

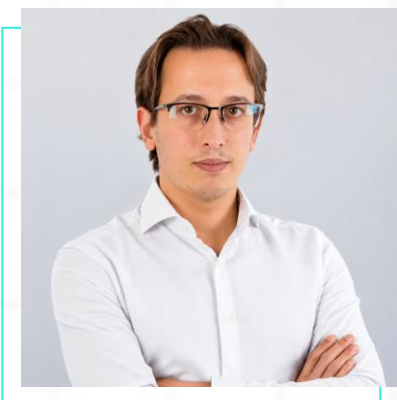
Flexible Multi-Source Data Pipeline for Precision Irrigation

From Knowledge to Competitiveness

Portugal-US Partnerships 2026 Summit

Space, Climate and Sustainable Systems

MIT Portugal Program



Carlos Lemonde¹

carloslemonde@isa.ulisboa.pt

Supervisor(s): Teresa Afonso do Paço¹, João Rolim¹

1. LEAF - Linking Landscape, Environment, Agriculture and Food Research Center, Associate Laboratory TERRA, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal

Research Objectives

Develop an open-source data pipeline capable of integrating and processing heterogeneous multi-sensor data into a common time-series infrastructure, supporting real-time monitoring, sensor calibration, and the discovery of patterns across related variables.

Challenges

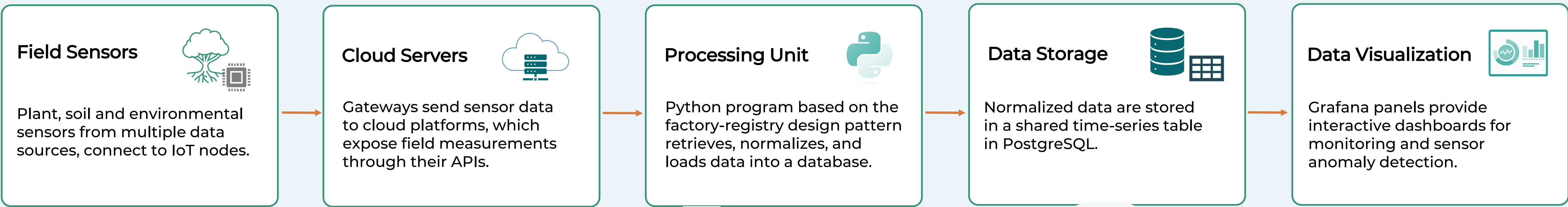
- Shifting temperature and precipitation patterns make freshwater availability less predictable, increasing the need for data-driven irrigation.
- Sensor vendors expose different data formats and sampling intervals, making coordinated monitoring and analysis difficult.

