Flusso di lavoro basato su Python per l'elaborazione automatizzata e il controllo di qualità dei dati ottenuti mediante sensori di crescita dei frutti

F. Pinerolo de Septis[1*], A. Boini[1], I. Ansaloni[1], E. Tarpinelli[1], C. Franceschini[1], G. Perulli[1], L. Manfrini[1]

[1] Università di Bologna,
Dipartimento di scienze e tecnologie agro-alimentari

[*] felix.pinerolode2@unibo.it

Gruppo di Lavoro in Sensoristica ed Intelligenza Artificiale in Ortoflorofrutticoltura

II Giornate Tecniche SOI
Bari 10-11 settembre 2026

1. Introduzione:

Fruttometro: la crescita di un frutto in diretta



Il fruttometro (fruit gauge) è un **sensore** da applicare sul **frutto** che misura ad **alta frequenza** ed in **continuo** le variazioni del **diametro** durante il giorno e la stagione.
È composto da un **potenziometro lineare** il cui output, attraverso appropriate conversioni, viene **elaborato** prima in **diametro** e poi **indicatori** di **velocità** di **crescita** (**AGR/RGR**).
L'elaborazione di questi dati permette di osservare il comportamento dei frutti in **risposta** a **stimoli ambientali** e fornendo importanti dati sullo **stato fisiologico** della **pianta**.



1. Introduzione: Flussi vascolari

I **fruttometri, oltre** a misurare **frutti attaccati**, possono osservare la **variazione** del **diametro** di **frutti anulati** e **distaccati**, i **primi** rimangono attaccati alla pianta ma **viene reciso il floema** tramite **anulatura** e **defogliati**, i **secondi** vengono completamente **separati** dalla **pianta**. Tramite le **misure** dei diametri di queste **3 tipologie** di frutto è possibile **stimare** in modo differenziale i **flussi vascolari** che influiscono sul frutto e la sua crescita ed effettuare una **comparazione** tra i diversi **trattamenti**.

Le equazioni prevedono l'uso di RGR o AGR per tipo di frutto

Floema: Attaccato – Anulato

Xilema: Anulato –Distaccato

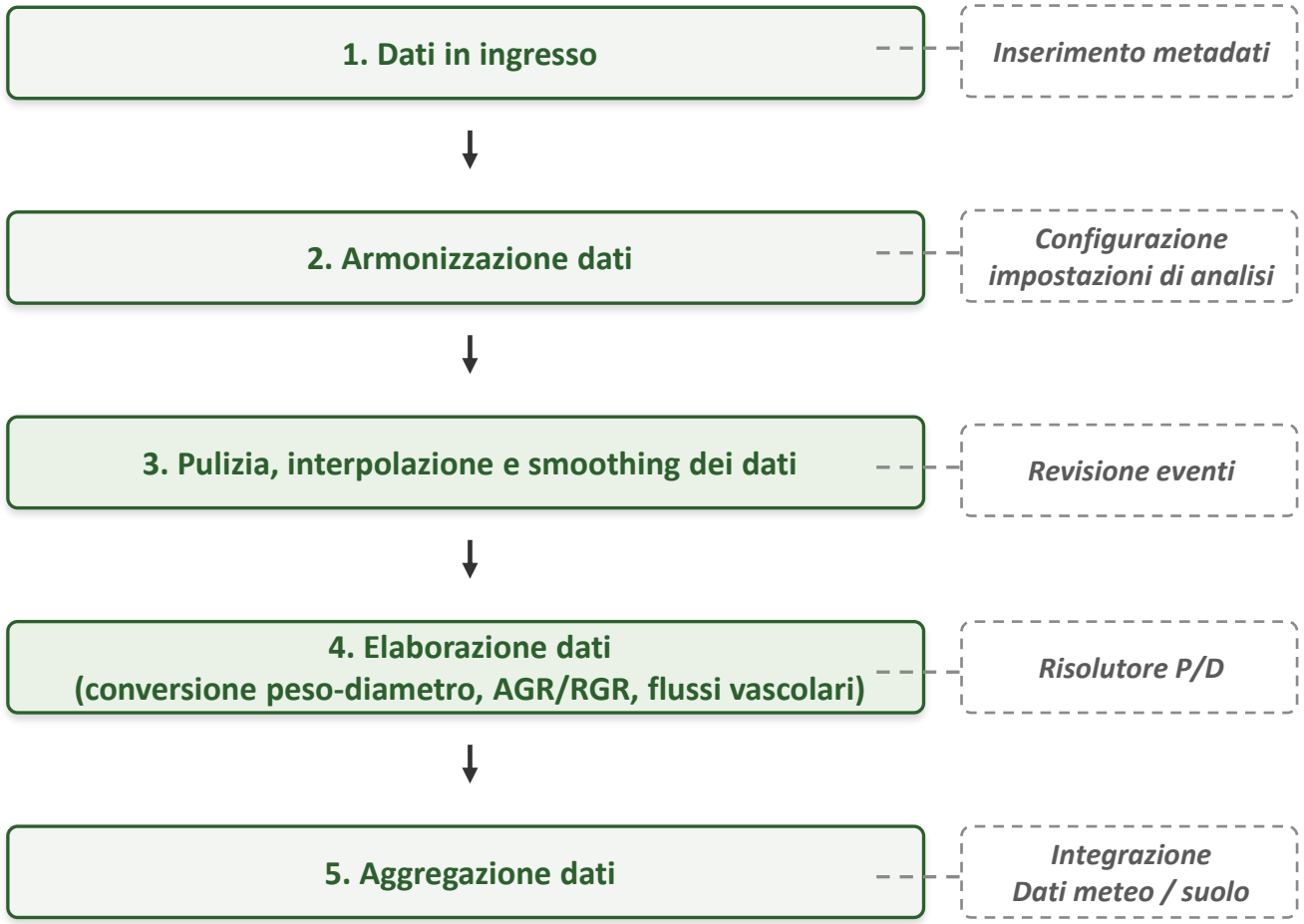
Traspirazione: Distaccato



Anulatura del floema

2. Materiali e metodi

Passaggi della pipeline da raw a AGR e flussi vascolari



La **pipeline** segue una **sequenza** vincolata di **14 fasi:** dalla **validazione** degli input, al **parsing** dei file raw, **compilazione** dei **metadati** utili per chiarire le unità sperimentali, alla **pulizia** del **segnale** (fisica o statistica), alla **costruzione** di una **griglia temporale** fissa, **interpolazione** e **smoothing** fino al calcolo di **AGR/RGR** e all'**aggregazione** finale con i proxy di **flussi vascolari** (sap-flux).
Ogni fase separa evidenza non distruttiva da operazioni che modificano la serie garantendo un alta tracciabilità.

2. Materiali e metodi

Struttura file e conversione peso-diametro

Sono richiesti i seguenti file:

- **CSV** con i **dati originali** dei nodi IoT-sensore,
- **Diario di campo** (Trial Diary) che definisce le finestre temporali (Round) di osservazioni e le CV
- **Diametri iniziali** di reset per ciascun sensore-round.
- **Parametri di conversione Peso-Diametro**.

File di input	Funzionalità	Output finale	Tipo
Dati Fruttometro (Fruit Gauges)	Contiene i dati grezzi del sensore	Tassi di crescita e flussi linfatici	**Obbligatorio**
Diario di campo (Trial diary)	Registra le finestre temporali dell'esperimento	Calendario dei Round	**Obbligatorio**
Diametri iniziali	Metadati e diametro iniziale/reset di ogni Round	Codici univoci, abilitazione elaborazione dati	**Obbligatorio**
Dati di conversione peso (W/D)	Stima dei parametri di conversione peso-diametro	Coefficienti di conversione specifici per trattamento	**Obbligatorio**
Dati meteo	Monitoraggio macro e microclimatico	Deficit di pressione di vapore ed evapotraspirazione	**Ancillare**
Dati di umidità del suolo	Monitoraggio del suolo multi-livello	Contenuto idrico del suolo	**Ancillare**

La **conversione** da **diametro** a **peso**/sostanza secca (W/D) usa una **regressione log-log** (**$W = \alpha \times D^{\beta}$**), con **parametri** risolti **per specie, cultivar, zona, anno e trattamento** tramite un **registro** aggiornabile su cui possono essere **caricati dati specifici** per trattamento legati alla **prova sperimentale o accedere ad un database** su cui sono **registrati** i **parametri di alcune CV**.

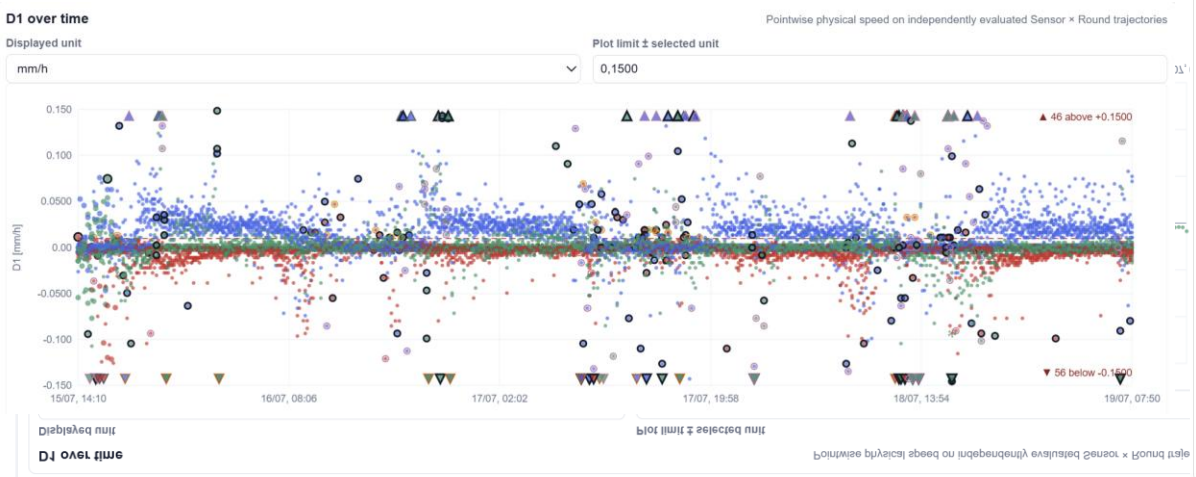
Gruppo di Lavoro in Sensoristica ed Intelligenza Artificiale in Ortoflorofrutticoltura

2. Materiali e metodi

Pulizia dati e interpolazione dati

I **dati grezzi** vengono **puliti** con **strategie configurabili**, integrate da **statistici**, protezione degli **spike** isolati e dei vicini.
Ogni cella registra con un **flag** la presenza di uno o più **elementi** concorrenti alla **pulizia**.
Le **lacune** brevi interne vengono **colmate** per **interpolazione** lineare o PCHIP.

Per la stesura di **questo abstract** sono state **utilizzate** e **comparate diverse tipologie** di **pulizia** incentrate principalmente sulla **velocità** e **l'accelerazione** dell'output.





2. Materiali e metodi

Registrazione eventi e azioni

Ogni **sensore/round** viene analizzato per **rilevare eventi di integrità del segnale**

Gli **eventi** sono divisi tra **locali** ed **estesi**, e hanno delle **caratteristiche ben definite che ne caratterizzano il livello di affidabilità (0-1)** su cui si **basano decisioni** autonome di **flag**, **revisione dell'operatore** ed **esclusione automatica**.

Un **evento non modifica** direttamente i **dati**: per ogniuno di essi l'operatore decide se mantenerlo come flag, se correggere l'offset con diverse opzioni o escludere il segmento, garantendo tracciabilità completa delle azioni sui dati.

Evento rilevato	Descrizione	Flag/strumenti che concorrono
Spike	Variazione isolata anomala	Pulizia fisica, Rolling MAD, protezione vicini
Clipping	Segnale bloccato al limite di saturazione	Range/endpoint protection
Intermittenza	Interruzioni ripetute del segnale	Rilevamento integrità segnale
Assestamento iniziale	Instabilità nella fase iniziale dopo installazione	Rilevamento integrità segnale
Salto di livello (level shift)	Offset improvviso tra segmenti	Rilevamento integrità segnale, revisione evento
Perdita del frutto	Interruzione permanente per caduta/distacco del frutto	Rilevamento integrità segnale
Deriva lenta (drift)	Tendenza lenta non biologica nel tempo	Rilevamento integrità segnale
Anomalia su AGR/RGR	Valore di crescita implausibile dopo la ricostruzione	Consenso tra ≥2 strumenti AGR (IQR, Z-score, limiti fisici, Rolling MAD, Bridge) o 1 flag AGR confermato da un endpoint WD

Obiettivo dello studio e procedura di validazione

L'obiettivo dello studio è stato **sviluppare** e **valutare** un flusso di lavoro Python specifico per i fruttometri, **capace** di **trasformare** le misure grezze in **diametro**, **peso** (W), **AGR**, **RGR** e proxy di **flusso vascolare**, mantenendo la **tracciabilità completa** delle elaborazioni.

