UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea Triennale in
Ingegneria Informatica e dell'Automazione

Sviluppo di un sistema per l'automazione dei processi di raccolta e analisi dei dati idrici per il monitoraggio della sorgente di Gorgovivo

Development of an automated system for water data collection and analysis for the monitoring of the Gorgovivo spring

Relatore:
Dr. Galdelli Alessandro

Laureando:
Piemontese Alessandro

Correlatore:
Prof. Zingaretti Primo

Anno Accademico 2025-2026

Indice

Sommario

La gestione sostenibile delle risorse idriche rappresenta una sfida nel contesto di crisi climatica, caratterizzata da una variazione del ciclo idrologico e da una crescente pressione antropica sulla risorsa idrica. Questo scenario comporta una crescente incertezza legata alla variabilità delle precipitazioni e alla ricarica delle falde acquifere per tutte le sorgenti d'acqua sotterranee che costituiscono un'importante fonte di approvvigionamento idrico per molti territori.

Una situazione così descritta impone il superamento delle metodologie di monitoraggio tradizionale delle sorgenti idriche a favore di sistemi intelligenti, caratterizzati dall'analisi di dati, dal cloud computing e dall'Intelligenza Artificiale, al fine di raggiungere una migliore gestione della risorsa.

La presente tesi è inserita all'interno del progetto di ricerca "Gorgovivo 4.0", sviluppato in collaborazione tra il Consorzio di Gorgovivo, Viva Servizi S.p.A. e l'Università Politecnica delle Marche. Il progetto si occupa dello sviluppo di un sistema informativo tramite tecniche di cloud computing e Intelligenza Artificiale come supporto decisionale per la gestione della sorgente di Gorgovivo.

La tesi si inserisce nella fase di elaborazione dati al fine di produrre report informativi funzionali e strutturati. Il presente lavoro ha come obiettivo l'automazione dei processi di elaborazione dei dati pluviometrici acquisiti dalle stazioni presenti nei comuni serviti dalla sorgente di Gorgovivo. In particolare, il contributo descritto si è focalizzato sullo sviluppo di script in linguaggio Python per la generazione automatica di reportistica tecnica. L'infrastruttura del progetto, in cui si inserisce il lavoro di tesi, integra AWS S3 per l'archiviazione sicura dei dati e AWS Lambda per l'esecuzione periodica dello script. La logica del lavoro sviluppato si suddivide in pre-elaborazione del formato dei dati raccolti dalla rete sensoristica, elaborazione dei dati in aggregazioni temporali giornaliere, mensili e annuali e calcolo di indicatori statistici di supporto.

Il risultato del lavoro elaborato su Python e integrato con i servizi di AWS è la produzione sistematica di report strutturati in formato Excel, archiviati e aggiornati quotidianamente su cloud. I report integrano tabelle riassuntive, serie storiche e visualizzazioni grafiche per un'analisi accurata dei trend temporali e stagionali dei livelli di acqua.

L'implementazione effettuata garantisce elevata scalabilità, robustezza e continuità operativa, fornendo uno strumento decisionale efficace per l'analisi del-

le tendenze pluviometriche, l'individuazione di anomalie e la valutazione della capacità di ricarica delle falde acquifere.

Il sistema si configura come un supporto tecnologico per la governance moderna, trasparente e resiliente della risorsa idrica regionale, contribuendo concretamente agli obiettivi di gestione sostenibile definiti dalle normative Europee e nazionali.

Abstract

Sustainable water management represents a critical challenge within the current context of climate crisis, based on structural alterations in the hydrological cycle and increasing anthropogenic pressure on water supplies. This scenario entails growing uncertainty regarding precipitation variability and groundwater recharge for the subterranean springs that serve as essential water sources for many territories.

This situation necessitates moving beyond traditional spring monitoring methologies in favor of intelligent systems, characterized by data analytics, cloud computing and Artificial Intelligence, to achieve more effective resource management.