Contributions

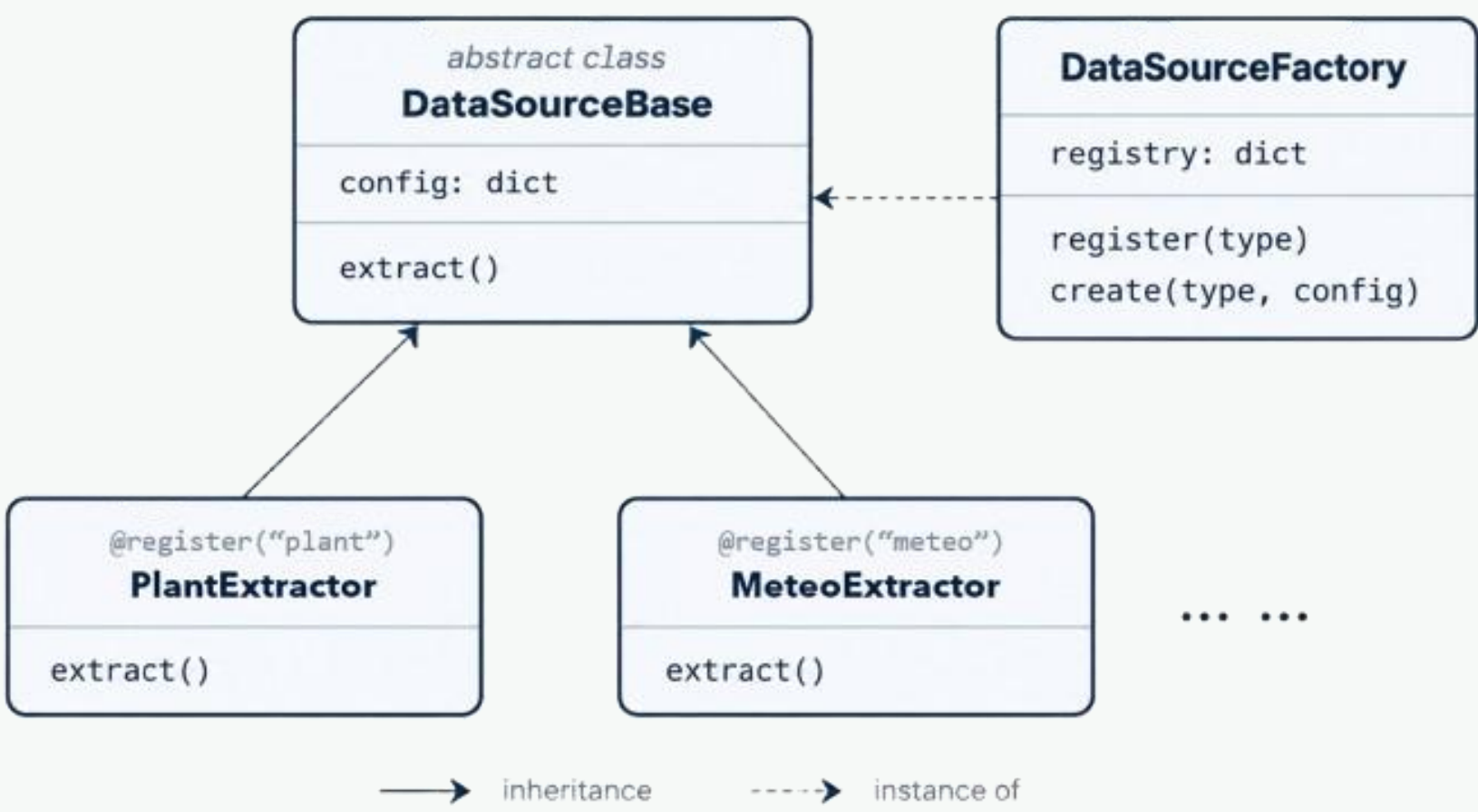
This work contributes to **APICC** (Adapting Precision Irrigation to Climate Change) project by complementing **SOFIS**, a low-cost IoT system designed to control water use in orchards and vineyards.

SYSTEM ARCHITECTURE

1 MONITORING DATA FLOW



The **Factory-Registry Pattern** decouples data source-specific extractors from the core pipeline logic. A centralized factory creates the appropriate extractor, while a decorator registry automatically associates extractor classes with their data source identifiers. This modular design enables new data sources to be integrated without modifying the core pipeline.

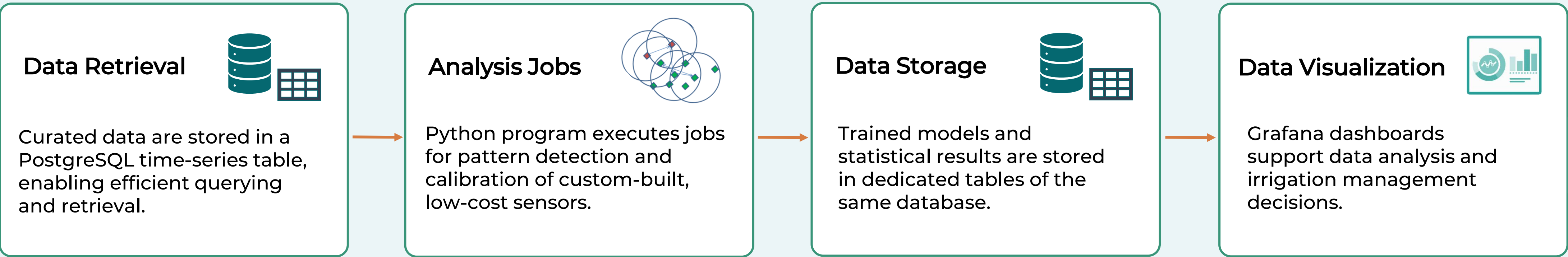


The **Time-Series Schema** follows an Entity-Attribute-Value (EAV) model, comprising two metadata tables and a timestamped data table. By storing attributes as rows rather than fixed columns, it enables new attributes to be added without significant structural changes.