Per **verificarne** la **robustezza** sono state eseguite **21 run controllate su 3 dataset** secondo un disegno ad "**ablation**":
una **run di riferimento (R0)** e **sei varianti** che modificano una sola componente alla volta

Run	Pulizia ΔmV	Azione integrità	Riempimento gap	Denoising	Effetto testato
R0	Pulizia fisica	Flag	Lineare	Off	Riferimento
A	Pulizia stocastica	Flag	Lineare	Off	Pulizia adattiva
B	Pulizia fisica	Rimozione	Lineare	Off	Rimozione basata su integrità
C-off	Pulizia fisica	Flag	Off	Off	Effetto del riempimento gap
C-PCHIP	Pulizia fisica	Flag	PCHIP	Off	Interpolazione polinomiale
D	Pulizia fisica	Flag	Lineare	Savitzky-Golay	Denoising
STACK	Pulizia fisica	Rimozione	PCHIP	Savitzky-Golay	Effetto combinato degli strumenti

- **Pulizia statistica** adattiva (A) per tipologia di frutto.
- **Rimozione** basata **sull'integrità** del **segnale** (B).
- **Disattivazione** dell'**interpolazione** (C-off).
- **interpolazione polinomiale** PCHIP (C-PCHIP).
- **Denoising Savitzky-Golay** (D).
- **Combinazione** (STACK)

3. Risultati

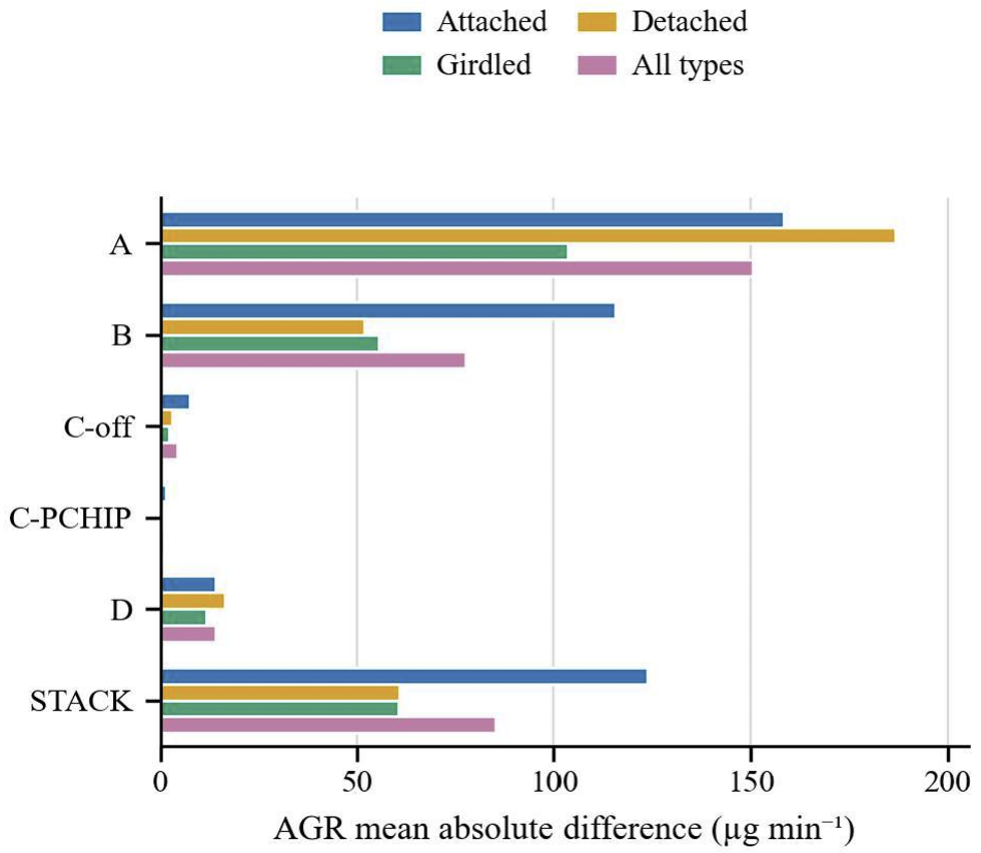
Comparazione delle prestazioni dei diversi strumenti dello script

Gruppo di Lavoro in Sensoristica ed Intelligenza Artificiale in Ortoflorofrutticoltura

Le 21 run differenti hanno permesso di quantificare come ciascuna scelta di elaborazione influisca sulla **disponibilità** di **celle AGR valide**, sulla **magnitudine dell'AGR**.

Magnitudine di AGR post elaborazione:

- La run A produce le differenze più piccole per tutti i tipi di frutto (0.3 – 1.25 µg/min)
- A mostra le differenze più marcate, in particolare sui frutti Attaccati (186.59 µg/min);
- B (integrità) ha l'effetto complessivo maggiore sulla rimozione dei dati, seguita da STACK; entrambe ne sono soprattutto sui frutti Attaccati.
- Le due varianti di interpolazione (C-off, C-PCHIP) hanno effetto trascurabile sulla disponibilità di celle.
- Il denoising (D) mostra un lieve aumento (+0.35 pp).