This thesis is developed within the framework of the "Gorgovivo 4.0" research project, a collaborative initiative involving the Gorgovivo Consortium, Viva Servizi S.p.A., and the Marche Polytechnic University. The project focuses on the development of an information system utilizing cloud compunting and Artificial Intelligence techniques to serve as a decision support tool for the Gorgovivo spring management.

The thesis addresses the data processing phase to produce functional and structured information reports. The work focuses on automating the processing of pluviometric data collected from stations serving the municipalities supplied by the Gorgovivo spring. Specifically, this research focuses on the development of Python-based scripts designed for the automated generation of technical reports. The project infrastructure leverages a cloud-native architecture, integrating AWS S3 for secure data storage and AWS Lambda for the periodic execution of analytical tasks. The logic of the developed system is divided into three stages: preprocessing of raw data acquired from the sensor network, data aggregation into daily, monthly, and annual intervals, and the calculation of supporting statistical indicators.

The result of the work, developed via Python and integrated with AWS services, is the systematic production of structured Excel reports, stored and updated daily on the cloud. These reports integrated summary tables, time series analysis, and graphical visualizations for accurate monitoring of temporal and seasonal trends in water levels.

The implemented solution ensures high scalability, resilience and operatio-

nal continuity, providing an effective decision-support tool for analyzing rainfall trends, detecting anomalies, and assessing aquifer recharge capacities.

Ultimately, this system serves as a technological foundation for a modern, transparent, and resilient governance of regional water resources, contributing directly to the sustainable management objectives defined by both European and national regulations.

Capitolo 1

Introduzione

L'acqua potabile è una risorsa essenziale per la sopravvivenza ed evoluzione degli ecosistemi, per la salute umana e per lo sviluppo sociale ed economico mondiale [1]. Tale risorsa è parte integrante del ciclo idrologico, il processo continuo che vede l'acqua circolare tra l'atmosfera e la terraferma. Attraverso le fasi di evapotraspirazione, precipitazione e scorrimento superficiale, l'acqua viene costantemente ridistribuita sul pianeta. Inoltre, la fase di infiltrazione permette la ricarica degli acquiferi sotterranei che fungono da serbatoi naturali per le sorgenti fondamentali del territorio. Nonostante il ciclo idrologico la renda una risorsa rinnovabile nel nostro pianeta, essa non è infinita. Le pressioni naturali e antropiche ne mettono a rischio l'equilibrio, rendendola limitata e vulnerabile. La sua disponibilità non è garantita né uniforme. Secondo il rapporto congiunto di OMS e UNICEF [2], circa 2,2 miliardi di persone, pari a un quarto della popolazione mondiale, non hanno ancora accesso a servizi di acqua potabile gestiti in modo sicuro.

La consapevolezza del valore di questa risorsa ha portato la comunità internazionale a riconoscerla non solo come risorsa ambientale, ma come bene sociale ed economico strategico. Tale visione è alla base dell'Agenda 2030 per lo Sviluppo Sostenibile dell'ONU [3], che pone l'obiettivo di garantire a tutti la disponibilità e la gestione sostenibile delle risorse idriche e dei servizi igienico-sanitari.

L'Organizzazione Meteorologica Mondiale (acronimo in inglese WMO) evidenzia come la crisi idrica sia il risultato di una combinazione di fattori climatici, ambientali e antropici [4]. Il surriscaldamento globale rappresenta uno dei principali fattori ambientali alla base dell'attuale crisi idrica. L'aumento delle temperature globali, attribuibile in larga parte all'incremento delle concentrazioni di gas dannosi per l'ambiente di origine antropica, altera il ciclo idrologico in maniera significativa, influenzando e riducendo la distribuzione spaziale e temporale delle risorse idriche superficiali e sotterranee. Gli eventi meteorologici osservano una crescente irregolarità, periodi di siccità sempre più prolungati alternati a precipitazioni intense e concentrate in brevi intervalli temporali. Tali dinamiche riducono l'efficacia dei processi di infiltrazione e ricarica delle falde acquifere.