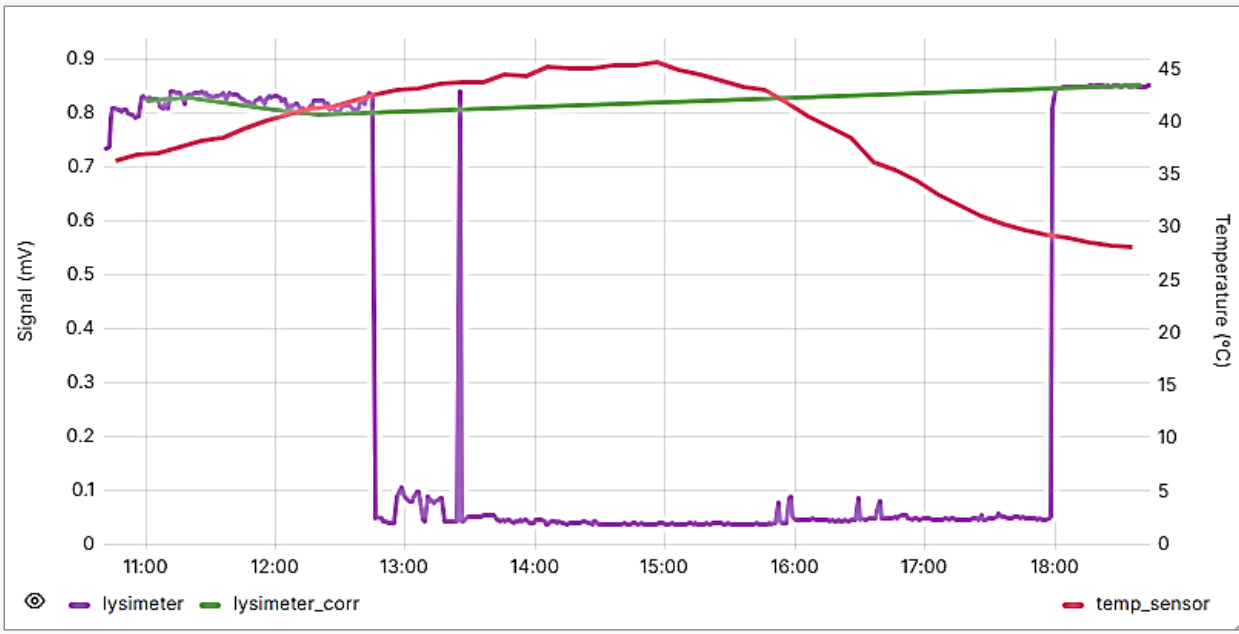
devices			metrics		
Column	Type	Constraint	Column	Type	Constraint
device_id	bigint	PK, identity	metric_id	bigint	PK, identity
name	varchar(50)	not null, unique	name	varchar(50)	not null, unique
device_type	varchar(30)	not null	unit	varchar(20)	not null
location	varchar(100)	—	description	text	—

sensor_data		
Column	Type	Constraint
time	timestamp	PK (composite)
device_id	bigint	FK → devices, PK
metric_id	bigint	FK → metrics, PK
value	double precision	not null

