoritmo e dei relativi iper-parametri, la quantità dei dati di addestramento e la strategia di ricampionamento. FLAML riesce a gestire la dipendenza tra queste variabili, creando un sistema economico, robusto ed efficiente in grado di ottimizzare il tradeoff costo-errore, partendo da configurazioni economiche e modelli imprecisi per virare gradualmente verso configurazioni costose e modelli accurati.

Il processo di ottimizzazione può essere formalizzato come segue: dati L algoritmi, ciascuno con il proprio spazio di iper-parametri H_l , ad ogni esplorazione viene scelto un algoritmo l , i valori dei suoi iper-parametri $\mathbf{h} \in H_l$, la dimensione del campione dei dati di addestramento s e la strategia di ricampionamento $r \in \{cv, holdout\}$ ⁵. Una configurazione di apprendimento è

⁵Rispettivamente *k-fold cross-validation* e *hold-out*, sono due tecniche di suddivisione dei dati finalizzate alla convalida della bontà di un certo modello di predizione.

quindi definita come la tupla $\chi = (l, \mathbf{h}, s, r)$ che se esplorata genera in output un errore di validazione $\tilde{\epsilon}(\chi)$ e un modello $M(\chi)$.

FLAML offre due diversi algoritmi di ottimizzazione degli iper-parametri:

- **Frugal Optimization for Cost-related Hyperparameters (CFO):** utilizza la ricerca diretta randomizzata FLOW², un metodo di tipo iterativo in grado di muoversi rapidamente verso regioni *low-loss* dello spazio di ricerca e di evitare l'esplorazione di regioni ad alto costo se non strettamente necessario. Ha il vantaggio di implementare una ricerca locale con un certo tasso di convergenza e un costo limitato. La sua applicazione è consigliata negli scenari in cui prima dell'ottimizzazione è noto un punto iniziale a basso costo, che viene sfruttato come partenza per poi spostarsi al bisogno verso regioni a più alto costo.
- **BlendSearch:** combina la ricerca locale con quella globale, sfruttando la frugalità di CFO da un lato e la capacità di esplorazione dei metodi di ricerca globale (come l'ottimizzazione bayesiana) dall'altro. Ciò gli permette di capire in modo intelligente ed economico quali regioni dello spazio visitare, anche se questo è più grande del necessario. Il vantaggio da parte dell'utente è quindi considerevole, visto che può specificare uno spazio di ricerca più ampio senza doverlo restringere manualmente a priori. Come CFO, richiede in input un punto iniziale a basso costo (se esiste) dal quale partire, senza però dover attendere che la ricerca locale converga completamente prima di provare altri punti, che vengono suggeriti dal metodo di ricerca globale. La sua applicazione è da preferire se lo spazio di ricerca è così complesso tale per cui CFO rimarrebbe bloccato in un punto di ottimo locale.

Capitolo 3

Orchard3D-Lab

Orchard3D-Lab è un sistema integrato che unisce un modello numerico tridimensionale ad una procedura di assimilazione basata su griglie di sensori. Per avere una corretta previsione dell'irrigazione è infatti necessario fare affidamento a un modello di bilancio idrico che sia in grado di calcolare il contenuto volumetrico di acqua nel suolo (SWC) o il potenziale idrico (SWP) nella zona radicale. In base ai valori che si ottengono da questo modello è quindi possibile procedere con una stima del fabbisogno idrico del suolo, utile a mantenere la pianta coltivata in condizioni ottimali per la crescita e la maturazione del frutto. Sfruttare un modello di bilancio idrico presuppone la misurazione di molteplici processi fisici, tra cui il deflusso superficiale, la percolazione profonda nelle falde acquifere, la risalita capillare dalle stesse, il flusso laterale sotterraneo, l'evaporazione del suolo e la traspirazione delle piante [65]. Dato che misurare tutte queste variabili non è normalmente fattibile per la maggior parte delle aziende agricole, un *sistema di simulazione*, quale è Orchard3D-Lab, è essenziale per poterle stimare, utilizzando come input i dati meteorologici e di umidità del terreno, che possono essere raccolti con sensori low-cost installati nel profilo di suolo, oppure, nel caso delle variabili meteo, sfruttando i dati resi disponibili dagli enti pubblici o da aziende private.

In Figura 3.1 è raffigurato l'approccio di simulazione adottato e i diversi moduli che compongono il sistema. Ad alto livello, è possibile distinguere due diverse fasi:

- **fase offline:** si occupa di determinare i parametri ottimali del modello numerico. Il modulo di tuning dei parametri ricerca i parametri migliori andando a confrontare i dati simulati dal modello numerico (CRITERIA-3D) con i dati storici provenienti dai sensori posizionati sul campo, che devono quindi coprire lo stesso periodo dello scenario di simulazione.

- **fase online:** una volta trovati i parametri ottimali, il sistema procede con la simulazione dei dati di umidità del terreno, sfruttando un modulo di assimilazione dei dati utile a inizializzare il modello di simulazione allo stato reale del campo.

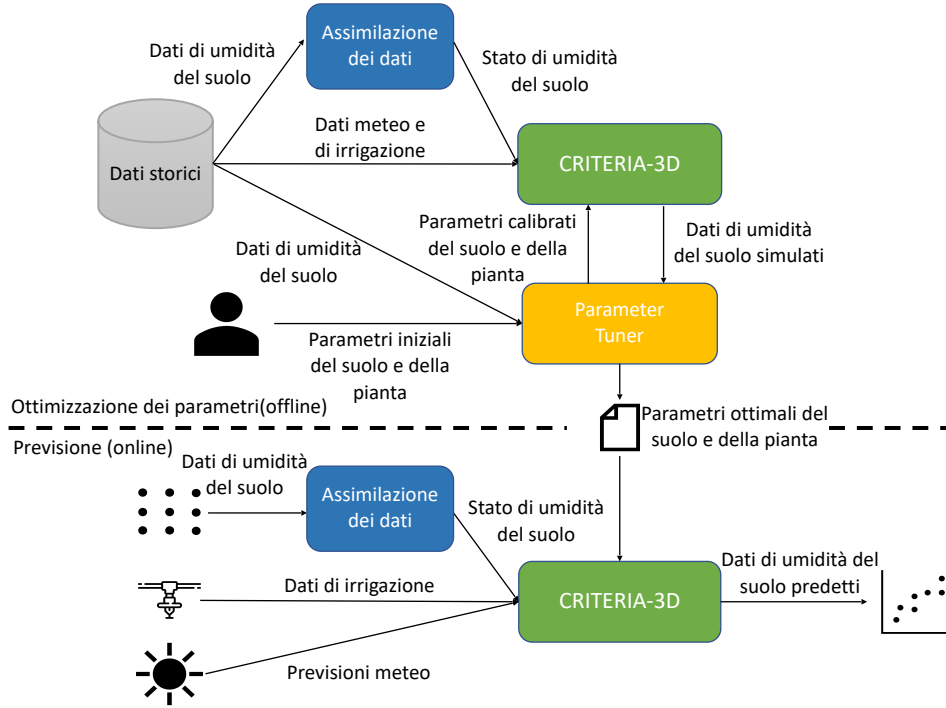


Figura 3.1: Panoramica e principali componenti di Orchard3D-Lab

Nella Tabella 3.1 vengono elencati i parametri di simulazione richiesti: quelli evidenziati in grassetto sono automaticamente inferiti dalla fase di tuning (fase offline), mentre i restanti devono essere forniti dall'utente.

Nelle successive sezioni vengono presentati nel dettaglio i diversi moduli che compongono Orchard3D-Lab e che ne determinano il funzionamento: il modello di simulazione (sezione 3.1), il modulo per l'assimilazione dei dati (sezione 3.2) e il modulo per il tuning dei parametri (sezione 3.3).

3.1 Modello di simulazione

Orchard3D-Lab si avvale del modello di simulazione CRITERIA-3D ([29, 30, 31]), un simulatore numerico che modella i flussi superficiali e sotto-superficiali di acqua e di calore nel suolo in un dominio tridimensionale e che richiede in input i dati meteorologici orari.

Tipo	Parametro	Descrizione
Campo	Posizione	Latitudine, longitudine e altitudine del campo [m]
	Pendenze	Pendenze del campo simulato [$m \cdot m^{-1}$]
	Tumulo	Pendenza [$m \cdot m^{-1}$] e larghezza [m] del tumulo
	Falda freatica $\in [1.5...5]$	Profondità di una falda [m]
Strato di suolo	Spessore	Spessore dello strato di suolo [m]
	Tessitura	Quantità di sabbia, limo e argilla [%]
	$K_{sat} \in [10^{-7}...10^{-3}]$	Conducibilità idraulica satura [$m \cdot s^{-1}$]
	$\theta_s \in [0.2...0.7]$	Contenuto di acqua satura [–]
	$\theta_r \in [0.02...0.7]$	Contenuto di acqua residua [–]
	$\alpha \in [0.5...5]$	Aspirazione con entrata d'aria inversa di Van Genuchten [m^{-1}]
	$n \in [1...2]$	Distribuzione dimensionale dei pori di Van Genuchten [–]
	$he \in [0...15]$	Potenziale matriciale dell'aria di Van Genuchten/Campbell [mH_2O]
Pianta	$b \in [1...9]$	Parametro di ritenzione empirica di Campbell
	Posizione	Posizione delle piante [m]
	Raggio $\in [1...3]$	Raggio della chioma della pianta [m]
	Leaf Area Index (LAI) $\in [3...5]$	Superficie fogliare verde per unità di superficie di terreno
	Profondità delle radici $\in [0.5...1.2]$	Profondità iniziale e massima delle radici [m]
	Deformazioni delle radici $\in [0...2]$	Fattori di deformazione delle radici sugli assi x e z
	$K_{cmax} \in [0.8...4]$	Coefficiente culturale massimo
	$f_{RAW} \in [0...1]$	Frazione di acqua del suolo disponibile per la pianta
Irrigazione	Tipo	Tipologia di irrigazione
	Posizione	Posizione degli irrigatori
Modello	Geometria	Lunghezza $\times$ larghezza $\times$ profondità del campo simulato [m]
	Dimensione della cella	Dimensione di una singola unità di simulazione di parallelepipedi [m]
	Curva di ritenzione $\in \{Campbell, VanGenuchten\}$	Modello di ritenzione idrica
	Media di conducibilità idraulica $\in \{log, harm, geom\}$	Metodo della media della conducibilità dell'acqua
	Rapporto di conducibilità idraulica $\in [0...1]$	Rapporto di conducibilità orizzontale/verticale dell'acqua

Tabella 3.1: Parametri richiesti per la simulazione numerica

Per calcolare il flusso d'acqua, il modello di simulazione CRITERIA-3D risolve l'equazione di Richards. Lo stato dell'umidità del suolo viene memorizzato come una matrice 3D, in cui ogni cella rappresenta una porzione del volume di suolo assieme alla sua SM simulata. Per ogni cella viene quindi calcolata la variazione del potenziale idrico e del contenuto volumetrico di acqua nel tempo. Ciò richiede alcune informazioni, tra cui le proprietà idrauliche del suolo, la dinamica dell'acqua nel terreno e il comportamento al confine del volume di suolo in esame.

Sfruttando il cosiddetto metodo delle differenze finite integrate, noto anche come schema a volumi finiti centrati sulla cella, nel calcolo del bilancio idrico di un frutteto Orchard3D-Lab porta i seguenti contributi:

- un modello idrico 3D capace di simulare il bilancio idrico del suolo;
- un modello di crescita della pianta, che tiene conto del suo sistema radicale e della calibrazione dei parametri dell'albero;
- l'applicazione del calcolo tridimensionale, specifico per un frutteto con terreni in pendenza e con un sistema di irrigazione a goccia;
- l'assimilazione dei dati sfruttando una griglia 2D di sensori.

L'algoritmo 1 descrive il processo iterativo che porta al calcolo dell'umidità del terreno. Nel dettaglio, il simulatore prende come input lo stato iniziale

Algorithm 1 Orchard3D-Lab

Require: SM_{Ass} : assimilated soil moisture, $\vec{W}$: weather time series, $\vec{I}$: irrigation time series, X : simulation parameters
Ensure: $\vec{SM}$: time series of simulated soil moisture

```

1:  $\vec{SM}_0 \leftarrow SM_{Ass}$                                 ▷ Inizializza lo stato del simulatore
2:  $\vec{SM} \leftarrow []$                                     ▷ Inizializza l'accumulatore
3: for each  $t \in \vec{W}, \vec{I}$  do                                ▷ Per ogni timestamp ...
4:    $SM_{t+1} \leftarrow \text{Criteria-3D}(w_t, i_t, SM_t)$         ▷ ... calcola il prossimo stato
5:    $\vec{SM} \leftarrow \vec{SM} \cup (t + 1, SM_{t+1})$             ▷ ... e lo aggiunge
6: end for
7: return  $\vec{SM}$                                            ▷ Restituisce la serie temporale di SM

```

SM_{Ass} calcolato tramite assimilazione, le serie temporali che descrivono le condizioni meteorologiche $\vec{W}$ e le irrigazioni $\vec{I}$ registrate a timestamp successivi, infine l'insieme di parametri X di simulazione, già presentati in Tabella 3.1. Affinché la simulazione proceda correttamente, è necessario che le due serie temporali $\vec{W}$ e $\vec{I}$ siano allineate e abbiano la stessa granularità.

Dopo aver inizializzato lo stato del simulatore con i dati assimilati e la serie temporale che conterrà la SM simulata (linee 1 e 2), per ogni timestamp considerato viene aggiornato lo stato del simulatore usando l'equazione di Richards tramite CRITERIA-3D (linea 4), per poi essere accumulato (linea 5). Per farlo, si considerano lo stato attuale del meteo, dell'irrigazione e del simulatore. Infine, viene restituita la stima dell'umidità del terreno lungo tutto il periodo considerato (linea 7).

Proprietà idrauliche Il calcolo della curva di ritenzione idrica e della funzione di conducibilità idraulica sono di primaria importanza per potere applicare con successo l'equazione di Richards nella stima del bilancio di massa dell'acqua. Tale curva descrive il processo di essiccazione di un certo terreno, correlando il contenuto volumetrico di acqua nel suolo con il rispettivo potenziale idrico. Sebbene il modello di Van Genuchten-Mualem sia quello più comunemente utilizzato, esso può condurre a stime errate. Per questo motivo, Orchard3D-Lab sfrutta una formulazione modificata del modello che include un valore di ingresso dell'aria che ostacola l'entrata dell'acqua e determina la seccatura del terreno [66].

La funzione di conducibilità idraulica, invece, rappresenta la facilità con cui l'acqua attraversa le fratture e gli spazi porosi del terreno. L'abilità del terreno di condurre l'acqua quando i pori sono saturi è veicolata dalla variabile K_{sat} , il cui valore decresce non appena il suolo inizia a desaturare, cioè ad essiccarsi.

Evaporazione e traspirazione Il flusso di acqua in un terreno è fortemente influenzato da due fenomeni che qui si svolgono: l'evaporazione e la traspirazione. Mentre il primo riguarda direttamente il terreno, il secondo è dovuto alla presenza delle piante, le quali, attraverso le radici, dapprima assorbono l'acqua dal suolo per trasmetterla agli apparati fogliari, poi, attraverso i pori e gli stomi, la rilasciano all'atmosfera sotto forma di vapore. Tale processo è innescato dalla fotosintesi, durante cui le piante tengono aperti gli stomi, per potere raccogliere l'anidride carbonica, dai quali fuorisce poi l'acqua.

La somma della quantità di acqua persa dal suolo per evaporazione e quella rilasciata dalle piante per traspirazione determina il fenomeno dell'evapotraspirazione. Si denota con ET_0 l'evapotraspirazione potenziale di riferimento, che viene calcolata tramite l'equazione di Penman-Monteith in relazione a un campo con caratteristiche standard.

Utilizzando alcune variabili aggiuntive è possibile ricavare la massima traspirazione potenziale (T_{pot}) e la massima evaporazione potenziale (E_{pot}). T_{pot} viene suddiviso tra tutti i volumi di suolo (*i.e.*, porzione di suolo che contiene la maggior parte della radici di una pianta) per calcolare l'assorbimento radicale in un punto specifico. Maggiore è la densità delle radici nel volume, tanto più sarà la quantità di T_{pot} che traspira in quel punto.

Infine, per potere risolvere numericamente le equazioni differenziali alle derivate parziali che governano il flusso dell'acqua nell'equazione di Richards, è opportuno considerare le condizioni al contorno per ridurre il numero di variabili sconosciute. Ciò significa definire il comportamento del suolo ai confini del volume simulato in modo da poter determinare la soluzione dell'equazione (*i.e.*, il valore del flusso) nella parte di interesse. Orchard3D-Lab consente di impostare le condizioni al contorno di Dirichlet e Neumann, che definiscono rispettivamente il valore specifico che il flusso deve assumere lungo il confine e la derivata desiderata.

Apparato radicale La forma delle radici più diffusa è quella a "fittone", caratterizzata da una parte centrale più lunga e robusta rispetto alle ramificazioni laterali, ed è anche l'aspetto assunto dalle radici degli alberi di kiwi. Formalmente, l'assetto delle radici può essere descritto considerandone la densità lungo i diversi assi. La Figura 3.2 mostra la variazione della densità delle radici di una pianta di kiwi sul piano xy (a sinistra) e sul piano xz (a destra). Nello specifico, lungo l'asse x la densità radicale varia in modo lineare rispetto alla distanza dal gocciolatore, lungo l'asse y è costante, mentre lungo l'asse z è non lineare. Spostandosi dal centro, la densità delle radici diminuisce sia orizzontalmente che verticalmente.

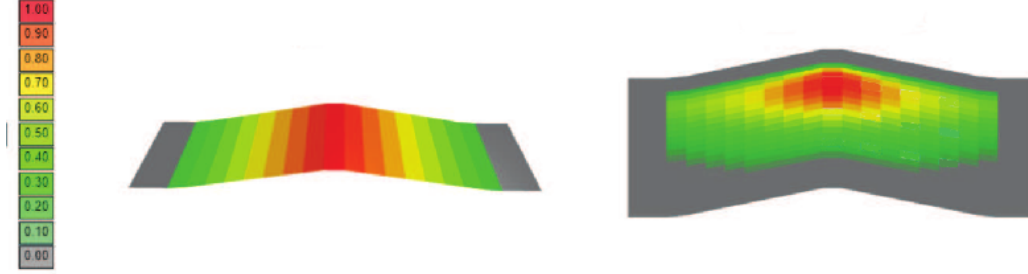


Figura 3.2: Proiezioni della densità delle radici sui piani xy (a sinistra) e xz (a destra)

3.2 Modulo di assimilazione dei dati

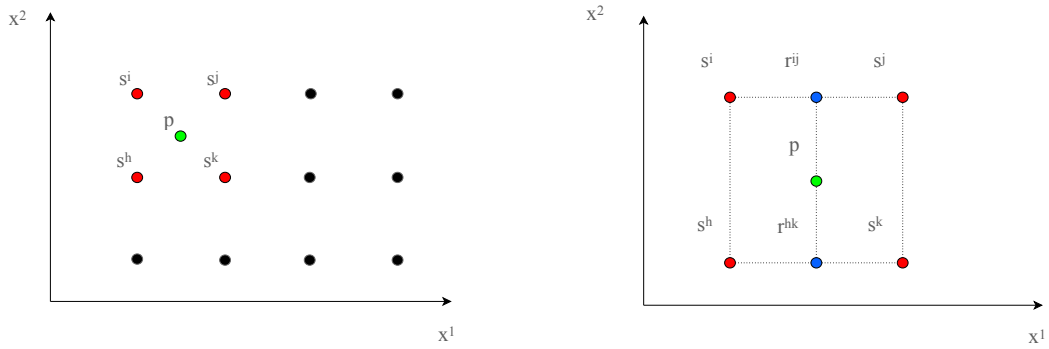
Il modulo di assimilazione dei dati permette di fornire come input al sistema le proprietà basilari del suolo (*i.e.*, SWC e SWP) provenienti dai sensori installati in loco, in modo da impostare lo stato del suolo ad una situazione realmente osservata. Nelle situazioni reali, infatti, non esistono stime perfette e tanti piccoli errori portano il modello a deviare progressivamente, soprattutto per i domini in cui molti fenomeni diversi interagiscono tra di loro. Abbattere in partenza ogni occasione di errore è quindi importante, e l'assimilazione rende possibile inizializzare ogni simulazione con i valori reali di umidità del terreno, in modo da partire da una situazione senza errore. Il processo di assimilazione viene altresì impiegato nel corso della simulazione, per riportare lo stato del suolo alla situazione reale, annullando nuovamente ogni errore.

Scendendo nello specifico, il processo di assimilazione consiste in una tecnica che mappa i dati provenienti da una griglia a grana grossa di sensori del campo nella griglia a grana fine di Orchard3D-Lab, che viene usata come stato iniziale della simulazione e per annullare l'errore nel corso della simulazione. Più sono i sensori e maggiore è la dimensionalità della griglia di partenza, migliore è la granularità del campionamento e quindi l'accuratezza dello stato iniziale.

La funzione di assimilazione si basa sul concetto di *funzione di profilazione*, attraverso la quale è possibile approssimare l'umidità del suolo campionata da una griglia di sensori a grana grossa a una griglia raffinata con granularità più fine:

Definizione 4. Data una griglia di sensori n -dimensionale S e un profilo di umidità P , una funzione di profilazione $f : S \rightarrow P$ approssima i valori del profilo di umidità P a partire dalla griglia di sensori S .

Prendendo come esempio una griglia 2D di sensori, la funzione di profilazione effettua una interpolazione bi-lineare dell'umidità del terreno e procede in due fasi, illustrate in Figura 3.3. Per ciascun punto da calcolare, vengono innanzitutto individuati i quattro sensori che determinano il rettangolo di delimitazione minimo che lo racchiude (prima fase, Figura 3.3a), in seguito si effettua l'interpolazione lungo il primo asse per ottenere i valori di umidità sui punti intermedi (r^{ij} e r^{hk}) che vengono poi a loro volta interpolati lungo l'altro asse per ottenere il valore finale desiderato (p) (seconda fase, Figura 3.3b).



(a) Rettangolo di delimitazione: in verde il punto da approssimare, in rosso i sensori che lo racchiudono

(b) Interpolazione bi-lineare: prima vengono interpolati i punti blu, poi il punto verde

Figura 3.3: Esempio di funzione di profilazione su una griglia 2D di sensori

3.3 Modulo di tuning dei parametri

Il modulo di regolazione dei parametri viene adottato per calibrare in maniera automatica i parametri del modello su un frutteto specifico. Tarare a mano il modello non è infatti un approccio fattibile, a causa dell'elevato numero di parametri da scegliere (Tabella 3.1).

Orchard3D-Lab è l'unico simulatore numerico basato su un'ottimizzazione bayesiana [67], capace di apprendere un modello dello spazio di ricerca al fine di esplorare le regioni più promettenti. Durante l'esplorazione, l'ottimizzatore aggiorna il suo modello finché non viene restituita la configurazione migliore trovata.

Queste tecniche tipicamente sono vincolate a un budget di esplorazione, espresso in termini di tempo di esecuzione o di numero di configurazioni da visitare. Partendo a valutare le prestazioni di una configurazione iniziale scelta in modo potenzialmente casuale, convergono progressivamente verso

combinazioni migliori di parametri, andando di volta in volta ad esplorare in maniera intelligente una diversa porzione dello spazio di ricerca [68].

Algorithm 2 Parameter Tuner

Require: $\vec{W}$: weather time series, $\vec{I}$: irrigation time series, X_{fix} : fixed parameters, X_{tun} : tunable parameters, $budget$: maximum iterations, $\vec{S}$: sensor soil-moisture time series

Ensure: X_{tun}^* : best parameter configuration

```

1:  $SM_{Ass} \leftarrow DataAssimilation(\vec{S}_0)$  ▷ Assimila lo stato iniziale
2:  $H \leftarrow \emptyset$  ▷ Inizializza lo storico delle configurazioni esplorate
3:  $iteration \leftarrow 0$  ▷ Iterazione corrente
4: while  $iteration < budget$  do ▷ Finché rimane del budget ...
5:    $iteration \leftarrow iteration + 1$  ▷ ... incrementa l'iterazione corrente
6:    $X'_{tun} \leftarrow BayesianOpt(H)$  ▷ ... ottimizza i parametri "liberi"
7:    $\vec{SM} \leftarrow Criteria-3D(SM_{Ass}, \vec{W}, \vec{I}, X'_{tun} \cup X_{fix})$  ▷ ... esegue il simulatore
8:    $e \leftarrow error(\vec{SM}, \vec{S})$  ▷ ... ottiene l'errore
9:    $H \leftarrow H \cup \{(X'_{tun}, e)\}$  ▷ ... aggiunge la configurazione allo storico
10: end while
11: return  $argmin_{X_{tun}^* s.t. (X_{tun}^*, e) \in H(e)}$  ▷ Restituisce la migliore configurazione

```

L'algoritmo 2 formalizza il comportamento del tuning, che prende come input le condizioni meteorologiche $\vec{W}$, le irrigazioni $\vec{I}$ e l'umidità del terreno $\vec{S}$ registrate dai sensori a timestamp successivi, un budget massimo di iterazioni $budget$ e infine i parametri X_{fix} fissati manualmente e quelli X_{tun} da trovare con il tuning. L'output restituito rappresenta la configurazione di parametri che permette a Orchard3D-Lab di produrre la migliore approssimazione possibile dei sensori reali. L'errore tra i valori stimati e i valori reali è misurato attraverso il Root-Mean-Square Error (RMSE), di cui si considera la media su tutti i campioni della serie temporale.

Nello specifico, l'algoritmo parte assimilando lo stato iniziale del simulatore sulla base del primo snapshot dell'umidità del terreno campionato (linea 1) e inizializzando la lista delle configurazioni esplorate e il numero di iterazioni (linee 2 e 3). Dopodiché, fin quando il budget non si esaurisce, si procede con l'ottimizzazione bayesiana (linea 6) e con la previsione della SM con CRITERIA-3D (linea 7). Contestualmente, viene calcolato l'errore (linea 8) e la configurazione appena esplorata viene archiviata assieme al relativo errore (linea 9). Quando il budget si esaurisce, la configurazione migliore viene restituita (linea 11).

Capitolo 4

Approccio

Lo scenario di applicazione del sistema di previsione può essere riassunto in Figura 4.1. Inizialmente (fase offline), l'agronomo fornisce l'insieme dei parametri del suolo e della pianta che caratterizzano il frutteto del quale desidera ottimizzare la gestione, assieme a un'indicazione sulla modalità e sulla cadenza delle irrigazioni. Tali parametri vengono quindi forniti, assieme alle variabili meteorologiche, al simulatore Orchard3D-Lab che genera i dati di umidità del terreno in esame. Se i dati di irrigazione non sono noti, è possibile farli generare direttamente al simulatore attraverso la definizione di una specifica regola di irrigazione. Dopodiché, i dati simulati appena creati e le condizioni meteorologiche vengono usati come input per l'algoritmo di Machine Learning, trovato e parametrizzato tramite il meccanismo di SMBO. Ottenuto il modello di riferimento, è possibile procedere con la sua applicazione vera e propria in uno scenario reale (fase online): al modello vengono forniti in input i valori reali di *soil moisture* e di irrigazione provenienti da una certa campagna produttiva del campo e le corrispondenti condizioni meteo per predire i valori futuri di SM.

Avvalersi del simulatore porta con sé alcuni vantaggi, primo tra tutti la possibilità di avere a disposizione i dati di umidità del terreno prima ancora che la campagna produttiva sia iniziata. In tal modo, nel momento in cui questa prende il via, è già disponibile un modello in grado di predire con una certa accuratezza i valori di SM futuri, in base ai quali attuare decisioni più oculate. In questa situazione, il simulatore può essere d'aiuto anche nello stimare preventivamente i parametri del terreno tramite l'apposito modulo di tuning, utilizzandoli poi durante la generazione dei dati di umidità. In ogni caso, la produzione di dati tramite simulazione torna utile per la successiva fase di addestramento del modello di ML, che richiede una discreta mole di dati per poter apprendere e generalizzare correttamente l'andamento di umidità del suolo.

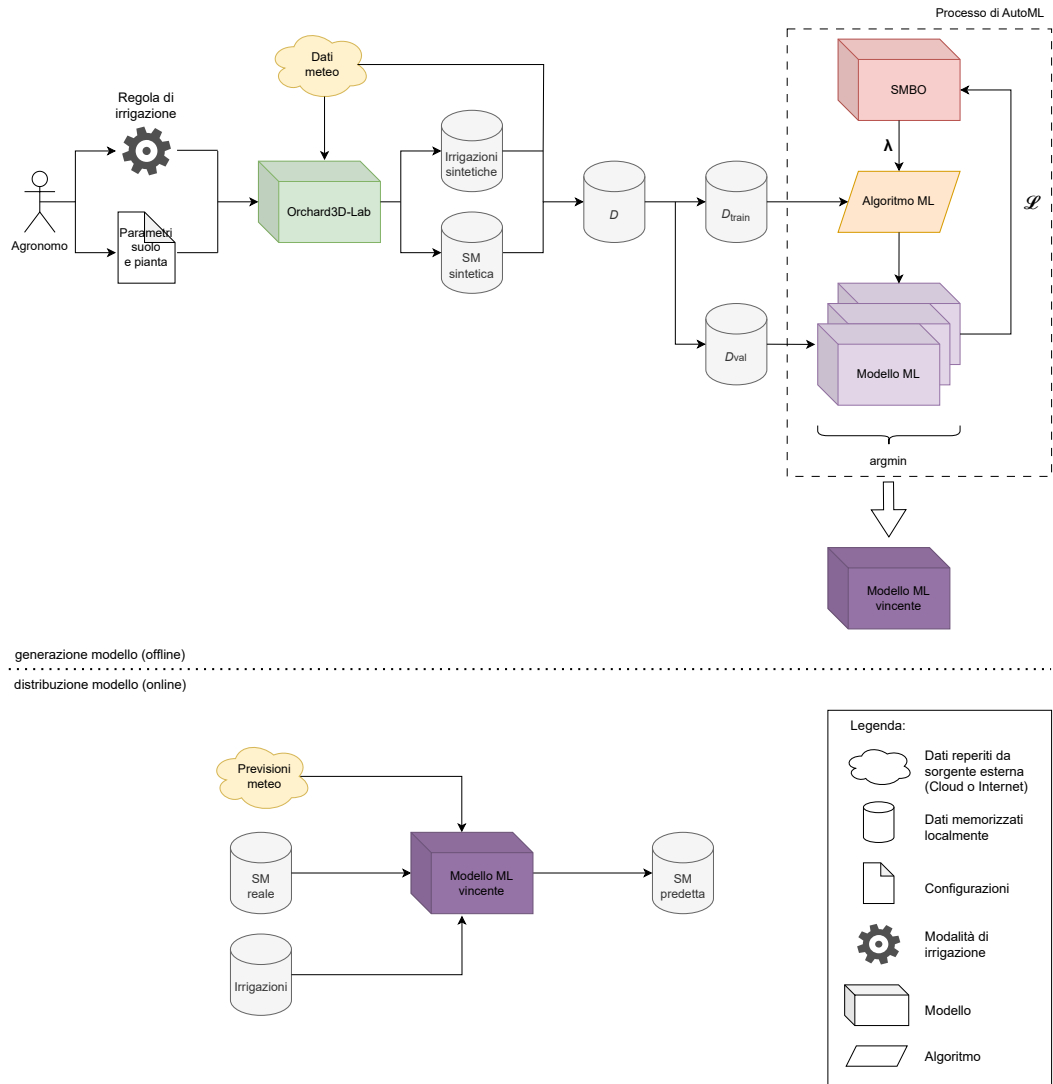


Figura 4.1: Scenario di applicazione del sistema di previsione dell'umidità del suolo

Lo scenario appena illustrato può essere ampliato per supportare un processo periodico di adattamento e miglioramento del modello trovato in prima istanza, come raffigurato in Figura 4.2. Tale meccanismo può applicarsi alle seguenti necessità:

- man mano che la campagna produttiva avanza, si vogliono migliorare le capacità predittive del modello fornendogli periodicamente i dati reali provenienti dal campo.

- si vuole applicare il modello precedentemente addestrato a un altro campo, che può in generale presentare una tessitura e altre proprietà fisiche differenti. In questo caso, l'adattamento avviene utilizzando i dati provenienti dal nuovo campo, i quali dovrebbero far convergere il modello alle nuove caratteristiche del terreno con le successive iterazioni.
- si vuole applicare il modello precedentemente addestrato a piani irrigui diversi da quello di partenza, al fine di adattarsi ad eventuali cambiamenti che potrebbero avvenire negli schemi di irrigazione nel tempo.

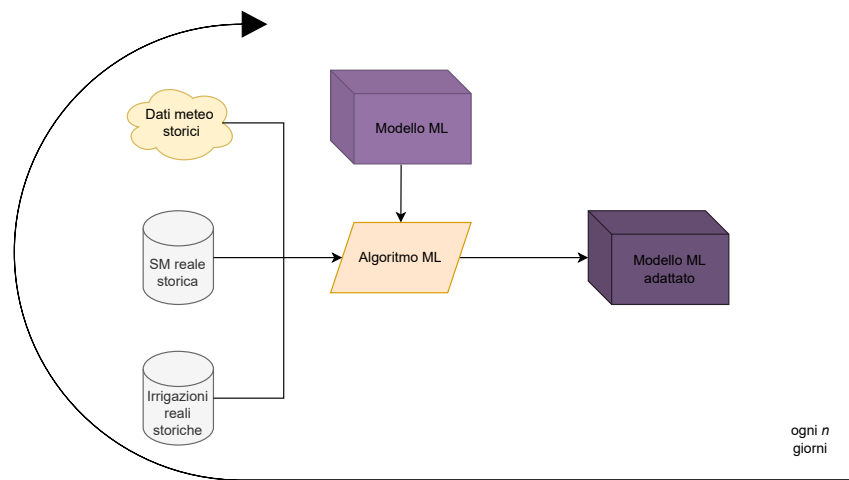


Figura 4.2: Scenario di adattamento del sistema di previsione dell'umidità del suolo

Nel seguito di questa sezione viene presentata la creazione della *rolling window* (sezione 4.1), la definizione del piano irriguo (sezione 4.2) e quindi il funzionamento del modello di previsione, sia nella formulazione base (sezione 4.3) che nei contesti di adattamento (sezione 4.4).

4.1 Creazione rolling window

Avendo a che fare con serie temporali di dati che descrivono l'andamento delle variabili meteorologiche, di irrigazione e di umidità del suolo nel tempo, è opportuno scegliere un'ideale tecnica di analisi e rappresentazione. La *rolling window*, talvolta chiamata *sliding window*, è una tra queste [69]. Essa prevede la definizione di una finestra scorrevole di dimensione m prefissata, la quale rappresenta il numero di osservazioni consecutive della serie che si vogliono includere nella finestra. La scelta della dimensione è condizionata dalla

numerosità T del dataset: più è corta la serie temporale, minore è la dimensione che si tende dare alla finestra, anche se questo può portare in generale a delle stime meno uniformi in confronto a finestre più grandi. Dopodiché, si sceglie l'orizzonte di predizione h e il passo, che determina di quante posizioni la finestra deve scorrere in avanti. Se il passo scelto è 1, allora il dataset verrà suddiviso in $N = T - m + 1$ finestre sovrapposte, che hanno il vantaggio di generare più dati, anche se tra di loro correlati. Così facendo, la prima finestra conterrà le osservazioni dall'istante 1 all'istante m , la seconda finestra quelle che vanno da 2 a $m + 1$ e così via, come illustrato in Figura 4.3.

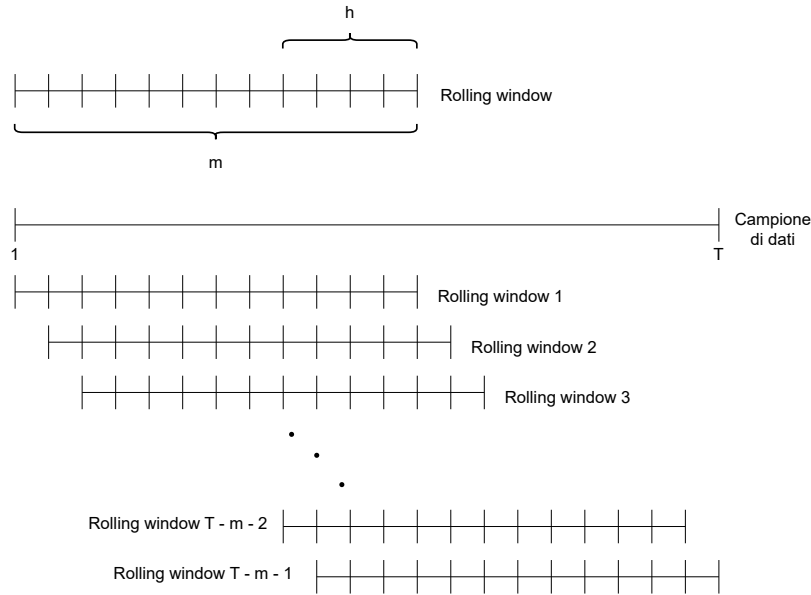


Figura 4.3: Processo generale di creazione delle finestre sovrapposte

In questo lavoro, per la predizione dell'umidità del suolo si considerano congiuntamente le serie temporali relative alle variabili sopracitate, pertanto viene creata una sorta di *rolling window* composta, caratterizzata dalla compresenza di questi valori in ciascuna finestra. Ogni finestra conterrà quindi i valori passati e futuri di temperatura dell'aria, umidità dell'aria, velocità del vento, radiazione solare, precipitazioni, irrigazioni e i valori passati di umidità del terreno per ciascun sensore considerato.

È inoltre possibile applicare una aggregazione (*stride*) ai dati per consentire la riduzione della dimensione delle finestre e rendere la computazione più gestibile. Prendendo come riferimento i tre orizzonti scelti, non effettuare alcuna aggregazione vorrebbe dire ottenere delle finestre con dimensione pari a $(|input\ variables| \cdot horizon \cdot 2) + (|sensors| \cdot horizon)$, chiaramente

troppo complesse. Applicando invece uno *stride*, tale dimensione si riduce a $(|input\ variables| \cdot (\frac{horizon}{stride}) \cdot 2) + (|sensors| \cdot (\frac{horizon}{stride}))$.

4.2 Definizione del piano irriguo

Le variabili di irrigazione usate come input vengono generate mediante Orchard3D-Lab. Sfruttando il calcolo orario dell'evapotraspirazione ET_0 del simulatore, è stato definito e implementato l'algoritmo 3 per la messa a punto di un piano irriguo che sia in grado di riportare il suolo in una condizione di umidità favorevole. La regola di irrigazione scatta ogni d giorni e controlla lo stato di umidità delle celle del volume di suolo: se nelle ultime 12 ore meno della metà delle celle ha avuto un'umidità pari o superiore a -100 cbar (valore corrispondente a uno stato ideale di umidità per la cella) e meno di un quarto ha avuto un'umidità pari o superiore a -30 cbar (valore corrispondente a una cella molto bagnata) allora si può procedere con l'irrigazione, ma solo se la quantità di pioggia cumulata nello stesso lasso di tempo non ha superato i 7 mm (linea 1). In questa situazione, la quantità di acqua fornita sarà pari alla somma di ET_0 degli ultimi $i \cdot 2$ giorni (linea 2) e verrà rilasciata nel frutteto a partire dalle ore 05:00 del mattino, coerentemente con la portata dell'impianto, che è capace di somministrare fino a 4 L di acqua in un'ora.

Algorithm 3 Irrigation Plan

Require: $\vec{W}$: weather time series, C_{SM} : soil moisture map cells, d : days interval, X : simulation parameters
Ensure: $\vec{I}$: time series of simulated irrigation plan
1: **if** $\frac{|C_{SM} \in [-100, 0]|}{|C_{SM}|}$ last 12 h < 0.5 **and** $\frac{|C_{SM} \in [-30, 0]|}{|C_{SM}|}$ last 12 h < 0.25 **and** $\vec{W}_{rainfall}$ last 12 h $\leq$ 7 mm **then**
2: $\vec{I} \leftarrow \sum_{n=1}^{d \cdot 2} ET_0$ ▷ Irriga con tale quantità di acqua
3: **else**
4: $\vec{I} \leftarrow 0$ ▷ Non irriga
5: **end if**
6: **return** $\vec{I}$ ▷ Resituisce la serie temporale del piano di irrigazione

4.3 Modello di previsione

Per l'ottenimento del modello di previsione è stata applicata una pipeline di Machine Learning prestabilita (Figura 4.4), ossia il workflow che porta alla produzione del modello stesso, che nell'ordine comprende: (i) l'acquisizione dei dati, (ii) l'imputazione, (iii) la creazione della rolling window, (iv) la normalizzazione dei dati, (v) la costruzione del modello di ML e (vi) la validazione del modello.

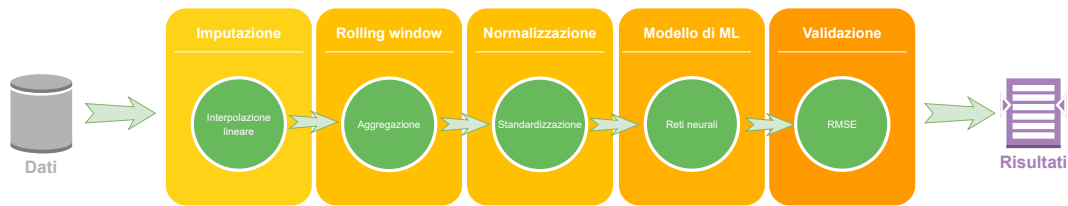


Figura 4.4: Pipeline di ML adottata

Inizialmente, vengono acquisiti i dati, memorizzati su database, riguardanti le variabili meteorologiche, di irrigazione e di umidità del suolo, le ultime due generate tramite Orchard3D-Lab.

Rispettando la granularità oraria del campionamento, nella fase di imputazione vengono aggiunti eventuali timestamp mancanti. Per ognuno di essi, i valori delle variabili corrispondenti vengono calcolati tramite una interpolazione lineare in entrambe le direzioni, ossia ogni valore mancante viene stimato sulla base di quello precedente e di quello seguente.

La successiva creazione della *rolling window* ha l'effetto di trasformare il problema in una predizione di campioni indipendenti: ogni riga della finestra, corrispondente a un certo timestamp, contiene infatti tutte le informazioni utili a predire il valore target di *soil moisture* all'orizzonte desiderato. Queste informazioni includono i valori meteorologici e di irrigazione passati e futuri e i valori passati di SM aggregati in base allo *stride* scelto.

La normalizzazione consente di trasformare i dati in modo che siano più facilmente gestibili dall'algoritmo di ML. La tecnica adottata è quella della standardizzazione, che effettua una trasformazione di scala alle variabili in input, in modo che diventino omogenee.

L'approccio scelto per la costruzione del modello è basato sulle reti neurali, la cui applicabilità nel contesto della predizione dell'umidità futura del suolo è stata dimostrata in letteratura con risultati soddisfacenti. Ci si è affidati dunque a FLAML per il tuning degli iper-parametri, andando a specificare per ognuno di essi un opportuno range di valori potenzialmente ottimali. Ad ogni iterazione del processo di auto-tuning, FLAML prova una specifica istanziazione dello spazio,

procedendo in base alle evidenze riscontrate dal processo di ottimizzazione, fino ad ottenere il modello migliore.

Infine, la fase di validazione permette di misurare e valutare la bontà del modello realizzato, esaminando la sua capacità di generalizzare. Allo scopo, i dati a disposizione vengono suddivisi in tre dataset distinti, che vengono rispettivamente utilizzati per le fasi di addestramento, validazione e testing finale. Al modello viene fornito unicamente il dataset di addestramento, mentre i dati di validazione vengono usati solo dal processo di auto-tuning per scegliere il modello migliore. Il dataset di testing viene quindi usato solamente per il testing, al fine di misurare l'effettiva performance di generalizzazione.

4.4 Adattamento del modello di previsione nel tempo

La Figura 4.5 mostra il processo di adattamento: al modello di base M_0 ottenuto in precedenza viene fornito un nuovo dataset (blocco azzurro) e quindi riaddestrato per ottenere un nuovo modello M_1 , potenzialmente più accurato del precedente, visto che per la sua costruzione sono stati considerati dati aggiornati del campo. Questo processo può essere ripetuto più volte, fino ad ottenere il modello finale M_n .

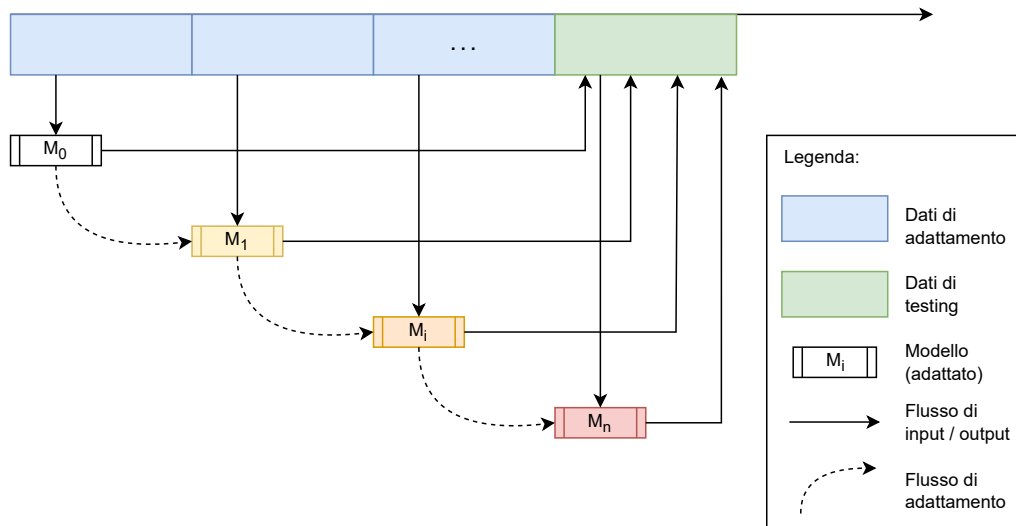


Figura 4.5: Processo di adattamento del modello di previsione nel tempo

Nello specifico, ogni ciclo di adattamento consiste in una ulteriore fase di training applicata alla rete già addestrata in precedenza. In questo modo, i pesi si ricalibrano per consentire l'assimilazione delle nuove informazioni

provenienti dall'ultimo blocco di dati fornito. Tali informazioni possono essere semplicemente i dati aggiornati provenienti dal campo di partenza e alle stesse condizioni di irrigazione, oppure possono fare riferimento allo stesso campo al quale però è stato cambiato il piano irriguo, o ancora possono derivare da un altro campo con differenti caratteristiche di tessitura. L'obiettivo è quello di adattarsi alle nuove condizioni, senza però che la precedente conoscenza acquisita venga perduta. È possibile ottenere un comportamento di questo tipo andando a diminuire l'iper-parametro del *learning rate* della rete.

Capitolo 5

Validazione

In questo capitolo viene presentata la validazione dell’approccio proposto, andando come prima cosa a descrivere nel dettaglio l’effettiva configurazione scelta (sezione 5.1), per poi analizzare i risultati ottenuti dagli esperimenti eseguiti (sezione 5.2).

5.1 Configurazione

Nella modellazione del problema si considerano diversi fattori ambientali che in vario modo sono in grado di influenzare l’umidità del terreno. La struttura del campo, l’impianto di irrigazione e la posizione dei sensori di riferimento sono quelli descritti nella sezione 1.4, relativi a una piantagione di kiwi che si trova ad Errano, in provincia di Ravenna. Tuttavia, la necessità di avere a disposizione una maggiore quantità di dati per l’addestramento preliminare del modello, rende fondamentale l’uso di Orchard3D-Lab per ottenere la quantità di dati richiesta senza dovere aspettare il trascorrere delle campagne produttive. Inoltre, il simulatore risulta essenziale per generare dati provenienti da più zone con caratteristiche fisiche, di tessitura e di irrigazione differenti, di cui non si ha un corrispettivo caso di studio reale. Nello specifico, il simulatore in questione è stato sfruttato per generare i dati relativi all’umidità del terreno e all’irrigazione che verranno forniti come variabili di input per la successiva applicazione degli algoritmi di Machine Learning. Non potendo interagire con il piano irriguo scelto dal tecnico agrario per il campo di Errano, dei dati sintetici vengono sfruttati anche per la fase di testing.

Di seguito, nella sottosezione 5.1.1 si offre una panoramica dei dati meteorologici reali utilizzati, nella sottosezione 5.1.2 vengono elencati i piani irrigui scelti e la modalità di generazione dei dati sintetici di irrigazione, nella sottosezione 5.1.3 si elencano le tessiture selezionate, nella sottosezione 5.1.4 viene descritta la modalità con cui i dati sono stati suddivisi a formare i dataset

di training, validation e testing, nella sottosezione 5.1.5 viene presentata una breve disamina degli algoritmi selezionati e dei loro spazi di ricerca, nella sottosezione 5.1.6 si descrive la modalità di valutazione dei modelli, infine nella sottosezione 5.1.7 si entra nello specifico della procedura di adattamento del modello.

5.1.1 Analisi dei dati meteorologici

Come si accennava all'inizio, i valori futuri di umidità del suolo non solo dipendono dai rispettivi valori passati, bensì anche dalle condizioni meteo e di irrigazione del campo stesso. Per quanto riguarda le condizioni meteorologiche, sono stati considerati i dati reali misurati e raccolti da ARPAE [70] relativamente a due distinte località della regione Emilia-Romagna, rispettivamente Martorano (in provincia di Forlì-Cesena) e Bologna, per i motivi esposti sopra. Tali dati comprendono la temperatura dell'aria ($^{\circ}\text{C}$), l'umidità dell'aria (%), la velocità del vento (Km/h), la radiazione solare (W/mq) e le precipitazioni (mm). Ciascuno di essi è campionato con cadenza oraria ed è codificato nel formato $\langle \text{timestamp}, \text{value} \rangle$, dove il primo attributo rappresenta l'ora della misurazione espressa in *unix time*¹ e il secondo il valore della specifica variabile nell'unità di misura di riferimento. In Figura 5.1 è riportato l'andamento di ciascuna variabile misurata a Martorano nel periodo che va da giugno a settembre 2017. Il valore associato ad ogni timestamp fa riferimento alla *media* dei valori assunti da ciascuna variabile nel corso dell'ora, ad eccezione della variabile associata alle precipitazioni, il cui valore fa riferimento alla quantità di acqua *cumulata* rispetto al precedente timestamp. I dati testimoniano un periodo notevolmente caldo e soleggiato con picchi di 38.7°C e con un'umidità che ha raggiunto il 91%. I mesi considerati sono stati inoltre caratterizzati da scarso vento e da piogge complessivamente modeste, anche se quest'ultime si sono concentrate in brevi periodi di qualche ora, determinando così dei picchi di moderata intensità che hanno superato gli 11 mm in un'ora.

5.1.2 Scelta dei piani irrigui

Sono stati considerati tre distinti piani irrigui che si differenziano per l'intervallo di irrigazione d prescelto: 2, 4 e 7 giorni.

Le informazioni relative all'irrigazione somministrata alla coltura vengono stimate attraverso il simulatore e ricavate sulla base della quantità di acqua evapotraspirata (ET_0) (vedi sezione 4.2). Nel dettaglio, per avviare una simulazione, Orchard3D-Lab prende come input le variabili meteo sopra menzionate,

¹Lo *unix time* è un formato di rappresentazione del tempo, espresso come numero di secondi intercorsi dalla mezzanotte (UTC) del 1° gennaio 1970 ad oggi.

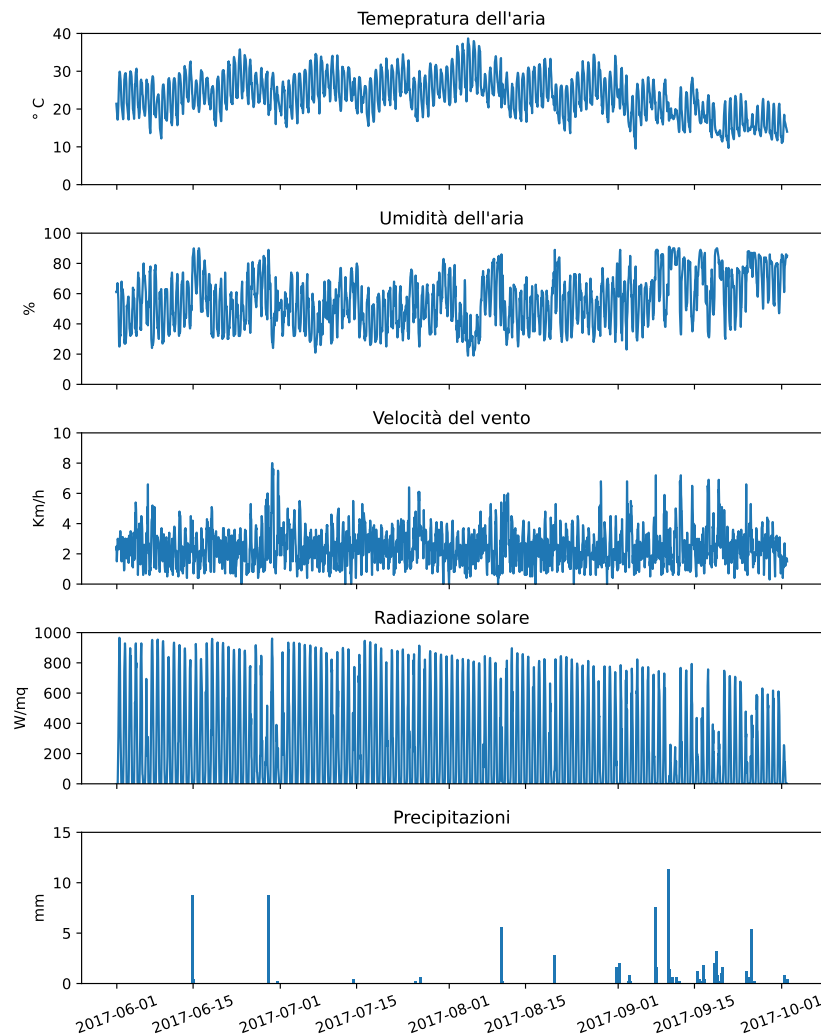


Figura 5.1: Dati meteorologici reali rilevati a Martorano nel periodo giugno-settembre 2017

assieme ai parametri specifici che identificano un particolare tipo di terreno, possibilmente derivanti da una precedente fase di tuning condotta dal relativo modulo preposto (vedi sezione 3.3). Se le irrigazioni non vengono fornite, come in questo caso, è possibile impostare uno stato iniziale di umidità del terreno che verrà usato dal simulatore per avviare i calcoli sulla dinamica dell'acqua. Si è così scelto di inizializzare lo stato del suolo ad un potenziale idrico di -30 cbar, corrispondente alla *field capacity*.

La variabile riguardante l'irrigazione, misurata in litri, segue lo stesso campionamento e codifica delle informazioni meteo, mentre il valore associato

ad ogni timestamp segue la stessa semantica delle variabili di precipitazione. In Figura 5.2 è riportato il dettaglio del suo andamento nel corso di un paio di giorni, ottenuto facendo riferimento alle condizioni meteo di Martorano, alla tessitura del campo di Errano e a un intervallo di irrigazione pari a 2 giorni. Si nota come la quantità massima di acqua erogata sia 4 L/h, che corrisponde all'effettiva portata del gocciolatore dell'impianto di riferimento.

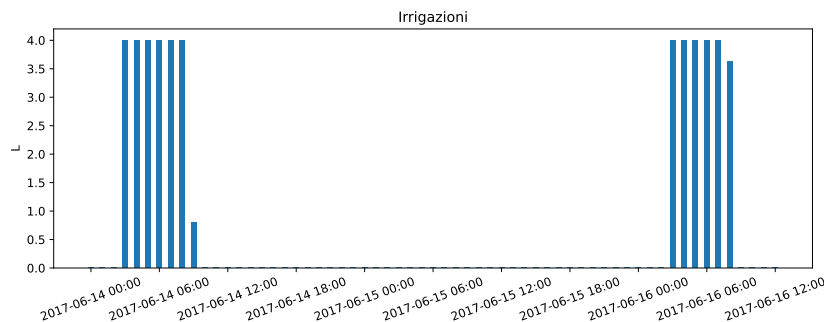


Figura 5.2: Dati di irrigazione simulati mediante Orchard3D-Lab, in riferimento allo scenario meteo di Martorano, alla tessitura di Errano e all'intervallo di irrigazione di 2 giorni (periodo dal 14 giugno al 16 giugno 2017)

La variazione dell'intervallo con cui si irriga il campo influisce sull'andamento di asciugatura del terreno stesso. Con riferimento alla configurazione citata sopra, la Figura 5.3 mostra l'andamento di umidità specifico dei nove sensori considerati dal sistema predittivo, ossia quelli posti alle profondità di 20, 40, 60 centimetri e alle distanze di 0, 25 e 50 centimetri dal gocciolatore. Il valore associato ad ogni timestamp rappresenta l'umidità puntuale (simulata) in quell'istante di tempo. Si può osservare chiaramente come i sensori che si trovano in prossimità del gocciolatore e negli strati più superficiali del suolo (in alto a sinistra) registrino una situazione di costante ed elevata umidità, che risponde prontamente alle irrigazioni somministrate. Allontanandosi dalla fonte di irrigazione e spostandosi verso il basso, si assiste invece ad una progressiva essiccazione del terreno, data dal fatto che l'acqua ha più difficoltà a raggiungere queste zone.

Aumentando l'intervallo a 4 giorni (Figura 5.4) si può osservare che le fluttuazioni per i sensori superficiali iniziano a diminuire in numero, essendo le irrigazioni più rade, e allo stesso tempo la maggiore quantità d'acqua fornita al campo fa sì che anche gli strati più in profondità siano influenzati dal piano irriguo, perciò i relativi sensori non rilevano più la graduale perdita di contenuto idrico di prima. L'andamento appena descritto risulta ancora più evidente considerando un intervallo di 7 giorni (Figura 5.5), con il quale emergono picchi di asciugatura ancora più marcati e radi.

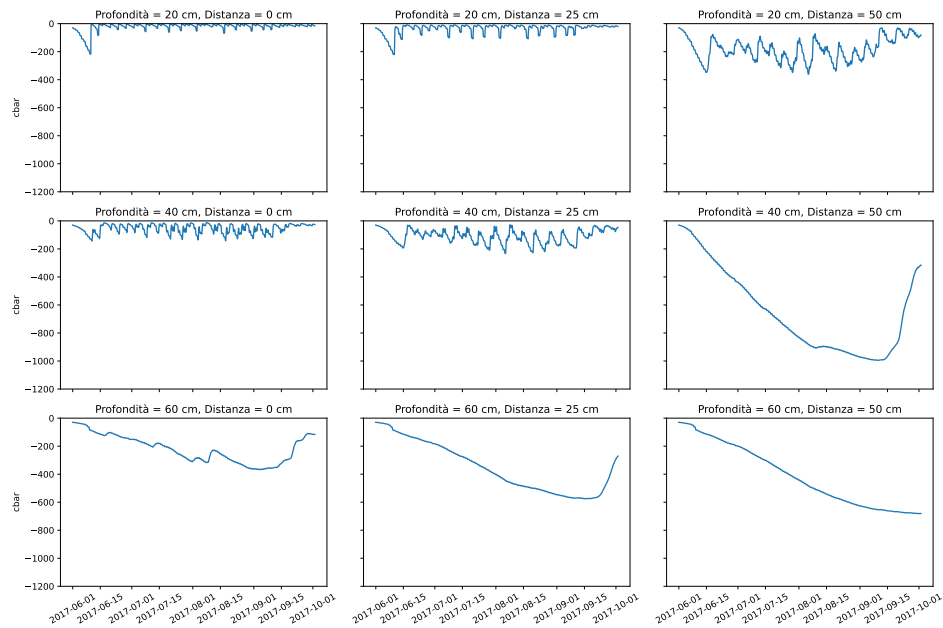


Figura 5.3: Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 2 giorni

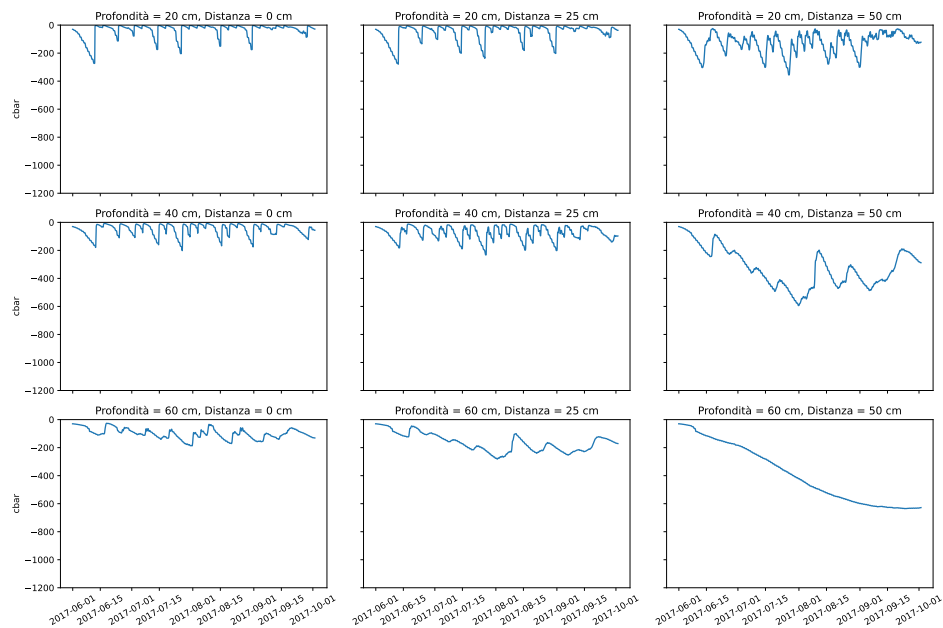


Figura 5.4: Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 4 giorni

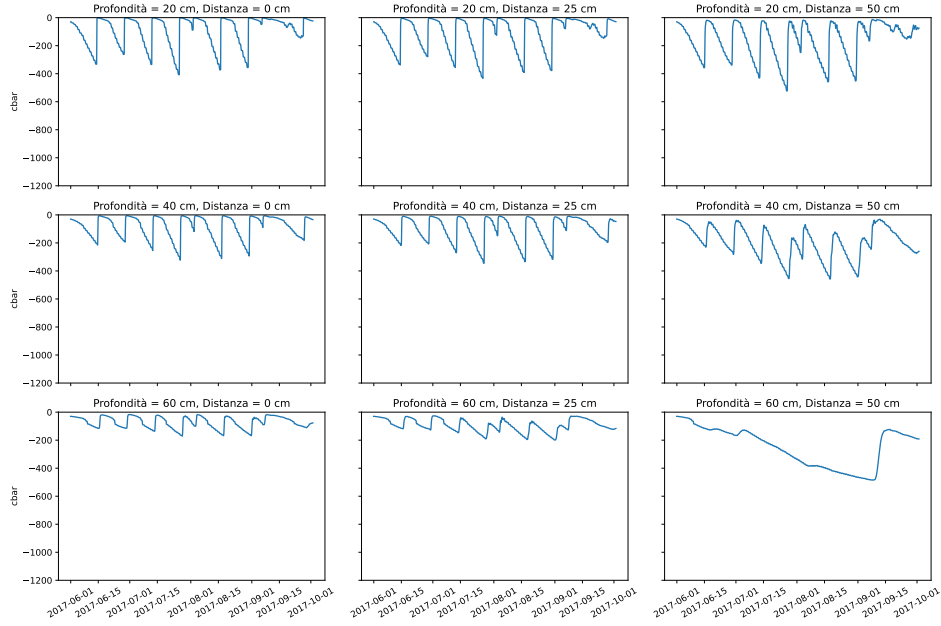


Figura 5.5: Andamento dell'umidità del suolo per ciascun sensore con intervallo di irrigazione pari a 7 giorni

5.1.3 Scelta delle tessiture

Le diverse tipologie di suolo considerate in questo lavoro, ognuna localizzabile in un punto specifico del triangolo (Figura 5.6), e le rispettive proprietà fisiche, sono le seguenti:

- **CLAYEY** (campo di riferimento situato ad Errano)
 - tessitura franco argillosa (*clay loam*): 30% sabbia, 30% limo e 40% argilla (Figura 5.6a)
 - $he = 0.23$, $\alpha = 1.9$, $n = 1.5$, $\theta_r = 0.03$, $\theta_s = 0.5$, $K_{sat} = 1.1 \cdot 10^{-6}$
- **SANDY**
 - tessitura franco sabbiosa (*sandy loam*): 65% sabbia, 25% limo e 10% argilla (Figura 5.6b)
 - $he = 0.15$, $\alpha = 2.84$, $n = 1.4$, $\theta_r = 0.03$, $\theta_s = 0.350$, $K_{sat} = 1.1 \cdot 10^{-6}$
- **SILTY**
 - tessitura franco limosa (*silt loam*): 10% sabbia, 74% limo e 16% argilla (Figura 5.6c)
 - $he = 0.27$, $\alpha = 1.37$, $n = 1.3$, $\theta_r = 0.03$, $\theta_s = 0.448$, $K_{sat} = 1.33 \cdot 10^{-5}$

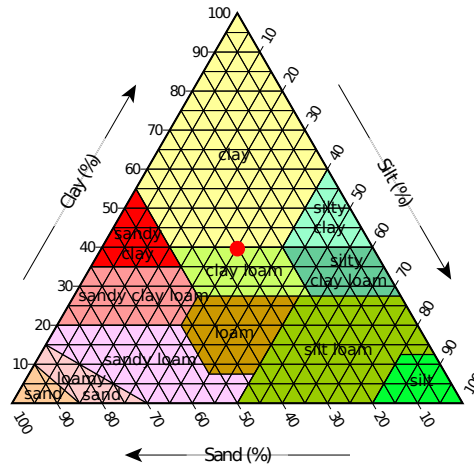
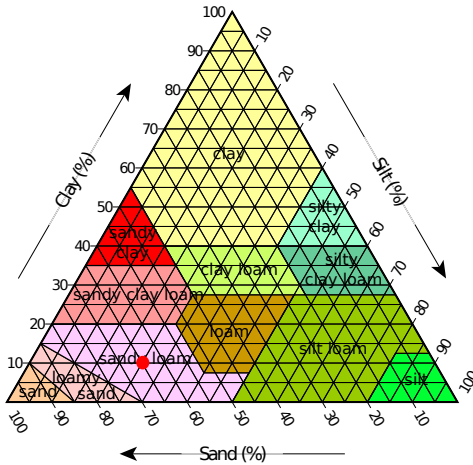
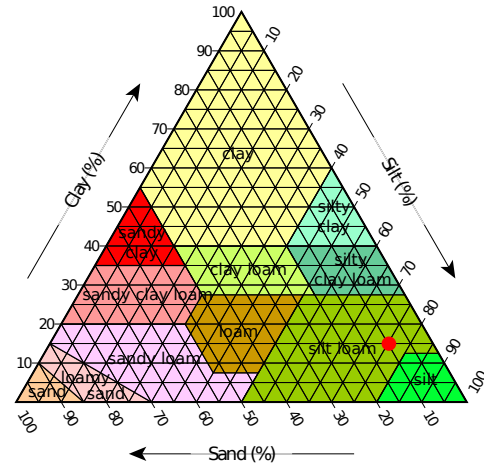
(a) Franco argillosa (**CLAYEY**)(b) Franco sabbiosa (**SANDY**)(c) Franco limosa (**SILTY**)

Figura 5.6: Tre differenti tessiture selezionate

Similmente ai piani irrigui, anche la scelta della tessitura influisce direttamente sulla dinamica dell'acqua nel suolo. Il terreno SANDY (Figura 5.7), per le specifiche proprietà fisiche, presenta scarse capacità di trattenere l'acqua e ciò si traduce in picchi più profondi e meno frequenti di asciugatura rispetto ad altri terreni, come il SILTY (Figura 5.8), che si mantiene invece a potenziali maggiori a quasi tutte le profondità e distanze. La maggiore frequenza di alti e bassi nell'andamento dell'umidità in quest'ultimo suolo conferma la sua maggiore sensibilità alle irrigazioni.

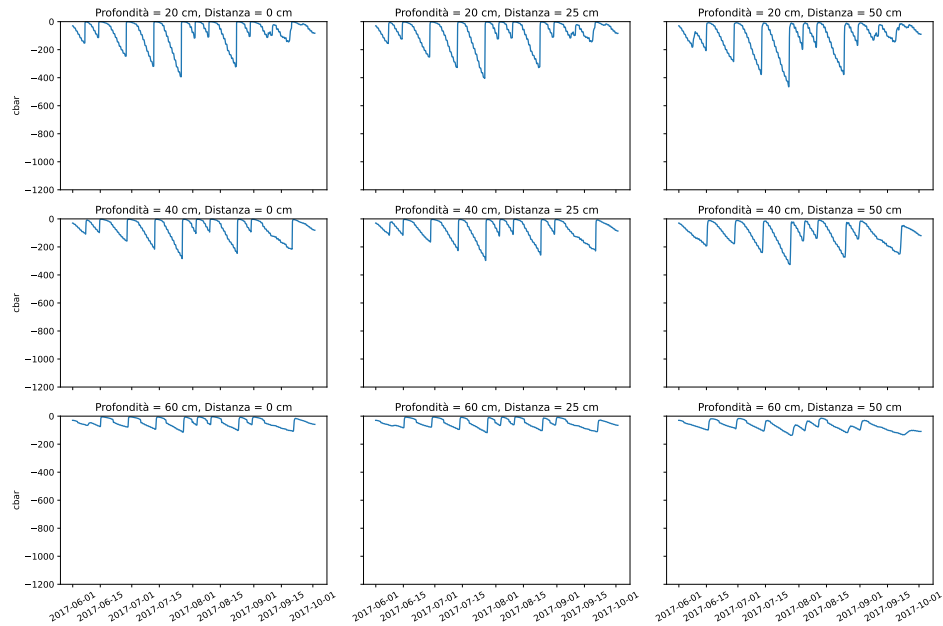


Figura 5.7: Andamento dell'umidità del suolo SANDY, relativamente allo scenario meteo di Martorano 2017 e all'intervallo di irrigazione di 7 giorni

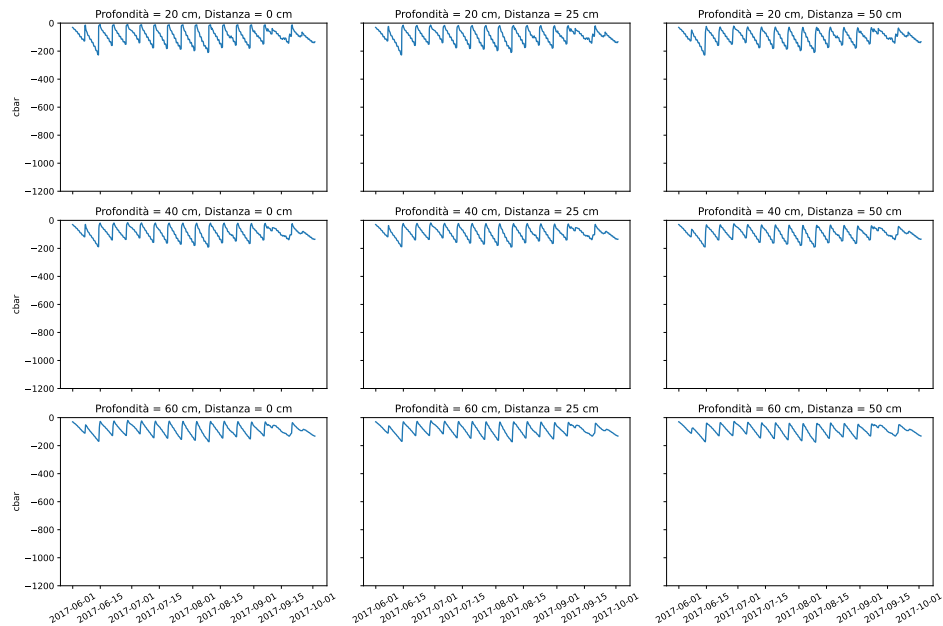


Figura 5.8: Andamento dell'umidità del suolo SILTY, relativamente allo scenario meteo di Martorano 2017 e all'intervallo di irrigazione di 7 giorni

5.1.4 Suddivisione dei dataset di training, validation e testing

Per il dataset di training vengono considerate le stagioni produttive di Martorano, che fanno riferimento al periodo che va da inizio giugno a fine settembre, relative agli anni 2017, 2018, 2019 e 2020. Per il dataset di validazione sono stati impiegati i dati di Bologna relativi agli anni 2016, 2017 e 2018, ciascuno dei quali copre il periodo di quattro mesi sopra menzionato. I dati di Bologna dell'anno 2019 sono invece stati utilizzati come dataset di testing, che non viene in nessun modo coinvolto nella precedente fase di addestramento e che serve per simulare il comportamento del modello sui dati reali.

Per quanto riguarda l'orizzonte di predizione (*horizon*), si è scelto di considerare tre istanti futuri differenti: 48, 96 e 168 ore, rispettivamente corrispondenti a 2, 4 e 7 giorni. Lo scopo è quello di osservare eventuali variazioni nella capacità predittiva all'aumentare dell'orizzonte di output considerato. I valori da predire fanno riferimento all'umidità del suolo espressa in termini di potenziale idrico.

In relazione all'aggregazione da applicare ai dati, è stato utilizzato uno *stride* pari al corrispondente *horizon*. Nello specifico, l'aggregazione considera la media dei valori assunti dalla temperatura e umidità dell'aria, dalla velocità del vento e dalla radiazione solare nelle precedenti/seguinti ore considerate dallo *stride*, la somma delle irrigazioni e precipitazioni avvenute nelle precedenti/seguinti ore e l'ultimo valore di umidità del suolo registrato da ogni sensore, cioè quello associato al timestamp corrente.

5.1.5 Selezione degli algoritmi e degli spazi di ricerca

Di seguito vengono elencati gli algoritmi forniti a FLAML, lo strumento di AutoML scelto per questo lavoro. Durante il processo di auto-tuning, per il quale si è scelto di utilizzare l'algoritmo BlendSearch, FLAML seleziona uno degli algoritmi, prende in input il corrispondente spazio di ricerca e ad ogni iterazione prova una specifica istanziazione dello spazio, procedendo in base alle evidenze riscontrate dal processo di ottimizzazione. Selezionati una serie di algoritmi classici come punto di partenza, l'obiettivo sarà quello di superare le loro performance attraverso l'approccio con reti neurali, su cui si basa il sistema adattivo.

Persistent System Non si tratta di un algoritmo addestrabile, bensì di un approccio naïf deterministico. In questo caso, non viene presa in considerazione nessuna variabile in input: la predizione all'orizzonte (*horizon*) desiderato sarà univocamente determinata a priori e corrispondente al valore di *soil*

moisture registrato *horizon* ore prima. Il risultato complessivo di tale predizione corrisponderà pertanto alla serie temporale di SM, traslata di *horizon* ore avanti. L'obiettivo è quello di superare in termini di performance i risultati ottenibili con questo meccanismo, che rappresenta il punto di partenza rispetto al quale migliorare le capacità predittive degli altri algoritmi.

Linear Regression La regressione lineare multivariata può essere impiegata per stimare il valore dell'umidità del suolo basandosi su più variabili che assieme ne influenzano la predizione, ovvero le variabili meteo, di irrigazione e le variabili di SM stesse. Per modellare la relazione tra le variabili, viene definito un iper-piano tramite la tecnica dei minimi quadrati, il cui obiettivo è quello di minimizzare la somma dei quadrati delle distanze tra i dati reali e l'iper-piano, in modo da trovare quello ottimale. Per la sua applicazione si è fatto riferimento all'implementazione fornita dalla libreria scikit-learn [71] di Python, che permette di scegliere un ristretto numero di iper-parametri con cui tarare il modello. Ogni iper-parametro e il relativo intervallo di valori, riportati in Tabella 5.1, viene quindi considerato dal modulo di AutoML per la ricerca della combinazione ottimale.

Iper-parametro	Descrizione	Intervallo di valori
<code>fit_intercept</code>	Stabilisce se includere il calcolo dell'intercetta nell'individuazione dell'iper-piano	{true, false}
<code>positive</code>	Stabilisce se imporre o meno ai coefficienti dell'iper-piano di essere positivi	{true, false}

Tabella 5.1: Spazio di ricerca per Linear Regression

Random Forest Questo algoritmo sfrutta tecniche di *ensemble* per combinare più modelli e quindi migliorare la capacità predittiva generale del sistema. Ad alto livello, l'approccio consiste nella creazione di un insieme di alberi decisionali, ciascuno dei quali esegue l'addestramento su un diverso sottoinsieme dei dati e considera una porzione casuale delle variabili di input. Le predizioni dei diversi alberi su un certo elemento di testing vengono quindi mediate per ottenere la predizione finale. Gli iper-parametri selezionati dall'implementazione di scikit-learn dell'algoritmo sono riportati in Tabella 5.2.

SVR Ogni elemento di addestramento viene rappresentato come un punto in uno spazio multi-dimensionale, nel quale ogni variabile di input rappresenta una dimensione. L'obiettivo è trovare un iper-piano che massimizzi la distanza tra i punti di addestramento e il piano stesso. Gli iper-parametri selezionati dall'implementazione di scikit-learn sono elencati in Tabella 5.3.

Iper-parametro	Descrizione	Intervallo di valori
<code>n_estimators</code>	Numero di alberi considerati	{10, 30, 50, 70, 100}
<code>criterion</code>	Funzione per stabilire la qualità di una suddivisione del dataset	{"squared_error", "absolute_error"}
<code>max_features</code>	Numero di variabili da considerare per la suddivisione migliore	{"auto", "sqrt", "log2"}
<code>bootstrap</code>	Stabilisce se effettuare la suddivisione del dataset	{true, false}
<code>oob_score</code>	Stabilisce se utilizzare campioni esterni per stimare il punteggio di generalizzazione	{true, false}
<code>warm_start</code>	Stabilisce se partire dalla soluzione precedente nella costruzione del successivo albero	{true, false}

Tabella 5.2: Spazio di ricerca per Random Forest

Iper-parametro	Descrizione	Intervallo di valori
<code>kernel</code>	Tipologia di <i>kernel function</i>	{"linear", "poly", "rbf", "sigmoid"}
<code>degree</code>	Grado della <i>kernel function</i> polinomiale	{2, 3, 4, 5}
<code>gamma</code>	Coefficiente del kernel	{"scale", "auto"}
<code>tol</code>	Termine indipendente della <i>kernel function</i>	$[10^{-4}, 10^{-2}]$
<code>C</code>	Parametro di regolarizzazione	{0.001, 0.01, 0.1, 1, 5}
<code>epsilon</code>	Parametro ϵ	[0.001, 1]
<code>shrinking</code>	Stabilisce se usare l'euristica di riduzione	{true, false}

Tabella 5.3: Spazio di ricerca per SVR

Reti neurali Per l'implementazione delle reti neurali si è fatto uso della libreria Keras [72], che funge da interfaccia per la sottostante libreria di back-end TensorFlow [73]. Affinché la rete riesca a convergere con più facilità, riducendo i rischi di scomparsa o esplosione del gradiente, viene applicata una normalizzazione ai dati di input, in modo da annullare la differenza di scala delle diverse variabili. La tecnica scelta è quella della standardizzazione, o *z-score*, che trasforma ciascuna variabile di input (x) in modo che abbia media (μ) zero e deviazione standard (σ) unitaria, applicando la formula $z = \frac{x-\mu}{\sigma}$. Ciò rende più stabile il processo di addestramento e può migliorare la capacità di generalizzazione della rete, evitando che certe variabili predominino sulle altre a causa delle loro scale originali. Quando l'addestramento termina, il risultato viene nuovamente de-normalizzato, in modo da ritornare alla scala di partenza.

Scendendo nel dettaglio, le reti implementate presentano una quantità di neuroni di input pari al numero di variabili di ingresso: sei nodi per i dati meteorologici, un nodo per i dati di irrigazione e nove nodi per i sensori considerati. Il livello di output è invece formato da nove nodi, che forniscono i valori di SM per ogni sensore. I livelli nascosti sono completamente connessi e terminano con un livello di *dropout*, attraverso il quale vengono disattivati in maniera casuale un certo numero di neuroni per prevenire l'overfitting. L'architettura di tali livelli è stata oggetto di tuning, cercando di realizzare

una struttura che ricordasse quella degli autoencoder, caratterizzata da un progressivo restringimento nel numero di neuroni dei livelli intermedi, seguito da una speculare parte di allargamento. Così facendo, la rete è costretta a comprimere i dati in input e quindi a catturarne gli aspetti più importanti.

In Tabella 5.4 vengono elencati nel dettaglio tutti gli iper-parametri considerati.

Iper-parametro	Descrizione	Intervallo di valori
num_epochs	Numero di epoche considerate nell'addestramento	{256, 512, 1024}
batch_size	Numero di dati gestiti in una singola iterazione di addestramento	{64, 128, 256, 512}
num_hidden_layers	Numero di livelli intermedi	{2, 3, 4, 5}
num_neurons	Numero di neuroni per ogni livello intermedio	{16, 32, 64, 128, 256}
encoder	Stabilisce se abilitare la creazione dell'architettura ad autoencoder	{true, false}
dropout	Percentuale di <i>dropout</i>	[0.0, 0.9]
activation_function	Funzione di attivazione dei livelli intermedi	{"sigmoid", "tanh", "ReLU"}
optimizer	Ottimizzatore dei parametri del modello	{"SGD", "Adadelta", "Adam"}

Tabella 5.4: Spazio di ricerca per le reti neurali

Per quanto riguarda il processo di auto-tuning di FLAML, si è scelto un budget che fosse coerente con la complessità dello spazio di ricerca di input. Per gli esperimenti che hanno impiegato gli algoritmi più classici, si è predisposto un budget di 1 ora, mentre per quelli che hanno fatto uso di reti neurali si è usato un budget di 24 ore, che ha permesso di raggiungere una copertura di esplorazione di circa il 2.6%.

5.1.6 Valutazione dell'errore del modello

Per valutare le performance dei modelli creati ci si è avvalsi del Root-Mean-Square Error, al fine di calcolare l'errore commesso nella predizione dei valori di umidità del suolo. Considerato il dataset di test con cardinalità n , il valore osservato y_i per una certa istanza i e il corrispondente valore predetto $\hat{y}_i$, l'errore complessivo di predizione sulle istanze del dataset può essere espresso come segue:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.1)$$

ovvero come la radice quadrata della media dei quadrati degli errori tra le previsioni del modello e i valori effettivi di SM. Tale misura rappresenta quindi l'errore che viene minimizzato dalle reti sul dataset di validation durante la fase di addestramento. Grazie alla fase di de-normalizzazione finale, tale valore corrisponde esattamente alla differenza in cbar tra il valore reale e quello predetto. Per ciascun esperimento eseguito si calcola sia il valore di RMSE specifico di ogni sensore, sia l'errore complessivo ottenuto facendo la media dei singoli errori dei sensori.

5.1.7 Adattamento del modello

I dati relativi all'anno 2019 di Bologna sono stati in realtà suddivisi in quattro parti, l'ultima delle quali è la porzione effettivamente utilizzata per il testing. Le porzioni precedenti sono state lasciate da parte nella costruzione del modello di base per poterle sfruttare nella fase di adattamento, simulando così l'acquisizione progressiva dei dati registrati dai sensori del campo in tempo reale, che descrivono l'avanzare della campagna produttiva corrente. Nello specifico, dell'intero periodo di riferimento, che va dal 16 giugno 2019 al 2 ottobre 2019, gli ultimi 33 giorni compongono il dataset di test, mentre i precedenti tre blocchi da 25 giorni l'uno rappresentano il dataset di training di ciascun ciclo di adattamento che il sistema prevede.

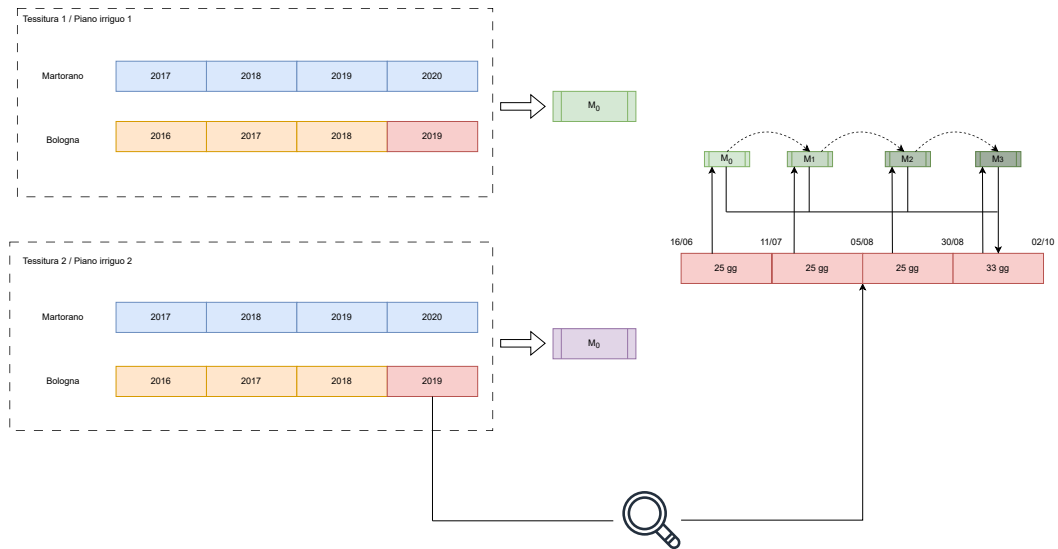


Figura 5.9: Processo di adattamento del modello di previsione da una tessitura e l'altra o da un piano irriguo e l'altro

La procedura di adattamento descritta nella sezione 4.4 si presta ad essere utilizzata nel caso si voglia adeguare il modello di base a un diverso piano

irriguo o a una differente tessitura rispetto a quelle utilizzate in partenza. Nel primo caso, i blocchi di dati forniti ad ogni ciclo di adattamento faranno riferimento al nuovo schema di irrigazione, nel secondo caso faranno invece riferimento al nuovo terreno. La Figura 5.9 esemplifica tale procedimento ad alto livello, per il quale sono stati previsti tre cicli successivi di addestramento. Nello specifico, ogni ciclo di addestramento ha previsto un nuovo allenamento della rete ottenuta in precedenza, utilizzando un *learning rate* pari a 0.0001, sufficientemente piccolo per non perdere la conoscenza pregressa, e un numero di epoche pari a 256.

5.2 Esperimenti

In questa sezione si presentano gli esperimenti condotti e i rispettivi risultati, procedendo in maniera incrementale. I modelli generati vengono innanzitutto testati considerando il medesimo scenario utilizzato per l'addestramento, caratterizzato dalla specifica tessitura e piano irriguo, oltre che dalle condizioni meteo (sottosezione 5.2.1). Successivamente, tali modelli vengono sfruttati per predire la *soil moisture* in scenari diversi da quello di partenza, nei quali a variare è la tipologia di terreno, lo schema di irrigazione o entrambi (sottosezione 5.2.2). Infine, i modelli vengono sottoposti al processo di adattamento, che si articola in tre cicli successivi di ri-addestramento delle reti neurali, al fine di osservare come cambiano le prestazioni sui nuovi scenari man mano che al modello vengono forniti sempre più dati specifici di questi ultimi (sottosezione 5.2.3).

5.2.1 Test dei modelli su scenari conosciuti

Per ogni algoritmo considerato, la Tabella 5.5 riporta i risultati ottenuti per ciascuna tessitura (CLAYEY, SANDY e SILTY), piano irriguo (con intervallo di irrigazione pari a 2, 4 e 7 giorni) e orizzonte (2, 4 e 7 giorni), evidenziando in grassetto i migliori risultati ottenuti per ogni combinazione di tessitura e piano irriguo. Ogni modello viene addestrato su un certo scenario, determinato dalla combinazione di tessitura e piano irriguo, a predire il valore dell'umidità del terreno diversi giorni avanti, e viene quindi testato sullo stesso scenario.

Come si può notare, le reti neurali ottengono buone prestazioni, superando in molti casi quelle degli altri algoritmi presi come riferimento. In generale, le performance diminuiscono all'aumentare dell'orizzonte scelto, dal momento che si tratta di un problema più complesso. Aumentando l'intervallo di irrigazione, invece, le prestazioni rimangono abbastanza costanti, ad eccezione di Persistent System per il quale peggiorano. Anche per tutte le altre combinazioni, quest'ultimo è l'algoritmo che ottiene gli errori più elevati. Relativamente al

terreno franco limoso (SILTY), si riscontrano valori di RMSE più bassi rispetto alle altre tessiture, e ciò può essere dovuto al suo andamento più regolare e compreso in un intervallo più ristretto di valori (vedi Figura 5.8).

Tessiture e piani irrigui		Algoritmi e orizzonti														
		PS			LR			RF			SVR			NN		
		2	4	7	2	4	7	2	4	7	2	4	7	2	4	7
CLAYEY	2	29.67	45.73	58.99	15.50	22.78	44.52	20.27	23.15	39.16	16.68	26.09	36.17	18.80	22.08	25.57
	4	45.49	64.43	84.51	27.43	35.47	43.99	35.14	39.92	46.56	28.72	37.88	46.46	23.14	25.15	41.40
	7	66.82	98.31	119.96	27.81	36.25	47.59	30.45	34.56	42.82	26.80	38.07	45.99	19.27	33.38	40.84
SANDY	2	30.02	44.01	54.33	16.41	21.91	28.63	22.22	25.74	30.39	16.57	23.16	32.46	14.93	20.46	29.99
	4	35.36	46.58	44.34	19.89	24.50	30.58	23.60	29.44	32.80	22.18	32.99	25.39	15.99	25.40	29.96
	7	60.07	85.08	101.44	22.16	39.92	50.12	23.17	21.37	20.43	18.94	21.76	25.86	13.09	23.25	27.93
SILTY	2	17.31	19.18	28.87	6.70	8.50	11.51	8.50	11.25	15.83	7.06	8.84	13.70	5.14	7.53	12.33
	4	34.14	22.47	35.76	10.14	9.65	13.87	13.08	14.59	16.98	10.08	9.72	14.87	7.65	8.27	13.39
	7	44.23	52.08	36.06	10.76	11.15	8.84	7.51	9.66	12.83	11.54	10.85	8.95	5.62	6.81	9.57

PS: Persistent System, LR: Linear Regression, RF: Random Forest, SVR: Support Vector Regression, NN: Neural Network. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.5: Risultati dei modelli di base per tutti gli algoritmi, le tessiture, i piani irrigui e gli orizzonti

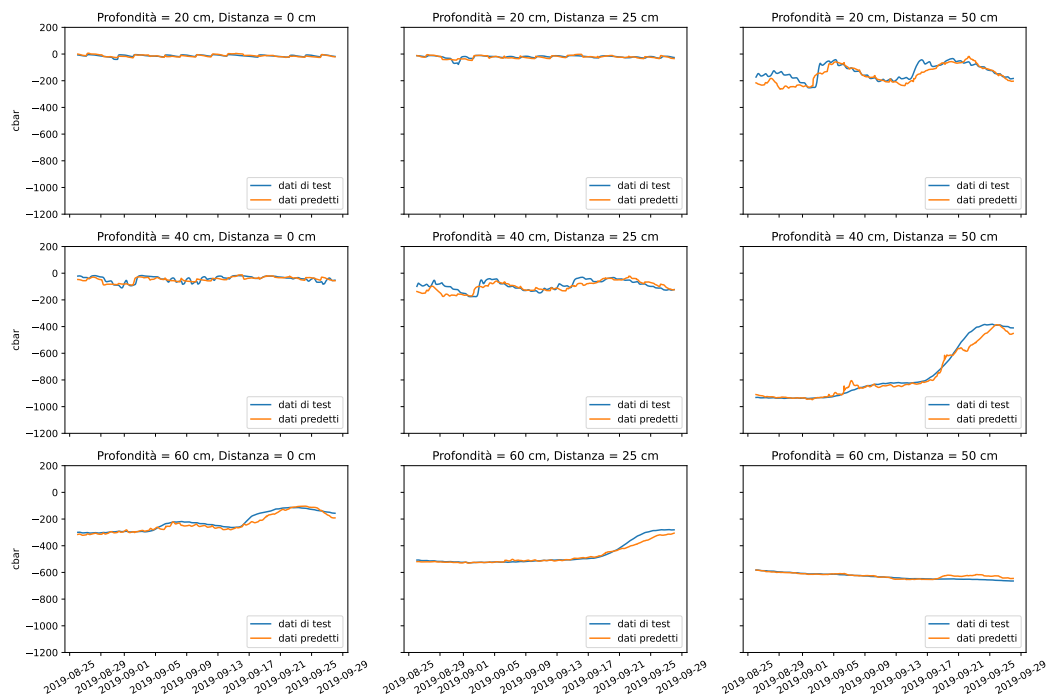


Figura 5.10: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore, relativamente alla tessitura franco argillosa, all'intervallo di irrigazione di 2 giorni e all'orizzonte di 4 giorni

In Figura 5.10 viene mostrato l'andamento della predizione delle reti neurali a confronto con i rispettivi dati osservati, relativamente alla tessitura franco argillosa (CLAYEY), all'intervallo di irrigazione di 2 giorni e all'orizzonte di 4 giorni.

5.2.2 Test dei modelli su scenari non conosciuti

In Tabella 5.6, Tabella 5.7 e Tabella 5.8 sono riportati i risultati degli esperimenti che hanno applicato il modello a una diversa tessitura e/o piano irriguo rispetto a quelli considerati durante l'addestramento, rispettivamente per ciascun orizzonte selezionato e facendo riferimento alle reti neurali, l'approccio risultato vincente.

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	18.80	35.39	42.91	28.35	28.78	40.34	32.93	41.56	50.56
	4	57.04	23.14	33.88	19.86	19.79	33.82	28.78	36.96	46.28
	7	153.59	45.06	19.27	35.30	29.17	23.41	28.09	41.69	42.70
SANDY	2	49.08	31.10	54.33	14.93	21.04	51.51	32.52	40.86	68.14
	4	11.68	44.60	51.03	111.68	15.99	29.53	43.26	46.75	51.58
	7	193.11	124.19	36.05	103.32	66.77	13.09	37.60	46.11	40.87
SILTY	2	195.36	123.97	70.49	105.55	74.88	44.89	5.14	12.12	13.54
	4	254.13	173.20	86.68	163.52	120.87	51.01	22.04	7.65	8.51
	7	254.08	188.13	113.69	186.72	155.03	50.57	25.32	16.73	5.62

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.6: Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 2 giorni

Rispetto ai precedenti risultati ottenuti su scenari conosciuti, in questo caso l'errore medio si assesta su valori più elevati, sia quando si passa a un differente piano irriguo, ma in maniera più evidente quando a cambiare è la tessitura. In particolare, passando dalla tessitura SILTY ad un'altra, l'errore è sensibilmente maggiore rispetto agli altri casi: ciò è dovuto al fatto che il comportamento di asciugatura di questo tipo di terreno si discosta in maniera più marcata rispetto agli altri due (vedi Figura 5.5, Figura 5.7 e Figura 5.8) e il suo intervallo di variazione di umidità è più ristretto, tale per cui un modello allenato su di esso non riesce a predire correttamente quando l'intervallo di variazione si allarga. Al contrario, spostandosi dalla tessitura SANDY o CLAYEY ad un'altra le prestazioni sono in generale migliori. Come osservato per i precedenti

esperimenti, anche in questo caso l'aumento dell'orizzonte della predizione tende ad influenzare in maniera negativa quest'ultima.

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	22.08	47.25	63.14	42.09	41.34	57.64	51.12	52.70	62.13
	4	110.01	25.15	41.30	29.12	25.66	43.14	40.35	43.08	57.40
	7	126.15	76.39	33.38	57.54	56.15	32.84	51.39	53.48	54.27
SANDY	2	61.29	42.57	60.77	20.46	25.96	53.93	48.41	52.26	62.29
	4	133.99	64.69	52.46	133.99	25.40	39.49	71.13	70.16	56.17
	7	227.94	130.70	46.68	100.17	63.02	23.25	53.41	53.38	51.33
SILTY	2	231.91	159.19	92.91	151.63	114.86	58.23	7.53	9.34	19.43
	4	185.76	141.81	84.23	118.57	88.75	56.66	9.68	8.27	24.73
	7	274.80	209.56	115.56	209.73	171.07	57.74	28.41	33.55	6.81

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.7: Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 4 giorni

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	25.57	60.52	89.47	50.15	44.67	78.79	53.11	58.71	67.94
	4	72.41	41.40	67.70	36.41	35.30	66.49	61.00	63.16	68.89
	7	149.05	82.50	40.84	67.45	63.19	41.62	46.76	47.42	50.26
SANDY	2	57.44	52.55	81.04	29.99	31.60	70.21	68.07	73.40	82.53
	4	143.37	73.01	54.20	143.37	29.96	53.50	61.08	65.02	72.07
	7	222.13	133.81	45.30	114.33	78.06	27.93	47.52	51.24	54.73
SILTY	2	228.30	151.33	84.10	139.96	103.18	57.59	12.33	15.59	22.89
	4	231.01	152.29	72.34	140.51	103.10	61.37	16.08	13.39	24.02
	7	179.75	122.15	88.32	99.54	73.01	51.25	16.18	17.78	9.57

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.8: Risultati dei modelli di reti neurali applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, per l'orizzonte di 7 giorni

Da tessitura franco argillosa a tessitura franco sabbiosa In questo caso, il modello generato per il terreno franco argilloso è stato applicato al terreno

franco sabbioso. In Tabella 5.9 sono riportati i risultati ottenuti considerando un intervallo di irrigazione di 7 giorni: il modello di base che sfrutta le reti ha buone prestazioni rispetto agli altri algoritmi e rappresenta il modello vincente, mentre Persistent System risulta essere il peggiore.

Orizzonte	Algoritmi				
	Persistent System	Linear Regression	Random Forest	SVR	Reti Neurali
2	60.07	29.09	39.87	40.49	23.41
4	85.08	36.31	53.01	49.30	32.84
7	101.44	50.82	69.28	58.68	41.62

Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.9: Risultati da tessitura franco argillosa a franco sabbiosa, con intervallo di irrigazione di 7 giorni

5.2.3 Adattamento dei modelli su scenari non conosciuti

Con tali esperimenti, i modelli di reti neurali generati in precedenza sono stati adattati in maniera progressiva ai nuovi scenari, fornendo loro in input i dati relativi a questi ultimi, da usare per ulteriori tre cicli di allenamento. In Tabella 5.10, Tabella 5.11 e Tabella 5.12 vengono mostrati i risultati ottenuti a seguito dei successivi cicli di adattamento delle reti, prendendo in considerazione l'orizzonte di 7 giorni.

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	64.79	72.25	102.62	101.14	55.88	67.73	34.20	46.52	67.86
	4	91.21	58.95	57.20	48.86	34.80	46.14	27.84	28.99	42.81
	7	122.67	63.09	57.06	46.79	50.74	38.98	31.85	27.31	29.26
SANDY	2	98.09	65.10	66.71	41.42	32.17	44.22	39.57	47.43	52.53
	4	105.20	63.55	50.26	105.20	41.88	42.05	32.39	32.21	31.45
	7	169.54	86.68	40.42	77.42	64.11	23.04	32.70	18.92	35.74
SILTY	2	185.64	139.90	63.79	101.31	85.39	31.91	14.83	14.66	15.86
	4	174.08	124.96	55.09	93.99	80.50	32.59	26.85	20.70	34.16
	7	173.43	115.61	69.53	97.67	70.89	39.31	16.21	16.64	12.76

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.10: Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il primo ciclo di adattamento

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	101.70	79.41	83.49	112.29	43.79	82.42	30.28	37.59	77.78
	4	67.52	63.22	58.15	64.80	29.99	39.88	24.48	22.89	41.47
	7	94.54	69.43	48.25	61.05	56.30	34.12	20.79	40.29	39.01
SANDY	2	57.02	60.35	58.21	46.86	37.09	50.54	31.27	35.89	50.86
	4	78.07	65.59	48.81	78.07	36.23	39.07	18.89	22.47	29.21
	7	82.41	82.94	38.53	67.09	60.68	32.55	15.51	31.06	37.72
SILTY	2	130.05	119.06	46.35	76.24	78.51	27.06	16.67	22.71	28.72
	4	115.47	84.17	39.79	67.22	70.14	36.68	23.09	16.60	36.33
	7	132.40	88.96	58.62	69.57	60.91	38.80	13.34	18.12	14.17

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.11: Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il secondo ciclo di adattamento

Tessiture e piani irrigui		Tessiture e piani irrigui								
		CLAYEY			SANDY			SILTY		
		2	4	7	2	4	7	2	4	7
CLAYEY	2	88.49	83.45	99.96	61.33	52.13	49.77	30.46	41.36	41.95
	4	62.06	45.74	75.25	34.27	26.32	34.50	14.85	20.17	29.76
	7	60.67	62.79	52.35	47.28	52.69	29.63	23.04	34.34	20.04
SANDY	2	58.59	53.28	58.04	31.30	40.12	44.36	23.67	20.90	32.37
	4	67.20	52.77	52.05	67.20	33.73	39.80	17.86	32.22	20.83
	7	66.75	63.70	40.74	50.50	42.42	24.81	16.59	49.37	13.34
SILTY	2	76.14	66.36	44.72	63.15	40.62	32.17	15.18	28.32	19.42
	4	68.52	66.35	37.13	55.80	47.60	31.34	20.00	13.53	25.69
	7	89.67	73.70	52.73	56.60	43.04	34.06	13.03	22.21	11.44

I valori sulla diagonale colorati in grigio si riferiscono agli esperimenti precedenti in cui lo scenario non cambia. Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.12: Risultati dei modelli di reti neurali, con orizzonte di 7 giorni, applicati a differenti tessiture e/o piani irrigui rispetto a quelli usati nell'addestramento, dopo il terzo ciclo di adattamento

Si osserva una generale diminuzione dell'errore all'avanzare dei cicli di adattamento, soprattutto per quanto riguarda le combinazioni più problematiche, evidenziate precedentemente, per le quali il valore di RMSE si riduce più della metà. Complessivamente, le predizioni migliori sono quelle relative ai modelli che sono stati testati sul terreno SILTY.

Da intervallo di 7 giorni a intervallo di 2 giorni con tessitura franco argillosa In questo caso, prima dell'adattamento Persistent System supera con netto vantaggio tutti gli altri algoritmi (Tabella 5.13). In seguito all'adattamento con le reti (Tabella 5.14), si assiste però ad un deciso e progressivo miglioramento, pur senza raggiungere le performance di Persistent System, che potrebbero essere superate aggiungendo ulteriori cicli di addestramento.

Orizzonte	Algoritmi				
	Persistent System	Linear Regression	Random Forest	SVR	Reti Neurali
2	29.67	122.98	153.97	198.08	153.59
4	45.73	159.47	167.71	204.37	126.15
7	58.99	289.42	168.37	196.35	149.05

Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.13: Risultati da intervallo di irrigazione di 7 giorni a 2 giorni, con tessitura franco argillosa

Orizzonte	Reti Neurali			
	Modello base	Primo ciclo di adattamento	Secondo ciclo di adattamento	Terzo ciclo di adattamento
2	153.59	191.89	100.84	44.62
4	126.15	136.68	90.72	63.97
7	149.05	122.67	94.54	60.67

Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.14: Risultati adattamento reti da intervallo di irrigazione di 7 giorni a 2 giorni, con tessitura franco argillosa

In Figura 5.11, Figura 5.12, Figura 5.13 e Figura 5.14 vengono riportati i grafici dell'andamento della predizione a 2 giorni di distanza per ciascun sensore, partendo dal modello base fino ad arrivare all'ultimo ciclo di adattamento.

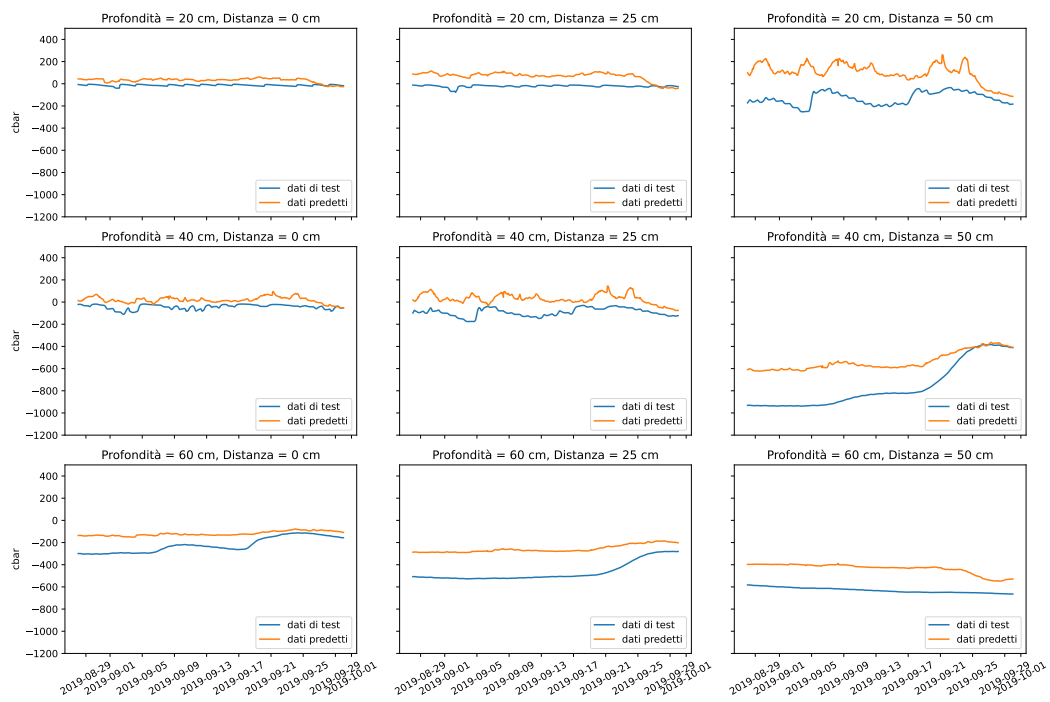


Figura 5.11: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore con il modello base, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni

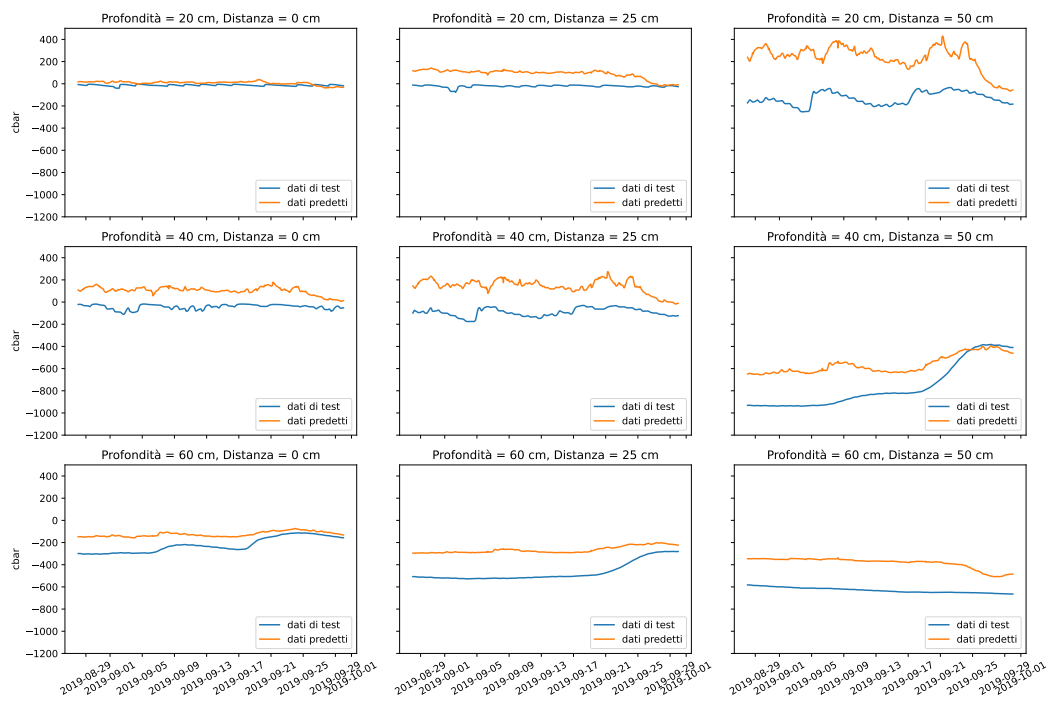


Figura 5.12: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il primo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni

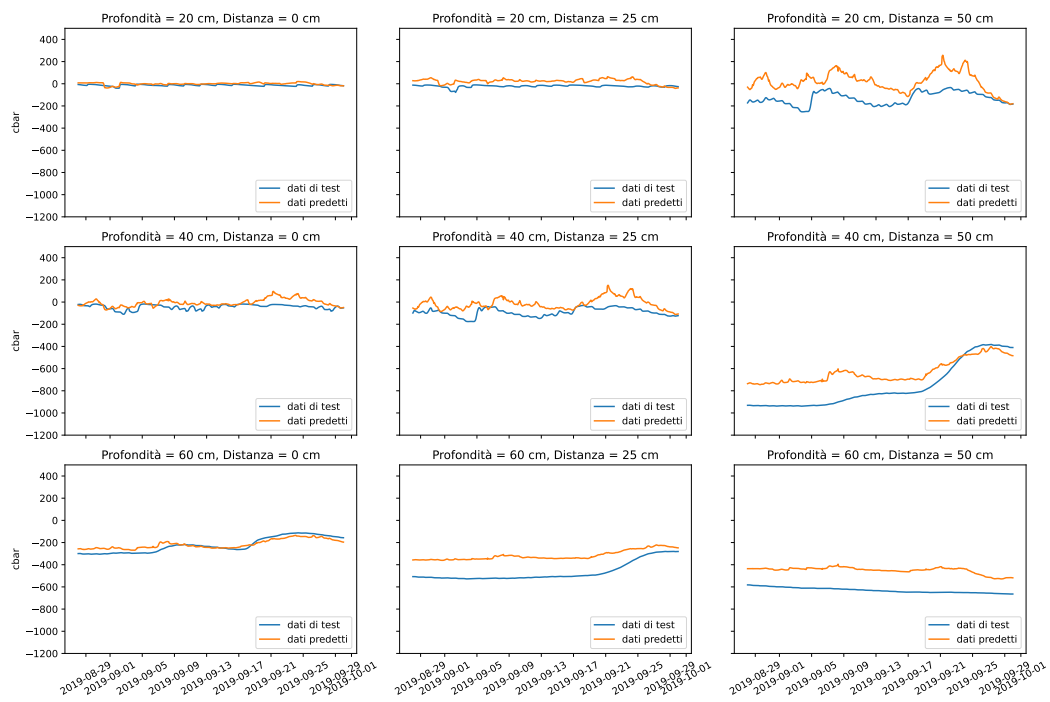


Figura 5.13: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il secondo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni

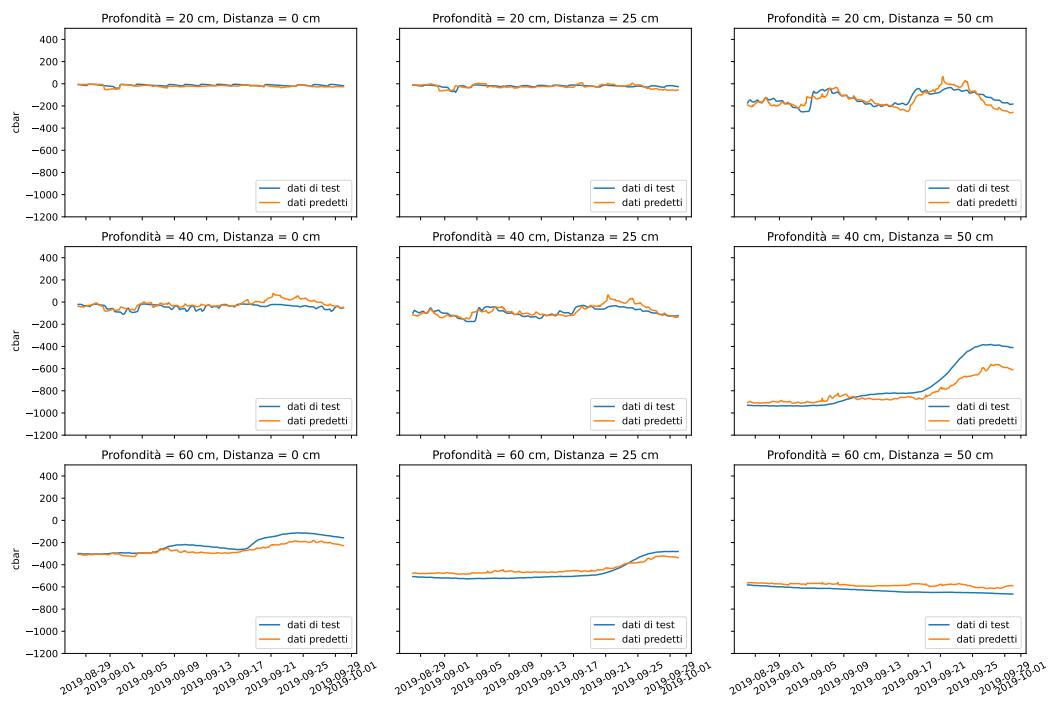


Figura 5.14: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il terzo ciclo di adattamento, passando dall'intervallo di 7 giorni a 2 giorni, con tessitura franco argillosa e orizzonte di 2 giorni

Da tessitura franco argillosa a tessitura franco limosa In questo caso, la tessitura usata nel processo di adattamento è quella franco limosa, con un comportamento di essiccazione che si discosta notevolmente da quella di partenza. I risultati corrispondenti, riportati in Tabella 5.15 e ottenuti fissando un piano con intervallo di irrigazione pari a 7 giorni, testimoniano un progressivo miglioramento dopo ogni ciclo. In Figura 5.15 sono riportati i grafici delle predizioni a 7 giorni ottenute dai vari sensori.

Reti Neurali				
Orizzonte	Modello base	Primo ciclo di adattamento	Secondo ciclo di adattamento	Terzo ciclo di adattamento
2	42.70	23.93	17.39	13.02
4	54.27	35.39	35.58	24.16
7	50.26	29.26	39.01	20.04

Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.15: Risultati adattamento reti da tessitura franco argillosa a franco limosa, con intervallo di irrigazione di 7 giorni

Da tessitura franco argillosa con intervallo di 7 giorni a tessitura franco limosa con intervallo di 2 giorni Anche in questa situazione, variando sia la tessitura che il piano irriguo, procedendo con gli adattamenti si osserva un trend di miglioramento progressivo (Tabella 5.16).

Reti Neurali				
Orizzonte	Modello base	Primo ciclo di adattamento	Secondo ciclo di adattamento	Terzo ciclo di adattamento
2	28.09	17.10	15.13	13.53
4	51.39	41.45	31.66	22.23
7	46.76	31.85	20.79	23.04

Per ogni scenario, il risultato migliore è riportato in grassetto.

Tabella 5.16: Risultati adattamento reti da tessitura franco argillosa con intervallo di irrigazione di 7 giorni a tessitura franco limosa con intervallo di irrigazione di 2 giorni

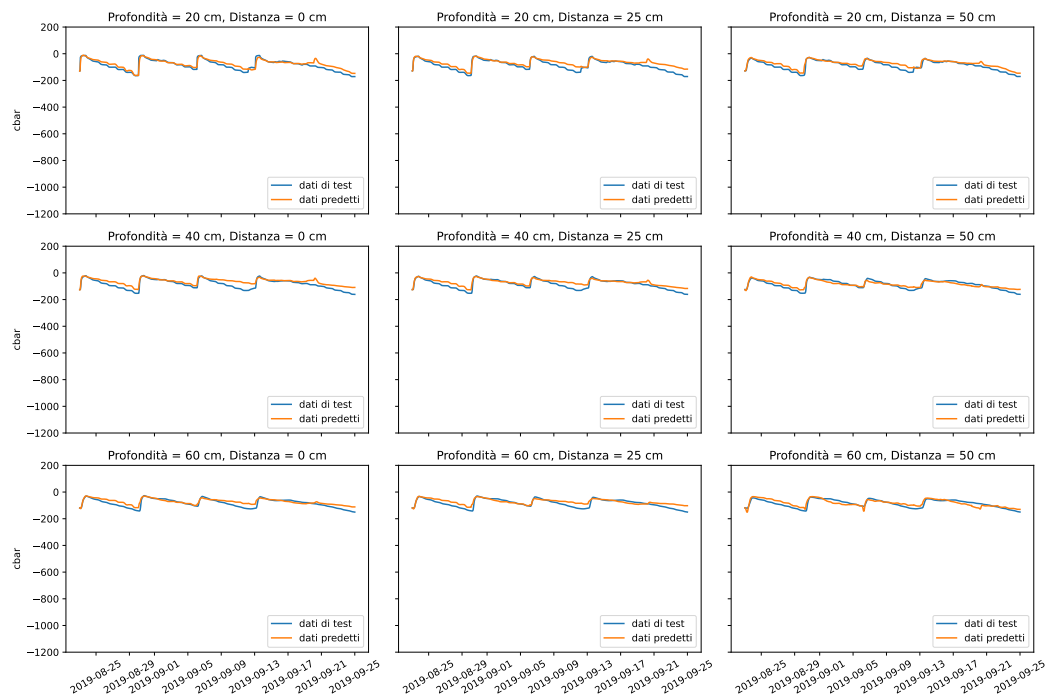


Figura 5.15: Andamento della SM osservata e predetta dalle reti neurali per ciascun sensore dopo il terzo ciclo di adattamento, passando dalla tessitura franco argillosa a quella franco limosa, con intervallo di irrigazione di 7 giorni e orizzonte di 7 giorni

Capitolo 6

Conclusioni e sviluppi futuri

L'applicazione di strategie di agricoltura di precisione è oggi di fondamentale importanza per fronteggiare le sempre più frequenti situazioni climatiche estreme, che hanno portato all'innalzamento della temperatura a livello globale con conseguenti episodi di siccità. In questo contesto, il lavoro qui presentato può fungere da punto di partenza per una soluzione di efficientamento agricolo, capace di ridurre gli sprechi di acqua tramite una gestione più oculata e mirata dell'irrigazione. Prendendo come riferimento un reale frutteto di kiwi e il relativo impianto sensoristico ed irriguo situato ad Errano, una località della regione Emilia-Romagna della provincia di Ravenna, il sistema proposto è in grado di predire i valori futuri dell'umidità del suolo. Tali risultati possono quindi essere sfruttati in seguito per implementare un sistema di *smart irrigation*, che sia in grado di risparmiare acqua rispetto al quantitativo che verrebbe fornito dal contadino, ma che allo stesso tempo garantisca la qualità del frutto e di conseguenza la buona riuscita del raccolto.

Attualmente, non sono stati condotti lavori di predizione dell'umidità del suolo che abbiano avuto come oggetto di studio una piantagione di kiwi. L'indagine di Grisafi *et al.* [74] mostra infatti come nel corso degli anni solo alcuni lavori ([75, 76, 77, 78, 79]) si siano concentrati in modo specifico sui campi di kiwi, ma questi si sono soffermati più che altro sulla stima del quantitativo di carbonio, non sulle necessità idriche. L'unico lavoro che si avvicina in tal senso è quello di Silva *et al.* [80], che ha analizzato la traspirazione di una coltivazione di kiwi.

Il punto di forza dell'impianto considerato è rappresentato dalla griglia bi-dimensionale di sensori installati direttamente nel sottosuolo, in grado di misurare in maniera puntuale il contenuto idrico del terreno a diverse profondità e distanze rispetto al punto in cui si trova la fonte di irrigazione. La maggior parte dei lavori noti in letteratura, infatti, si affida a un singolo sensore o al più a griglie mono-dimensionali, che si limitano a registrare i valori di umidità

in un punto specifico, a più profondità.

Un altro elemento fondamentale è stato l'uso del simulatore di flussi idrologici Orchard3D-Lab. Essendo basato sulla fisica, e quindi in grado di modellare con precisione il comportamento di un terreno sulla base di equazioni ben note nel settore, ha permesso di generare in maniera precisa e in tempi relativamente rapidi dei dati simulati di umidità del suolo e di irrigazione. Così facendo, è stato possibile ottenere la quantità di dati necessaria ai processi di addestramento dei modelli di Machine Learning considerati, senza dovere aspettare la conclusione delle varie campagne produttive e quindi l'accumularsi dei rispettivi dati reali.

Il contributo principale di questo lavoro è quello di fornire un meccanismo che consenta il rapido e progressivo adattamento di un modello di previsione dell'umidità del suolo precedentemente ottenuto, quando variano determinate condizioni. Per farlo, ci si è affidati agli approcci basati su reti neurali oramai diffusi in questo e altri contesti. Nello specifico, i risultati hanno evidenziato che è possibile adattare velocemente un modello a un suolo con caratteristiche fisiche o che segue schemi di irrigazione differenti rispetto a quelli di partenza. Ciò rende fattibile distribuire fin da subito il modello su un diverso campo, risparmiando tempo e senza aspettare di avere a disposizione la grande mole di dati storici, specifica del nuovo sito o del nuovo schema irriguo, necessaria per creare da zero il modello.

Gli sviluppi futuri potrebbero riguardare l'introduzione dei dati reali nella fase di testing del sistema, in modo da avere già un effettivo riscontro sulla situazione reale del campo di riferimento. Per quanto riguarda la modellazione e la risoluzione del problema, si potrebbero utilizzare tecniche di Deep Learning più avanzate, come gli approcci basati su reti LSTM, anche se per la loro applicazione sarebbe necessaria una quantità di dati ancora più ingente. In questo caso, cercare di ridurre le variabili di input potrebbe mitigare in parte la complessità del problema, sebbene tutte le variabili meteorologiche, di irrigazione e di umidità del suolo utilizzate contribuiscano in qualche modo alla predizione finale. Da ultimo, è possibile esplorare tecniche più avanzate di *continual learning*, ma che sono anche più acerbe in un contesto di regressione e di previsione di serie temporali, che permettono di adattare il modello variandone gli iper-parametri e trasformandone anche l'architettura.

Bibliografia

- [1] V. Filipović, G. Ondrasek, and L. Filipović. Modelling Water Dynamics, Transport Processes and Biogeochemical Reactions in Soil Vadose Zone. 2016.
- [2] E. Brochu, V. Cora, and N. Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. 2010.
- [3] Precision agriculture, 2015. URL: <https://www.agricolus.com/tecnologie/agricoltura-di-precisione/>. Accessed: August, 2023.
- [4] OECD/FAO. OECD-FAO Agricultural Outlook 2022-2031. *OECD Publishing*, 2022.
- [5] Trends in precision agriculture publications over time, 2023. URL: https://app.dimensions.ai/discover/publication?search_mode=content&search_text=precision%20agriculture&search_type=kws&search_field=full_search. Accessed: August, 2023.
- [6] Agro.Big.Data.Science, 2020. URL: <http://agrobigdatascience.it>. Accessed: August, 2023.
- [7] Top 10 Country Production of Kiwi fruit, 2023. URL: https://www.fao.org/faostat/en/#rankings/countries_by_commodity. Accessed: August, 2023.
- [8] ISTAT data on the cultivation and production of kiwifruit in Italy and in the Emilia-Romagna region, 2023. URL: <http://dati.istat.it/Index.aspx?QueryId=33705>. Accessed: August, 2023.
- [9] Kiwi cultivation, 2016. URL: <https://www.coltivazionebiologica.it/coltivazione-del-kiwi-semina-potatura/>. Accessed: August, 2023.
- [10] Water management strategies in kiwi orchards, 2023. URL: <https://rivistafrutticoltura.edagricole.it/coltura/kiwi/kiwi-una-coltura-sempre-piu-meridionale/>. Accessed: August, 2023.

-
- [11] Kiwi irrigation and fertilization systems, 2019. URL: <https://www.bonattiirrigazioni.it/it/culture/kiwi/>. Accessed: August, 2023.
 - [12] Kiwi irrigation systems, 2019. URL: <https://www.agricolplast.it/impianto-irrigazione-kiwi/>. Accessed: August, 2023.
 - [13] FAO. Soils and biodiversity. Soils host a quarter of our planet's biodiversity. *FAO Publishing*, 2015.
 - [14] Soil texture. URL: http://agricoltura.regione.campania.it/pedologia/pdf/tessitura_del_suolo.pdf. Accessed: August, 2023.
 - [15] R. K. Rai and V. P. Singh and A. Upadhyay. *Planning and Evaluation of Irrigation Projects. Methods and Implementation*. Academic Press, 2017.
 - [16] M. Bittelli. Measuring Soil Water Potential for Water Management in Agriculture: A Review. *Sustainability*, 2(5):1226–1251, 2010.
 - [17] Precision agriculture, 2021. URL: <https://www.mccormick.it/agricoltura-di-precisione/>. Accessed: August, 2023.
 - [18] P. Velusamy, S. Rajendran, R. K. Mahendran, S. Naseer, M. Shafiq, and J. Choi. Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges. *Energies*, 15:217, 2021.
 - [19] M. Francia, J. Giovanelli, and M. Golfarelli. Multi-sensor profiling for precision soil-moisture monitoring. *Computers and Electronics in Agriculture*, 197, 2022.
 - [20] J. J. Šimůnek, M. Šejna, H. Saito, M. Sakai, and M. T. Van Genuchten. The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media. Version 4.08. 2009.
 - [21] J. J. Šimůnek, M. Šejna, and M. T. Van Genuchten. The HYDRUS-2D Software Package for Simulating Water Flow and Solute Transport in Two Dimensional Variably Saturated Media. Version 2.0. 1998.
 - [22] M. Šejna and J. J. Šimůnek. HYDRUS (2D/3D): Graphical User Interface for the HYDRUS Software Package Simulating Two- and Three-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media. 2007.

- [23] H. Li, J. Yi, J. Zhang, Y. Zhao, B. Si, R. L. Hill, L. Cui, and X. Liu. Modeling of Soil Water and Salt Dynamics and Its Effects on Root Water Uptake in Heihe Arid Wetland, Gansu, China. *Water*, 7(5):2382–2401, 2015.
- [24] A.J. Zapata-Sierra, J. Roldán-Cañas, R. Reyes-Requena, and M.F. Moreno-Pérez. Study of the Wet Bulb in Stratified Soils (Sand-Covered Soil) in Intensive Greenhouse Agriculture under Drip Irrigation by Calibrating the Hydrus-3D Model. *Water*, 13(5), 2021.
- [25] L. A. Richards. Capillary conduction of liquids through porous mediums. *Physics*, 1(5):318–333, 1931.
- [26] Richards equation. URL: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/richards-equation>. Accessed: August, 2023.
- [27] Penman-Monteith equation. URL: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/penman-monteith-equation>. Accessed: August, 2023.
- [28] AquaCrop, 2009. URL: <https://www.fao.org/aquacrop/en/>. Accessed: August, 2023.
- [29] M. Bittelli, G. S. Campbell, and F. Tomei. *Soil Physics with Python: Transport in the Soil-Plant-Atmosphere System*. Oxford University Press, 2015.
- [30] M. Bittelli, F. Tomei, A. Pistocchi, M. Flury, J. Boll, E. S. Brooks, and G. Antolini. Development and testing of a physically based, three-dimensional model of surface and subsurface hydrology. *Advances in Water Resources*, 33(1):106–122, 2009.
- [31] CRITERIA - Model of water balance and crop development, 2020. URL: <https://www.arpae.it/it/temi-ambientali/meteo/scopri-di-piu/strumenti-di-modellistica/criteria>. Accessed: August, 2023.
- [32] E. V. Shein, A. Y. Mady, and L. I. Il'in. Validation of HYDRUS-1D for Predicting of Soil Moisture Content with Hysteresis Effect. *Biogeosystem Technique*, 6(1):59–64, 2019.
- [33] A. Basak, O. Mengshoel, K. Schmidt, and C. Kulkarni. Wetting and Drying of Soil: From Data to Understandable Models for Prediction. *Proceedings*

- of 5th International Conference on Data Science and Advanced Analytics, pages 303–312, 2018.
- [34] F. Karandish and J. J. Šimůnek. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology*, 543:892–909, 2016.
- [35] A. El Mezouari and M. Najib. A hadoop based framework for soil parameters prediction. *Proceedings of 15th International Conference on Signal Image Technology and Internet Based Systems*, pages 681–687, 2019.
- [36] A. Khanna and S. Kaur. Evaluation of Soil Moisture for Estimation of Irrigation Pattern by Using Machine Learning Methods. *Communications in Computer and Information Science*, 1441:343–352, 2021.
- [37] M. Gill, T. Asefa, M. Kemblowski, and M. McKee. Soil Moisture Prediction Using Support Vector Machines. *Journal of the American Water Resources Association*, 42:1033–1046, 2006.
- [38] Z. Hong, Z. Kalbarczyk, and R. K. Iyer. A Data-Driven Approach to Soil Moisture Collection and Prediction. *2016 IEEE International Conference on Smart Computing*, pages 1–6, 2016.
- [39] A. Goap, D. Sharma, A.K. Shukla, and C. Rama Krishna. An IoT based smart irrigation management system using Machine learning and open source technologies. *Computers and Electronics in Agriculture*, 155:41–49, 2018.
- [40] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- [41] A. Dubois, F. Teytaud, and S. Verel. Short term soil moisture forecasts for potato crop farming: A machine learning approach. *Computers and Electronics in Agriculture*, 180, 2021.
- [42] Y. Liu, L. Mei, and S. K. Ooi. Prediction of soil moisture based on Extreme Learning Machine for an apple orchard. *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, pages 400–404, 2014.
- [43] C. Kang, G. Diverres, M. Karkee, Q. Zhang, and M. Keller. Decision-support system for precision regulated deficit irrigation management for wine grapes. *Computers and Electronics in Agriculture*, 208, 2023.

- [44] A. Risheh, A. Jalili, and E. Nazerfard. Smart Irrigation IoT Solution using Transfer Learning for Neural Networks. *2020 10th International Conference on Computer and Knowledge Engineering*, pages 342–349, 2020.
- [45] Recurrent Neural Networks (RNN) - The Vanishing Gradient Problem, 2018. URL: <https://www.superdatascience.com/blogs/recurrent-neural-networks-rnn-the-vanishing-gradient-problem>. Accessed: August, 2023.
- [46] P. Gao, H. Qiu, Y. Lan, W. Wang, W. Chen, X. Han, and J. Lu. . *Agriculture*, 12(1), 2022.
- [47] O. Adeyemi, I. Grove, S. Peets, Y. Domun, and T. Norton. . *Sensors*, 18(10), 2018.
- [48] P. K. Kashyap, S. Kumar, A. Jaiswal, M. Prasad, and A. H. Gandomi. Towards Precision Agriculture: IoT-Enabled Intelligent Irrigation Systems Using Deep Learning Neural Network. *IEEE Sensors Journal*, 21(16):17479–17491, 2021.
- [49] A. F. Jiménez, B. V. Ortiz, L. Bondesan, G. Morata, and D. Damianidis. Evaluation of Two Recurrent Neural Network Methods for Prediction of Irrigation Rate and Timing. *Transactions of the ASABE*, 63(5):1327—1348, 2020.
- [50] P. Datta and S. A. Faroughi. A multihead LSTM technique for prognostic prediction of soil moisture. *Geoderma*, 433, 2023.
- [51] J. Yu, S. Tang, L. Zhangzhong, W. Zheng, L. Wang, A. Wong, and L. Xu. A Deep Learning Approach for Multi-Depth Soil Water Content Prediction in Summer Maize Growth Period. *IEEE Access*, 8, 2020.
- [52] Q. Li, Z. Li, W. Shangguan, X. Wang, L. Li, and F. Yu. Improving soil moisture prediction using a novel encoder-decoder model with residual learning. *Computers and Electronics in Agriculture*, 195, 2022.
- [53] S. Jiang, G. Chen, D. Chen, and T. Chen. Application and Evaluation of an Improved LSTM Model in the Soil Moisture Prediction of Southeast Chinese Tobacco-Producing Areas. *Journal of the Indian Society of Remote Sensing*, 2021.
- [54] Q. Li, Y. Zhu, W. Shangguan, X. Wang, L. Li, and F. Yu. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma*, 409, 2022.

- [55] T. Zhou, Y. He, L. Luo, and S. Ji. Research on Soil Moisture Prediction Based on LSTM-Transformer Model. *Communications in Computer and Information Science*, 1801:329–342, 2023.
- [56] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 440–778, 2016.
- [57] How Transformers Work, 2010. URL: <https://towardsdatascience.com/transformers-141e32e69591>. Accessed: August, 2023.
- [58] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, J. Thomas S. Coors, T. Ullmann, M. Becker, A.-L. Boulesteix, D. Deng, and M. Lindauer. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. 2021.
- [59] R. El Shawi, M. Maher, and S. Sakr. Automated Machine Learning: State-of-The-Art and Open Challenges. 2019.
- [60] F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential Model-Based Optimization for General Algorithm Configuration. *Proceedings of the 5th International Conference on Learning and Intelligent Optimization*, pages 507–523, 2011.
- [61] M. Ebdn. Gaussian Processes: A Quick Introduction. 2015.
- [62] S. Watanabe. Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance. 2023.
- [63] C. Wang, Q. Wu, M. Weimer, and E. Zhu. FLAML: A Fast and Lightweight AutoML Library. 2019.
- [64] FLAML - A Fast Library for Automated Machine Learning & Tuning, 2021. URL: <https://microsoft.github.io/FLAML/>. Accessed: August, 2023.
- [65] T. J. Marshall, J. W. Holmes, and C. W. Rose. *Soil Physics*. Cambridge University Press, 1996.
- [66] O. Ippisch, H.-J. Vogel, and P. Bastian. Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation. *Advances in Water Resources*, 29(12):1780–1789, 2006.

-
- [67] P. I. Frazier. A Tutorial on Bayesian Optimization. *ArXiv*, 1807.02811, 2018.
- [68] G. Luo, B. L. Stone, M. D. Johnson, P. Tarczy-Hornoch, A. B. Wilcox, S. D. Mooney, X. Sheng, P. J. Haug, and F. L. Nkoy. Automating Construction of Machine Learning Models With Clinical Big Data: Proposal Rationale and Methods. *JMIR Research Protocols*, 6(8), 2017.
- [69] E. Zivot and J. Wang. Rolling Analysis of Time Series. In *Modeling Financial Time Series with S-Plus*, pages 299–346. Springer New York, 2003.
- [70] Dext3r - application for extracting weather data recorded by ARPAE. URL: <https://simc.arpae.it/dext3r/>. Accessed: September, 2022.
- [71] scikit-learn - Machine Learning in Python. URL: <https://scikit-learn.org/stable/index.html> Accessed: August, 2022.
- [72] Keras, 2015. URL: <https://keras.io> Accessed: August, 2022.
- [73] TensorFlow, 2015. URL: <https://www.tensorflow.org/?hl=en> Accessed: August, 2022.
- [74] F. Grisafi, T. M. DeJong, and S. Tombesi. Fruit tree crop models: an update. *Tree Physiology*, 42(3):441–457, 2022.
- [75] J.G Buwalda, E.F. Walton, and G.S. Smith. A carbon balance model for the kiwifruit vine. *Acta Horticulturae*, pages 87–96, 1990.
- [76] D.H. Greer, A.N. Seleznyova, and S.R. Green. From controlled environments to field simulations: leaf area dynamics and photosynthesis of kiwifruit vines (*Actinidia deliciosa*). *Functional Plant Biology*, 31(2):169–179, 2004.
- [77] M. Cieslak, A.N. Seleznyova, and J. Hanan. Virtual kiwifruit: modelling annual growth cycle and light distribution. *Proceedings of the 5th international workshop on functional-structural plant models*, pages 41–46, 2007.
- [78] M. Cieslak, A.N. Seleznyova, and J. Hanan. A functional-structural kiwifruit vine model. *Proceedings of Plant Growth Modeling, Simulation, Visualization and Applications*, pages 206–213, 2009.
- [79] M. Cieslak, A.N. Seleznyova, and J. Hanan. A functional structural kiwifruit vine model integrating architecture, carbon dynamics and effects of the environment. *Annals of Botany*, 107(5):747–764, 2011.

- [80] R. Silva, T. Paço, M. Ferreira, and M. Oliveira. Transpiration of a kiwifruit orchard estimated using the granier sap flow method calibrated under field conditions. *Acta Horticulturae*, 792:593–600, 2008.