Studi recenti indicano infatti che l'aumento dell'intensità delle piogge, a discapito di precipitazioni moderate e costanti, favorisce il ruscellamento superficiale impedendo all'acqua di filtrare efficacemente nel sottosuolo [5]. In questo senso il surriscaldamento globale agisce come acceleratore del ciclo idrologico poiché aumenta i tassi di evaporazione degli specchi d'acqua e favorisce lo scorrimento superficiale (run-off) e l'erosione del suolo, a discapito delle disponibilità di acqua sotterranea. Inoltre, l'innalzamento delle temperature medie sposta la quota dello zero termico, riducendo l'accumulo nevoso montano invernale, che funge da riserva idrica a lento rilascio per i mesi primaverili ed estivi.

Di conseguenza, si osserva un aggravamento delle condizioni di stress idrico in molte aree, con particolare intensità nelle regioni più vulnerabili dal punto di vista climatico, quali il bacino del Mediterraneo.

Ad una componente climatica e ambientale si affianca la pressione esercitata dalle attività umane. La crescita demografica, l'urbanizzazione incrementale e l'aumento della domanda di acqua in molteplici settori come quello agricolo o industriale hanno portato ad un uso non sostenibile di questa risorsa. Il settore agricolo, da solo, assorbe circa il 70% dei prelievi di acqua dolce a livello globale [6], evidenziando elevati coefficienti di dispersione idrica a causa di tecniche di irrigazione basate su infrastrutture carenti di sistemi di monitoraggio in tempo reale e di tecnologie per l'irrigazione di precisione. Il settore industriale, invece, causa una impermeabilizzazione del suolo impedendo un'adeguata e sicura ricarica degli acquiferi e aumenta l'inquinamento delle acque superficiali. Le cause di questi danni risiedono nei processi di soil sealing, ovvero i processi di trasformazione di aree naturali in aree urbanizzate, come la costruzione di stabilimenti industriali, di infrastrutture di trasporto o la cementificazione di grandi luoghi naturali che riducono la capacità di assorbire acqua da parte del terreno sottostante. Nel frattempo, aumenta la domanda di acqua della popolazione mondiale: secondo il World Water Development Report delle Nazioni Unite del 2024 [7], la domanda globale di acqua è destinata ad aumentare del 20-30% entro il 2050. Tali condizioni hanno causato uno sfruttamento eccessivo delle falde acquifere e delle risorse disponibili, portando ad un depauperamento quantitativo e qualitativo delle risorse superficiali e sotterranee e compromettendo la proprietà di rinnovabilità intrinseca dell'acqua.

A livello globale, la crisi idrica si manifesta con dinamiche eterogenee ma parimenti distruttive. Da un lato la siccità strutturale del Corno d'Africa, l'inaridimento dei grandi bacini del Sud-Ovest degli Stati Uniti (come il Lago Mead e il fiume Colorado) o la siccità pluriennale in California e dall'altro le "alluvioni del secolo" in Pakistan, le inondazioni in Germania e in Belgio del 2021 o i GLOF (Glacial Lake Outburst Floods) in Asia, ovvero la ceduta di laghi instabili creati dal troppo veloce scioglimento dei ghiacciai.

Non sono più eccezioni climatiche, ma segnali del cambiamento strutturale ed estremizzato del ciclo idrologico. Persino arterie vitali per il commercio mondiale, che simboleggiavano stabilità in questo senso, come il canale di Panama,

hanno subito limitazioni operative a causa dei bassi livelli idrici. Questi fenomeni evidenziano come la stabilità climatica e la grande disponibilità idrica mondiale siano sempre più a rischio e meno scontate.

Figura 1.1: Punti critici del monitoraggio dello stress idrico globale [4].

Il contesto italiano riflette pienamente queste criticità. Situata in un'area geografica particolarmente vulnerabile agli effetti del surriscaldamento globale, l'Italia sta sperimentando un incremento sia nella frequenza che nell'intensità delle emergenze idriche. Secondo l'Osservatorio Città Clima di Legambiente [8], infatti, l'Italia è diventata un "hot-spot climatico". Nei primi 5 mesi del 2023, gli eventi climatici estremi sono aumentati del 135% rispetto a quelli di inizio 2022. Le regioni di Sicilia, Puglia, Emilia Romagna e Marche sono le più soggette ad episodi di siccità prolungata e alluvioni intense e frane. Casi emblematici della gravità della situazione italiana sono le alluvioni di Emilia Romagna e Marche tra il 2022 e il 2023 o il ciclone Harry, che ha colpito duramente la Sicilia a inizio 2026 [9]. Inoltre, secondo analisi di EAI-ENEA (Energia, Ambiente e Innovazione)[10], la disponibilità di acqua nel Paese negli ultimi decenni ha riscontrato una diminuzione stimata intorno al 20%, con impatti rilevanti sulla sicurezza degli ecosistemi. A questa situazione si aggiungono anche criticità strutturali legate alla gestione del servizio idrico. I dati ISTAT mostrano che la rete idrica nazionale presenta una perdita intorno al 42,4% dell'acqua immessa nelle condutture e il trend è in leggera crescita (42,2% nel 2020 e 40,1% nel 2002)[11].

Per evitare una maggiore frequenza di crisi idriche è opportuno mettere in atto una serie di provvedimenti per poter preservare e gestire nel modo più opportuno il patrimonio idrico nazionale e mondiale. Uno dei passi che le Nazioni Unite hanno indicato ai singoli Paesi come soluzione per l'arginamento delle carenze

idriche riguarda l'applicazione di approcci integrati nel processo decisionale di gestione intelligente dell'acqua.

Gli approcci tradizionali, spesso basati sul solo bilancio tra domanda e offerta, risultano oggi inefficienti a fronte delle incertezze introdotte dal surriscaldamento globale e dalle attività antropiche a danno di questa risorsa. ENEA, con un report del 2023 sulla gestione delle risorse idriche [12], sottolinea l'importanza di orientare l'organizzazione verso la riduzione degli sprechi, l'ottimizzazione degli usi e l'adozione di strumenti avanzati di monitoraggio e pianificazione. Questa esigenza è delineata dalla normativa europea definita "Direttiva Quadro sulle Acque" [13], che stabilisce un approccio integrato alla tutela e alla gestione delle risorse idriche. La normativa si concentra sull'importanza del monitoraggio sistematico delle acque superficiali e sotterranee, richiedendo la raccolta, l'archiviazione e l'analisi continuativa di dati quantitativi e qualitativi, garantendone la coerenza, l'affidabilità e la confrontabilità nel tempo.

In questo senso, il quadro regolatorio si è progressivamente arricchito per affrontare la complessità del ciclo idrologico sotto diversi fronti. La direttiva 2006/118/CE [14], indirizzata alla protezione delle acque sotterranee, impone parametri rigorosi per prevenire e combattere l'inquinamento e il deterioramento chimico. Successivamente, nel 2007, il parlamento Europeo scrive una direttiva concentrata sulla valutazione e gestione dei rischi idraulici per combattere il fenomeno alluvioni [15]. In ambito nazionale, il D.Lgs. 152/2006 (Testo Unico Ambientale)[16] rappresenta il pilastro normativo che definisce i Piani di Gestione dei Bacini Idrografici, mentre il recente D.Lgs. 18/2023 [17] ha aggiornato i requisiti di qualità per le acque destinate al consumo umano.

In tutto il contesto regolamentare, la direttiva più importante e fondamentale anche da un punto di vista sociale e politico, risulta essere del 2007 (Direttiva INSPIRE 2007/2/CE) [15]; essa impone ai Paesi membri di garantire l'interoperabilità e la condivisione dei dati geografici e ambientali. Di conseguenza nasce la necessità di strumenti informatici in grado di trasformare dati grezzi in informazioni standardizzate e accessibili, fungendo da Decision Support Systems per una governance moderna, trasparente e resiliente della risorsa.

Di fronte a queste considerazioni, risulta evidente come il monitoraggio e la gestione integrata delle risorse idriche rappresentino elementi centrali per affrontare in modo efficace la pressante necessità di intervento evidenziata dall'attuale crisi idrica globale.

1.1 Obiettivi e motivazioni

Alla luce del contesto descritto, di emergenza idrica e rigore normativo, la presente tesi, inserita nel progetto Gorgovivo 4.0, sviluppato in collaborazione con Viva Servizi S.p.A. [18] e il Consorzio Gorgovivo [19], si occupa di supportare i processi di gestione, analisi e reportistica dei dati idrici relativi alla sorgente di Gorgovivo.

La struttura del progetto si fonda su una solida pipeline AWS capace di acquisire e centralizzare i dati grezzi provenienti dalla rete di sensori ambientali. Essa è caratterizzata dalla raccolta delle informazioni che vengono aggiornate giornalmente su cloud tramite gli automatismi definiti nell'infrastruttura di AWS. Tale approccio Cloud-native permette di gestire grandi volumi di dati eterogenei provenienti dalle reti di monitoraggio, trasformandoli in informazioni strutturate, sicure e facilmente interpretabili.

Il contributo descritto in questa tesi si è integrato nel progetto favorendo un'estensione delle funzionalità di reportistica tramite l'utilizzo dei servizi AWS all'interno dell'infrastruttura cloud presente. Nello specifico, ci si è focalizzati sull'automazione del reporting delle informazioni, agendo sui dati archiviati su cloud, contenenti le informazioni pluviometriche giornaliere. Da un punto di vista operativo, il lavoro portato avanti si è concentrato nell'introduzione di script Python che automatizzano l'elaborazione dei dati grezzi, finalizzati alla produzione di report strutturati in formato Excel. Questo strumento facilita le attività di monitoraggio e pianificazione tecnica, supportando gli ingegneri e i tecnici del Consorzio Gorgovivo [19] e di Viva Servizi [18] nel processo decisionale di gestione della risorsa idrica.

Tale automazione consente di organizzare i dati in modo coerente e standardizzato, superando le criticità e le inefficienze operative di un'analisi manuale, sempre più complessa a fronte della crescente quantità di dati. L'adozione di una tale soluzione comporta quattro vantaggi principali:

- frequenza di aggiornamento dei dataset, garantendo una visione del sistema idrica in tempo quasi reale
- uniformità dei dati
- ottimizzazione delle risorse, soprattutto da un punto di vista temporale
- accessibilità delle informazioni in uscita, facilitando le attività di lettura e analisi della situazione idrica.

In questo modo, il lavoro svolto fornisce un contributo concreto alle attività di monitoraggio, pianificazione e gestione sostenibile delle risorse idriche.

1.2 Struttura della Tesi

Dopo questo primo capitolo introduttivo, nel quale vengono presentati gli obiettivi principali del lavoro svolto e un'introduzione sul contesto sociale, economico e normativo, la tesi è strutturata nei seguenti capitoli:

2. Stato dell'arte: descrive il contesto territoriale sul monitoraggio idrico, il ruolo delle serie temporali ambientali e dei sistemi informativi di supporto decisionale e introduce il progetto Gorgovivo 4.0, all'interno del quale si colloca il lavoro di tesi.

3. Caso studio: Sorgente di Gorgovivo: fornisce un adeguata introduzione e descrizione del contesto ambientale e storico della sorgente e della sua gestione.

4. Tecnologie utilizzate e architettura software: descrive i principali strumenti software utilizzati nell'implementazione del sistema, sottolineando l'importanza e le motivazioni delle scelte progettuali (Python, AWS, servizi cloud).

5. Progettazione e sviluppo: illustra l'architettura generale del sistema, la pipeline di elaborazione dati e le caratteristiche più importanti del task implementato (costruzione dei report e meccanismi di automazione).