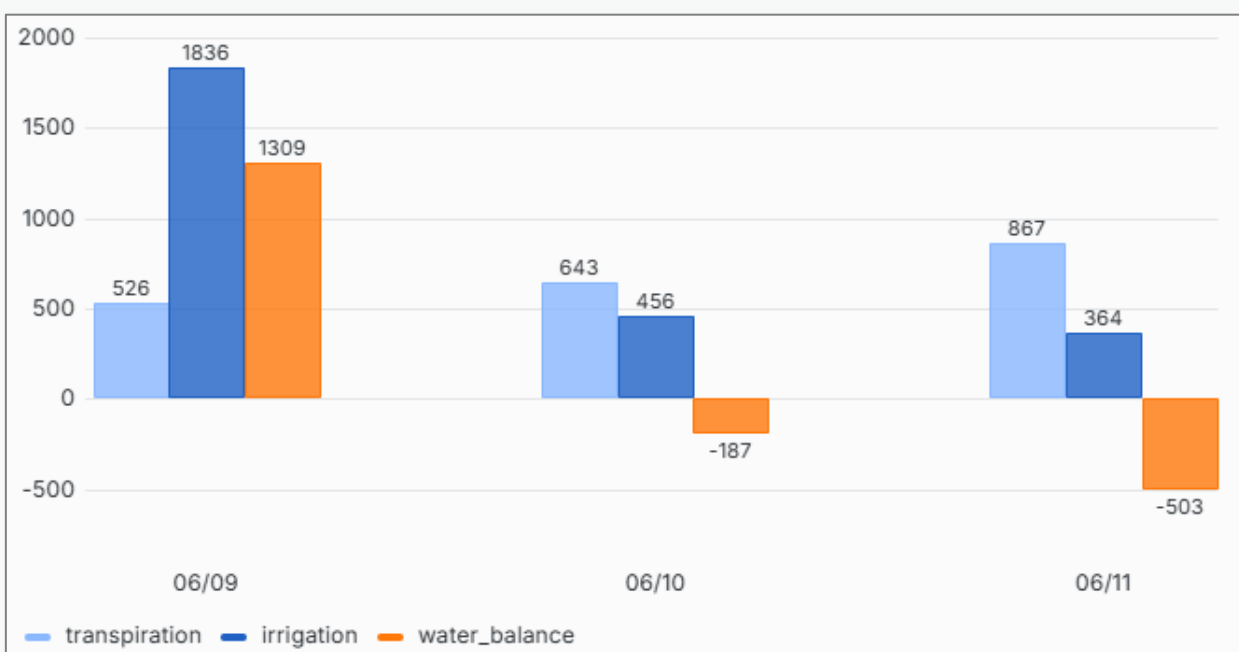
2 ANALYTICS DATA FLOW



Thermal drift correction



Daily transpiration



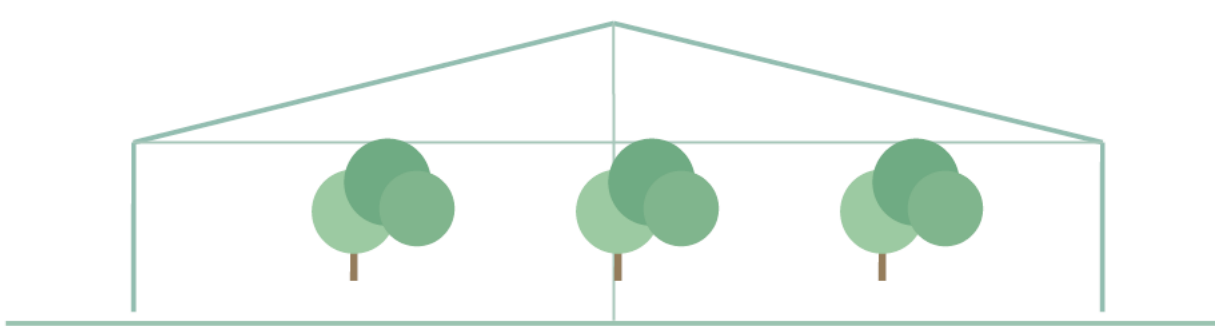
CONCLUSIONS

- The data pipeline demonstrated flexibility and modularity, easily integrating different data sources from the experimental setup and providing valuable insights for the SOFIS system.
- Further improvements are needed to enhance data processing and automation.

EXPERIMENTAL APPLICATION

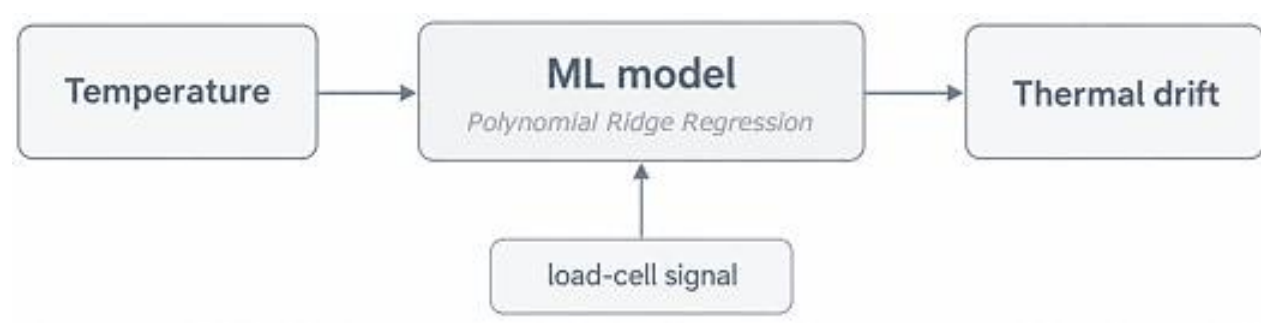
Greenhouse demonstration

Greenhouse experiment at *Instituto Superior de Agronomia* (Lisbon) using potted olive trees and integrating heterogeneous sensors while preserving temporal resolution.



Sensor correction

A thermal-drift correction algorithm was applied using environmental data and machine-learning methods to compensate for temperature artifacts on load-cell measurements.



Estimate tree transpiration

Model trained to calibrate sap flow sensors and directly estimate tree transpiration from their measurements, using corrected load-cell data as a reference.

$$\hat{T}_{tree} = f_{ML}(SF)$$

Funded by:

Coordination:

Collaboration:

