



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*

## **Decentralized Olive Orchard Monitoring with IoT and Cloud Computing - Cloud ETL Orchestration & 4G/5G Edge Connectivity**

*Antonio Francesco Gentile, Davide Macrì, Emilio Greco*

**RT-ICAR-CS-26-02**

**Gennaio 2026**



*The ICAR-CNR technical reports are published by the Institute of High Performance Computing and Networks of the National Research Council. These reports, prepared under the exclusive scientific responsibility of the authors, describe research activities of ICAR staff and collaborators, in some cases in a preliminary format before definitive publication elsewhere.*

# Decentralized Olive Orchard Monitoring with IoT and Cloud Computing - Cloud ETL Orchestration & 4G/5G Edge Connectivity

Antonio Francesco Gentile, Davide Macrì, Emilio Greco

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche (ICAR-CNR)

Via Pietro Castellino, 111 – 80131 Napoli, Italia

{antoniofrancesco.gentile, davide.macri, emilio.greco}@icar.cnr.it

## Abstract

Questo rapporto tecnico descrive l'architettura di comunicazione di un prototipo decentralizzato per il monitoraggio avanzato di frutteti di olivo, sviluppato presso il CREA-OFA e l'ICAR-CNR.

Il sistema integra sensori IoT distribuiti (Sap Flow Meters e stazioni metrologiche Campbell Scientific), l'infrastruttura cloud AWS per il servizio M2M ed altri sistemi per l'acquisizione, l'elaborazione e la visualizzazione dei dati agronomici. L'architettura privilegia l'indipendenza dalle infrastrutture locali LAN/WAN, utilizzando esclusivamente la copertura cellulare 4G/5G per la trasmissione dei dati.

Il design decentralizzato elimina i tradizionali single points of failure, garantendo la continuità operativa anche in ambienti remoti. La configurazione adattiva degli intervalli di acquisizione e di trasmissione in funzione della disponibilità energetica (radiazione solare, copertura nuvolosa, variabilità stagionale) assicura l'operatività autonoma del sistema in condizioni energetiche variabili. L'architettura rappresenta un approccio innovativo alla raccolta di dati agricoli di precisione, combinando flessibilità di deployment, resilienza operativa e costi ridotti dell'infrastruttura.

## Sommario dei Contenuti

1. Introduzione
2. Architettura di Comunicazione del Prototipo
3. Dispositivi di Campo e Data Logger
4. Trasmissione Dati e Connettività Cloud
5. Sincronizzazione Temporale e Visualizzazione
6. Algoritmo ETL e Orchestrazione dei Dati
7. Ottimizzazione Energetica e Timing Adattivo
8. Considerazioni sulla Resilienza e Affidabilità
9. Conclusioni e Sviluppi Futuri

# 1. Introduzione

L'agricoltura di precisione rappresenta uno dei settori chiave per aumentare la sostenibilità e l'efficienza produttiva. La gestione ottimale dei frutteti di olivo richiede il monitoraggio continuo di numerosi parametri fisiologici delle piante e delle condizioni microambientali locali. Tradizionalmente, tale monitoraggio è vincolato a infrastrutture locali proprietarie, che comportano elevati costi di implementazione, una limitata scalabilità geografica e una dipendenza da componenti hardware centrali che possono rappresentare single points of failure.

Il prototipo descritto in questo rapporto affronta queste limitazioni attraverso un'architettura radicalmente decentralizzata che sfrutta la copertura capillare della rete cellulare 4G/5G come backbone di connettività. Questo approccio consente il deployment di dispositivi di monitoraggio in qualsiasi ubicazione geografica con copertura cellulare, indipendentemente dall'esistenza di infrastrutture di rete locali.

L'architettura integra due tipologie di sensori specializzati: Sap Flow Meters (SFM1-X) per la misurazione del flusso linfatico intravascolare delle piante e stazioni meteorologiche Campbell Scientific multisensore. Entrambi i dispositivi trasmettono i dati direttamente al cloud AWS tramite moduli di comunicazione 4G/5G, eliminando la necessità di gateway locali o reti mesh proprietarie.

## 2. Architettura di Comunicazione del Prototipo

L'architettura di comunicazione è stata concepita per garantire la massima flessibilità operativa e l'indipendenza dalle infrastrutture proprietarie, sfruttando la copertura capillare della rete cellulare come backbone primario di connettività.