3. Discussione

I **vincoli fisici** restano la **principale salvaguardia** della pipeline, mentre gli **strumenti statistici sono complementari** nello screening del segnale. In assenza di un ground truth indipendente, la **differenza rispetto a R0 misura** la **sensibilità** del **metodo**, **non** la sua **accuratezza**: servirà una validazione con misure gravimetriche, diametri manuali o artefatti controllati.

- La **pulizia Statistica adattiva (A)** valuta **l'anomalia** rispetto alla **distribuzione locale** del segnale, mentre la **pulizia fisica** valuta il tasso di **variazione reale** implicato: un valore può quindi superare la pulizia adattiva ma essere comunque respinto a valle dai controlli fisici su W e AGR.

- La **rimozione** basata **sull'integrità** del segnale, a differenza del solo flag, **elimina osservazioni grezze prima** della **ricostruzione**, propagando **l'effetto lungo tutta** la **cascata**; la sensibilità **varia per tipo** di frutto (gli Anulati perdono più celle, gli Attaccati mostrano variazioni di AGR maggiori).

- **STACK non** è la **somma** delle singole varianti, **ma** riflette **l'interazione tra strumenti.**

- Disattivare **l'interpolazione** ordinaria o usare PCHIP **produce differenze limitate**, segno che le **lacune** ammissibili sono **generalmente brevi** e ben rappresentate da una transizione lineare;

- Il filtro **Savitzky-Golay migliora la continuità locale** ma **pochi** dei nuovi **valori** ricostruiti **superano** i controlli di **plausibilità** di **AGR**.

Conclusioni

Le **scelte di elaborazione influenzano gli output** del fruttometro sia in termini di **disponibilità** dei dati sia di **magnitudine dell'AGR**.

- I **vincoli fisici** rappresentano la principale salvaguardia contro **valori meccanicamente o fisiologicamente implausibili.**
- La **pulizia adattiva conserva più** osservazioni nelle fasi iniziali, ma parte di queste viene poi respinta dai controlli su W e AGR, poiché la **coerenza statistica non garantisce la plausibilità fisica.**

In sintesi, **interpolazione lineare e pulizia fisica** rappresentano le **scelte** di default più **conservative**, mentre gli **strumenti statistici e di smoothing** offrono un **supporto** complementare in condizioni di segnale specifiche senza tralasciare variabilità temporali e integrazione dati ancillari che possano validare limiti per trattamento.

Grazie per l'attenzione

Gruppo di Lavoro in Sensoristica ed Intelligenza Artificiale in Ortoflorofrutticoltura

Per il supporto:

A. Boini
I. Ansaloni
E. Tarpinelli
C. Franceschini
G. Perulli
L. Manfrini

II Giornate Tecniche SOI
Bari 10-11 settembre 2026

2. Materiali e metodi

Pulizia dati e interpolazione dati

Livello	Metodo	Funzionalità	Tipo
Dati grezzi	Protezione soglia e range	Segnala valori zero, a fine corsa o non validi prima di ogni calcolo derivato	Core
Dati grezzi	Rilevamento integrità del segnale	Registra spike, clipping, intermittenza, assestamento, salti di livello, perdita del frutto, deriva	Ablation
ΔmV	Pulizia fisica (physical bound)	Rigetta variazioni fuori dall'inviluppo fisico di velocità per tipo di frutto	Core
ΔmV	Pulizia stocastica	Rileva salti adattivi, micro-variazioni ripetute e anomalie MAD locali	Ablation
ΔmV	Interpolazione ordinaria	Colma lacune interne fino a 3 celle (lineare o PCHIP)	Ablation
ΔmV	Ricostruzione strutturale	Colma solo eventi strutturali classificati, entro limiti dedicati	Revisione manuale
ΔmV	Imputazione Savitzky-Golay	Stima solo le celle mancanti, non sovrascrive mai un valore osservato valido	Ablation
W (peso)	Inviluppo fisico del tasso	Segnala tassi di variazione del peso implausibili	Core
W (peso)	Rolling MAD	Segnala anomalie locali di peso	Core
AGR	Limiti fisici	Segnala tassi di crescita assoluti fuori dai limiti fisici per tipo di frutto	Core
AGR	IQR	Segnala valori distributivamente estremi	Core
AGR	Z-score	Segnala valori standardizzati estremi	Core
AGR	Rolling MAD	Segnala scostamenti locali dalla mediana mobile	Core
AGR	Bridge	Segnala anomalie isolate o brevi tra endpoint validi	Core
AGR	Inviluppo Attached	Segnala tassi fuori dall'inviluppo di riferimento dei frutti Attaccati	Core