### 2.1 Modello Decentralizzato Edge-to-Cloud

Il sistema adotta un modello edge-to-cloud completamente decentralizzato, in cui i dispositivi di monitoraggio operano come nodi autonomi della rete. Questa configurazione elimina i tradizionali single points of failure associati a gateway locali centralizzati o a router proprietari.

### 2.2 Caratteristiche Chiave dell'Architettura

- **Autonomia Energetica:** Alimentazione tramite pannelli solari e batterie di backup per operazione off-grid continua
- **Connettività Cellulare Decentralizzata:** Comunicazione diretta tramite moduli 4G/5G senza infrastrutture locali intermedie

- Indipendenza da protocolli proprietari: utilizzo di standard aperti MQTT e protocolli pubblici per l'eliminazione di vendor lock-in
- Sincronizzazione Temporale: gestione autonoma del timestamping mediante protocolli NTP standard
- Mitigazione Guasti: architettura senza punti singoli di failure, garantendo continuità dati anche in ambienti geograficamente isolati

## 3. Dispositivi di Campo e Data Logger

### 3.1 Sap Flow Meters SFM1-X

I Sap Flow Meters utilizzati nel prototipo sono basati sulla serie SFM1-X, una soluzione altamente specializzata per il monitoraggio del flusso linfatico intravascolare mediante la tecnica di dissipazione termica. Questi dispositivi incorporano un Universal Communications Module (UCM) che offre la massima flessibilità di connessione.



Specifiche tecniche di connettività:

- Moduli di trasmissione 4G/5G LTE-M / Cat-M1 integrati
- Scheda SIM dedicata per comunicazione cellulare autonoma
- Protocolli alternativi supportati: Bluetooth, USB, LoRaWAN (non utilizzati in questa configurazione)
- Formato dati aperto: assenza di codifiche proprietarie o decoding obbligatori

Il vantaggio dell'utilizzo dei moduli LTE-M/Cat-M1 consiste nella fornitura di dati in formato aperto, totalmente libero da vincoli di proprietà. Ciò consente il pieno controllo dei dati dal punto di acquisizione, garantendo totale flessibilità nella raccolta, nell'archiviazione e nella visualizzazione.

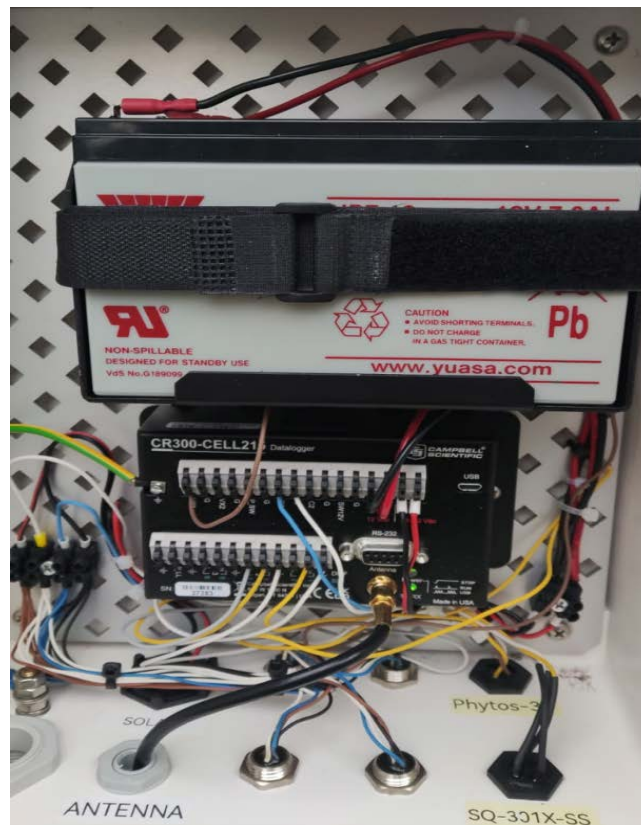
Questa caratteristica elimina la dipendenza da specifici software di decoding forniti dal produttore, riducendo notevolmente i costi operativi e aumentando l'interoperabilità del sistema.

### 3.2 Stazione Metrologica Campbell Scientific CR300

La stazione metrologica Campbell Scientific, configurata con il datalogger CR300 e il software PC400 (versione 5.10.x), costituisce la componente di acquisizione ambientale del prototipo. Questa stazione è equipaggiata con:

- Sensori integrati: Dendrometer Environmental Measuring Systems DRL26X; Leaf Temperature Sensor Apogee Instruments SF-421; Leaf Wetness Sensor Meter Phytos 31; Linear PAR Sensor Apogee Instruments MQ-301X; Stevens Water Monitoring Systems HydraProbe Soil Measurement Sensor (below-ground); Profile probe Sentek Drill&Drop (below-ground).
- Datalogger autonomo CR300 con memoria interna e gestione locale dei dati
- Antenna 4G con scheda SIM per trasmissione cellulare diretta
- Software PC400 per configurazione e gestione dei parametri di acquisizione

La stazione dispone di un modulo 4G LTE integrato con SIM dedicata, consentendo la trasmissione diretta dei dati verso AWS. Come il sistema SFM1-X, elimina completamente la dipendenza dalle infrastrutture di rete locali.



### 3.3 Data Logger Autonomi e Sincronizzazione Temporale

Entrambi i dispositivi (SFM1-X e stazione Campbell) incorporano datalogger autonomi che svolgono le seguenti funzioni critiche:

- Acquisizione locale: memorizzazione dei dati acquisiti con timestamp locale
- Buffering: mantenimento di dati in memoria in caso di temporanei problemi di connettività
- Sincronizzazione temporale: gestione del timestamping mediante protocolli NTP standard, garantendo allineamento temporale globale
- Gestione energetica: monitoraggio del livello di carica della batteria e regolazione dei cicli di acquisizione

La sincronizzazione temporale tra i datalogger è un aspetto critico per l'analisi correlata dei dati. Entrambi i dispositivi sincronizzano l'orologio interno tramite query SNTP (Simple Network Time Protocol) verso server pubblici via connessione 4G, assicurando che i timestamp associati a ogni acquisizione siano coerenti con una scala temporale comune.



## 4. Trasmissione Dati e Connettività Cloud

La trasmissione dei dati rappresenta il nodo critico della comunicazione tra il campo e l'infrastruttura di elaborazione. La strategia adottata privilegia l'affidabilità, l'efficienza energetica e l'indipendenza dalle infrastrutture proprietarie.

### 4.1 Connettività 4G/5G come Backbone Primario

I dati acquisiti dai dispositivi di campo vengono trasmessi al cloud AWS esclusivamente tramite moduli cellulari 4G/5G LTE-M/Cat-M1 integrati nei dispositivi stessi. Questa architettura elimina completamente la necessità di:

- Gateway locali
- Router proprietari centralizzati
- Reti mesh o infrastrutture di networking custom
- Reliance su operatori di rete privati

L'utilizzo della rete cellulare pubblica comporta vantaggi significativi:

- Copertura geografica illimitata (disponibile ovunque sia presente segnale cellulare)
- Costi di infrastruttura ridotti (nessun deployment di rete locale)
- Scalabilità geografica immediata (aggiunta di nuovi dispositivi in aree coperte da rete cellulare)
- Assenza di single points of failure legati a infrastrutture proprietarie locali

### 4.2 Middleware Cloud AWS e Riconciliazione Protocolli

La trasmissione dei dati verso il cloud AWS rappresenta un elemento architetturale critico per la seguente ragione: i moduli cellulari dei Sap Flow Meters e della stazione Campbell operano secondo protocolli di comunicazione parzialmente eterogenei. Il cloud AWS funge da middleware di comunicazione centralizzato, incaricato di:

- Riconciliazione protocolli: armonizzazione dei diversi formati di trasmissione provenienti dai due tipi di dispositivi
- Sincronizzazione dati: allineamento temporale dei flussi dati eterogenei in una scala temporale globale coerente
- Buffering e queuing: gestione delle variabilità nella disponibilità di connessione cellulare
- Persistenza: archiviazione affidabile dei dati in storage

I dati sono inviati tramite il protocollo MQTT (Message Queuing Telemetry Transport), uno standard aperto ottimizzato per trasmissioni a basso consumo energetico e a larghezza di banda ridotta. Queste caratteristiche sono essenziali per dispositivi alimentati da sistemi fotovoltaici in ambienti remoti. AWS implementa un broker MQTT che:

- Accetta messaggi dai dispositivi (publish)
- Li archivia in database relazionale MariaDB SQL per indicizzazione complessa e analisi a lungo termine
- Trigger funzioni Lambda per elaborazione in tempo reale e sincronizzazione temporale globale

## 5. Sincronizzazione Temporale e Visualizzazione

La sincronizzazione temporale rappresenta un aspetto critico per l'analisi integrata di dati provenienti da dispositivi eterogenei distribuiti geograficamente.

### 5.1 Meccanismi di Sincronizzazione

Il sistema implementa una strategia multi-livello per la sincronizzazione temporale:

#### **Livello 1 - Edge (Dispositivi di Campo):**

Ogni dispositivo (SFM1-X e stazione Campbell) sincronizza il proprio orologio interno tramite query SNTP verso server NTP pubblici, accessibili tramite connessione cellulare. Questo assicura che ogni timestamp generato localmente sia riferito a una scala temporale globale coerente, con accuratezza tipica di  $\pm 100-500$  ms.

#### **Livello 2 - Cloud (AWS):**

Il middleware cloud AWS implementa ulteriore sincronizzazione globale, verificando la coerenza temporale dei dati in ingresso e applicando timestamp server-side per garantire un ordine cronologico inequivocabile.

#### **Livello 3 - Visualizzazione (Grafana):**

L'interfaccia di visualizzazione di Grafana accede ai dati centralizzati in AWS e visualizza in tempo reale le acquisizioni sincronizzate provenienti da entrambi i dispositivi.

### 5.2 Interfaccia di Visualizzazione Grafana

Il Grafana Dashboard Service fornisce un'interfaccia web basata su dataset complessi, traducendoli in informazioni visuali immediatamente interpretabili dagli operatori.

Le caratteristiche principali includono:

- Visualizzazioni in tempo reale dei dati sincronizzati da SFM1-X e stazione Campbell
- Indicatori visivi per parametri critici (es., velocità vento, umidità, flusso linfatico)

- Alerts configurabili per condizioni anomale (es., velocità vento elevata che potrebbe causare deriva dei trattamenti)
- Accesso collaborativo: agronomi, ricercatori e gestori del frutteto possono simultaneamente visualizzare i medesimi dati sincronizzati
- Storico dati: capacità di analizzare trend a lungo termine e correlazioni tra variabili fisiologiche e ambientali

## 6. Algoritmo ETL e Orchestrazione dei Dati

L'elaborazione dei dati grezzo acquisiti dai dispositivi di campo viene gestita tramite un algoritmo ETL (Extract, Transform, Load) eseguito su un server Grafana dedicato. Questo processo automatizzato armonizza flussi di dati eterogenei e li rende disponibili per la visualizzazione e l'analisi.

### 6.1 Fase 1: Estrazione (Multi-Protocol Ingestion)

Lo script automatizzato effettua il retrieval dei dati dall'endpoint AWS secondo due protocolli distinti:

FTP Mirroring (Stazione Campbell CR300):

- Utilizza lftp con mirroring parallelizzato per il retrieval efficiente di file .dat e .csv
- Implementa delta-sync (trasferimento solo dati nuovo/modificati) per ridurre bandwidth

HTTP Retrieval (Sap Flow Meters SFM1-X):

- Utilizza curl con User-Agent Rotation per query agli endpoint AWS hosting dati SFM1-X
- Parsing automatico di response JSON con deserializzazione in strutture dati native

### 6.2 Fase 2: Trasformazione (Normalizzazione & Sanitizzazione)

Header Injection:

- Prepend di header standardizzato SAP\_FLOW\_HEADER a file raw SFM1-X
- Assicura compatibilità con strumenti di analisi statistica standard (R, Python, Excel)

Dynamic Metadata Merging (Campbell CR300):

- Parsing di file .dat originali
- Interleaving di nomi variabili e unità (es., Variable - (Unit))
- Ricostruzione di CSV leggibile ad umani con colonne semantiche

Data Cleaning:

- Normalizzazione line-ending (dos2unix) per cross-platform compatibility
- Rimozione quotes spurie (tr per character translation)
- Validazione timestamp mediante regex per identificazione righe malformate

### 6.3 Fase 3: Loading & Publishing

Symbolic Linking:

- Mapping del dataset più recente a latest.csv per accesso immediato

Real-time Availability:

- Automazione gestione permessi file (chown) per account servizio Grafana
- Notifiche event-driven che triggherano aggiornamento dashboard

### 6.4 Change Detection & Archiving

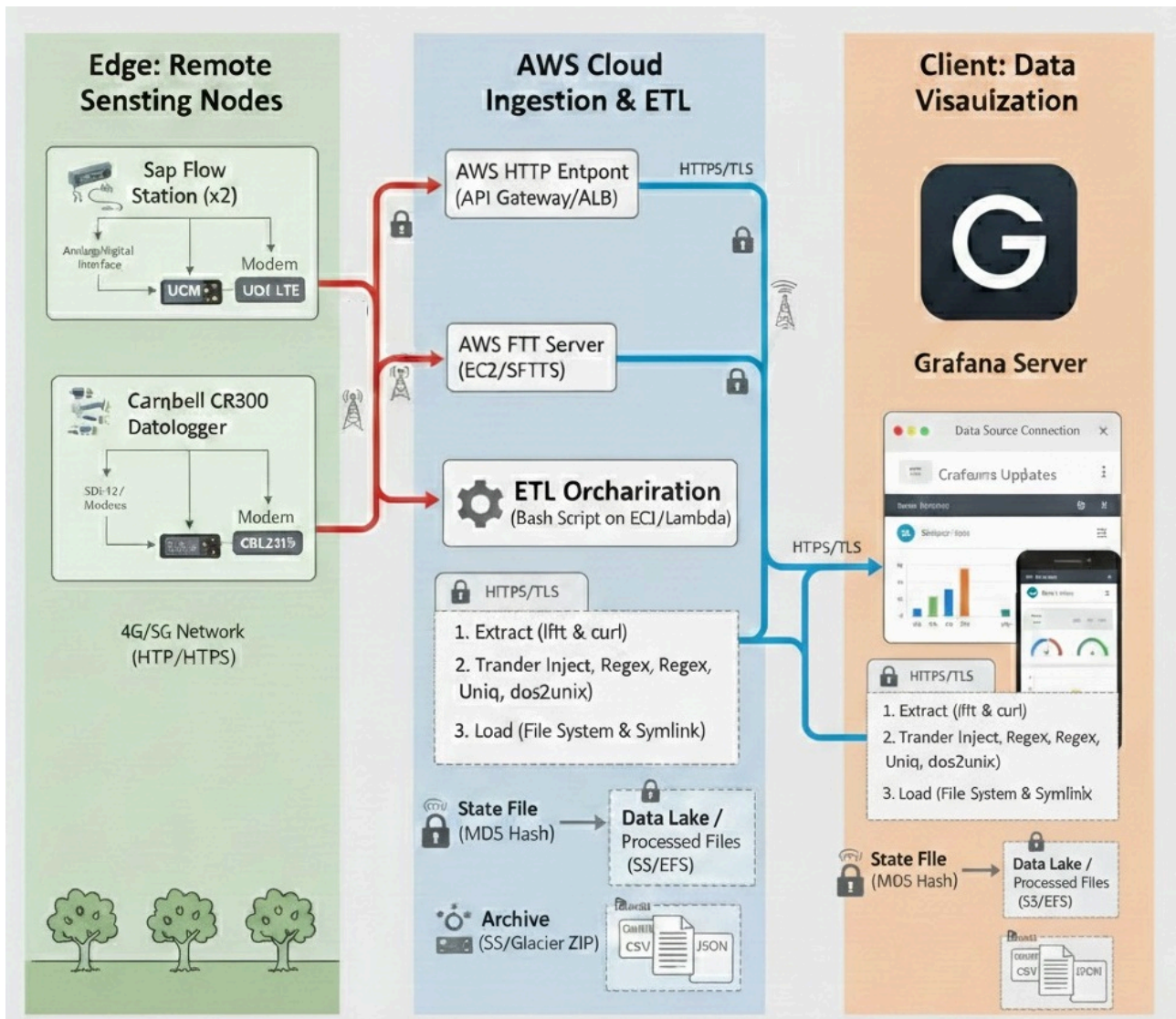
Per minimizzare overhead I/O e costi di storage cloud:

MD5 Cryptographic Hashing:

- Monitoraggio della directory source per rilevamento cambiamenti di stato
- Confronto hash garantisce detection di dati veramente nuovi vs. rielaborazione

Conditional ZIP Archiving:

- Generazione di backup con timestamp solo quando rilevati dati nuovi
- Assicura cronologia pulita e versionata della telemetria del frutteto



## 7. Ottimizzazione Energetica e Timing Adattivo

La criticità gestionale del prototipo risiede nella calibrazione dinamica degli intervalli di acquisizione e di trasmissione dei dati in funzione di molteplici variabili energetiche e ambientali.

### 7.1 Vincoli Energetici nei Sistemi Fotovoltaici Remoti

La disponibilità energetica è governata da fattori stagionali e geografici che modulano significativamente l'efficienza del sistema fotovoltaico:

- Radiazione solare incidente: variabile con stagione, latitudine, nuvolosità
- Angolo di incidenza dei raggi solari: diminuisce alle latitudini elevate, specialmente in inverno
- Copertura nuvolosa locale: riduce drasticamente la radiazione disponibile

- Variabilità stagionale della latitudine geografica: affezione periodo giorno/notte

## 7.2 Modello di Bilancio Energetico e Batching Strategy

Ogni ciclo di trasmissione cellulare comporta un consumo energetico articolato:

**Ewake:** Energia consumata per fase di registrazione alla rete cellulare (fase più costosa)

**Etrans:** Energia consumata per trasmissione del payload dati

**Dsize:** Dimensione del payload (correlata linearmente a Etrans)

La strategia di batching ottimale consiste nell'aumentare l'intervallo di trasmissione ( $T_{int}$ ) e ridurre la frequenza della fase alta-energia di "Wake/Registration" (Ewake). Questo approccio amplia significativamente l'autonomia della batteria in condizioni di radiazione solare ridotta (giornate coperte, stagioni invernali, latitudini elevate).

## 7.3 Calibrazione Adattiva dei Tempi

Gli intervalli di acquisizione e cicli di trasmissione devono essere configurati adattativamente in base a:

- ubicazione geografica dell'orto (latitudine)
- previsione radiazione media stagionale
- condizioni meteorologiche locali previste
- forecast nuvolosità
- disponibilità energetica misurata
- stato batteria tempo-reale

In scenari di radiazione ridotta:

- Incrementare intervalli di acquisizione (es., ogni 30 minuti invece di ogni 5 minuti)
- Ridurre frequenza trasmissione (es., ogni 4 ore invece di ogni 30 minuti)
- Mantenere bilancio energetico positivo: carica solare > consumo giornaliero

## 7.4 Threshold di Trasmissione Basato su Carica Batteria

La trasmissione dati avviene esclusivamente quando il livello di carica della batteria supera soglie critiche predefinite:

- Soglia minima: batteria deve avere carica  $\geq 40\%$  per abilitare trasmissione

- Soglia operativa: acquisizione continua, ma trasmissione differita fino al raggiungimento soglia
- Failsafe: disattivazione acquisizione se batteria scende sotto 20%, preservando memoria e circuiti

Questa strategia garantisce:

- Continuità operativa autonoma in condizioni energetiche variabili
- Prevenzione della disattivazione completa per insufficienza energetica
- Flessibilità di deployment in qualsiasi area geografica con copertura cellulare

## 8. Considerazioni sulla Resilienza e Affidabilità

L'architettura decentralizzata del prototipo conferisce proprietà di resilienza superiori rispetto ad approcci tradizionali centralizzati.

### 8.1 Eliminazione di Single Points of Failure

Assenza di gateways locali centrali: ogni dispositivo trasmette autonomamente su cloud  
 Assenza di router proprietari: eliminazione di componenti critiche single-failure  
 Indipendenza da reti mesh: nessuna dipendenza da catene di relay che potrebbero fallire.

### 8.2 Continuità Operativa in Ambienti Isolati

La strategia di batching energetico con soglie di trasmissione basate su carica batteria assicura:

- Acquisizione dati continua anche quando connettività cellulare è temporaneamente indisponibile
- Trasmissione recuperata automaticamente quando connessione cellulare ristabilita
- Buffering locale temporaneo per periodi di assenza rete

### 8.3 Standardizzazione Aperta

Utilizzo di protocolli standard aperti (MQTT, SNMP, FTP, HTTP) garantisce:

- Assenza di vendor lock-in
- Interoperabilità con ecosistemi software standard
- Facilità di manutenzione e evoluzione del sistema

- Riducibilità dei costi operativi a lungo termine

## 9. Conclusioni e Sviluppi Futuri

Il prototipo descritto rappresenta un significativo avanzamento nell'architettura di monitoraggio distribuito per applicazioni agronomiche di precisione. L'integrazione di tecnologie IoT decentralizzate, connettività cellulare pubblica e infrastruttura cloud di AWS fornisce una soluzione altamente flessibile, resiliente e cost-effective per il monitoraggio continuo dei frutteti.

Gli elementi innovativi chiave includono:

1. Decentralizzazione radicale: eliminazione di infrastrutture locali proprietarie a favore di reti cellulari pubbliche
2. Indipendenza energetica: estrattore autonomo in condizioni remote via fotovoltaico e batterie
3. Flessibilità di deployment: installazione possibile in qualsiasi ubicazione con copertura cellulare
4. Timing adattivo: calibrazione dinamica degli intervalli in funzione della disponibilità energetica
5. Standardizzazione: utilizzo esclusivo di protocolli aperti e formati dati proprietà-liberi

Sviluppi futuri potranno includere:

- Espansione a reti multi-sensore con ulteriori parametri fisiologici
- Integrazione di algoritmi ML per predictive analytics di malattie e stress idrico
- Estensione geografica a multiple aree di coltivazione con frontend unificato
- Ottimizzazione further di consumi energetici mediante hardware edge specializzato
- Integrazione con sistemi di controllo automatizzato per irrigazione e trattamenti

## Riferimenti Bibliografici

- [1] Campbell Scientific, Inc. (2023). CR300 Micrologger User Manual. Available at: <https://www.campbellsci.com/>
- [2] Campbell Scientific, Inc. (2023). CELL215 Industrial Modem: Technical Specifications and Integration Guide. Available at: <https://www.campbellsci.com/cell215>
- [3] Campbell Scientific, Inc. (2023). PC400 Data Retrieval Software: User Manual v5.10.x. Available at: <https://www.campbellsci.com/software>
- [4] Campbell Scientific, Inc. (2023). SFM1-X Sap Flow Meter: Technical Manual and Installation Guide. Available at: <https://www.campbellsci.com/sfm1x>
- [5] Campbell Scientific, Inc. (2023). Universal Communications Module (UCM): Connectivity Options and Protocols. Available at: <https://www.campbellsci.com/ucm>
- [6] Amazon Web Services (AWS) (2024). AWS IoT Core Documentation: MQTT Protocol and Message Queuing. Available at: <https://docs.aws.amazon.com/iot-core/>

- [7] Amazon Web Services (AWS) (2024). AWS Lambda Developer Guide: Event-Driven Processing and Real-Time Data Synchronization. Available at: <https://docs.aws.amazon.com/lambda/>
- [8] Amazon Web Services (AWS) (2024). Amazon S3 User Guide: Object Storage and Data Archiving. Available at: <https://docs.aws.amazon.com/s3/>
- [9] Amazon Web Services (AWS) (2024). Amazon Relational Database Service (RDS) for MariaDB: User Guide and Best Practices. Available at: [https://docs.aws.amazon.com/rds/latest/UserGuide/USER\\_MariaDB.html](https://docs.aws.amazon.com/rds/latest/UserGuide/USER_MariaDB.html)
- [10] IETF RFC 5905 (2010). Network Time Protocol Version 4: Protocol and Algorithms Specification. Mills, D., et al. Available at: <https://tools.ietf.org/html/rfc5905>
- [11] IETF RFC 4648 (2006). The Base16, Base32, and Base64 Data Encodings. Josefsson, S. Available at: <https://tools.ietf.org/html/rfc4648>
- [12] Grafana Labs (2024). Grafana Dashboard Service: User Guide and Visualization Configuration. Available at: <https://grafana.com/docs/grafana/latest/>
- [13] Grafana Labs (2024). Grafana Alerting: Real-Time Notifications and Threshold Management. Available at: <https://grafana.com/docs/grafana/latest/alerting/>
- [14] OASIS (2014). MQTT Version 3.1.1: Protocol Specification. Available at: <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/>
- [15] W3C (2004). XML Schema Part 2: Datatypes Second Edition. Available at: <https://www.w3.org/TR/xmlschema-2/>