6. Risultati: presenta i risultati ottenuti osservando come esempio la stazione di Cupramontana.

7. Conclusioni: valuta il progetto, la sua efficacia, alcune limitazioni e i possibili sviluppi futuri.

Capitolo 2

Stato dell'arte

Una gestione sostenibile ed efficiente dell'acqua non può prescindere da un monitoraggio continuo e sistematico della risorsa. Esso rappresenta un elemento essenziale per la tutela degli ecosistemi naturali. Attraverso una corretta, efficiente e tecnologica gestione dell'acqua è possibile valutare lo stato delle risorse disponibili, individuare situazioni di criticità e supportare le politiche decisionali di intervento.

Le attività di monitoraggio sono condotte a differenti scale spaziali e temporali, in funzione delle caratteristiche del territorio e degli obiettivi di gestione. In ambito locale, il controllo è focalizzato su specifiche fonti di approvvigionamento, come sorgenti, pozzi e piccoli bacini idrici. In scala più ampia riguarda interi bacini idrografici con tempistiche di raccolta più distribuite. In entrambi i casi, la disponibilità di dati affidabili e aggiornati rappresenta il primo requisito per una corretta valutazione dello stato della risorsa.

Dal punto di vista operativo, il monitoraggio idrico si concentra sull'acquisizione di parametri di diversa natura; i dati raccolti vengono, infatti, organizzati e suddivisi in parametri quantitativi (livello di falde, portata, stress idrico), qualitativi (temperatura, pH, nutrienti, torbidità, contaminanti) e biologici (biodiversità, vita acquatica, erosione delle sponde dei fiumi). Tecnicamente, l'acquisizione dei dati avviene attraverso diverse tipologie di strumenti e sistemi che danno vita a reti di monitoraggio su larga scala.

L'evoluzione tecnologica, negli ultimi anni, ha consentito l'implementazione di sistemi SCADA (Supervisory Control And Data Acquisition) nei processi di monitoraggio, consentendo un'acquisizione dati in tempo reale, un controllo remoto e un'archiviazione ottimizzata per costruire serie temporali storiche fondamentali in analisi e modellazioni future. L'organizzazione generale dei sistemi SCADA di monitoraggio si sviluppa principalmente in: *monitoraggio in sito* (piezometri, aste idrometiche e sensori IoT per parametri qualitativi) e *monitoraggio remoto* (immagini satellitari e droni). Inoltre questi sistemi prevedono che i dati raccolti in sito e da remoto vengano trasmessi tramite sistemi di telemetria e reti di

comunicazione dedicate, basate su tecnologie radio, reti cellulari o collegamenti IP.

La diffusione di reti di monitoraggio meteorologico e di sistemi così strutturati ha reso possibile l'integrazione dei dati idrologici con informazioni pluviometriche di precipitazioni. Questa relazione risulta particolarmente rilevante nello studio delle risorse idriche sotterranee, in quanto consente di avere una precisione maggiore nell'analisi delle fonti e di avere una visione completa sull'intero ciclo dell'acqua nel territorio. Nonostante i progressi tecnologici, i sistemi di monitoraggio delle risorse idriche presentano ancora delle criticità, in particolare per quanto riguarda la gestione e l'utilizzo efficace dei dati raccolti. Una delle principali problematiche è rappresentata dall'elevata eterogeneità delle informazioni raccolte, dovuta all'impiego di diverse tipologie di sensori e infrastrutture differenti; i dati possono quindi presentare formati, strutture e frequenze di acquisizione non uniformi, rendendo complessa la loro integrazione all'interno di un unico sistema informativo coerente.

Un'ulteriore criticità risiede nella discontinuità delle serie temporali. Interruzioni nel funzionamento dei sensori, problemi di trasmissione, manutenzioni o errori di misura possono generare dati mancanti o non affidabili, che richiedono operazioni di validazione e pulizia prima di poter costituire la base dell'analisi idrica. In assenza di strumenti automatizzati, una tale problematica può risultare dispendiosa e poco efficiente.

Inoltre, la crescente diffusione di sistemi di raccolta dati e l'aumento della frequenza di campionamento hanno portato a una produzione di dati sempre più elevata. Una così grande quantità di dati può rappresentare un elemento critico se non correttamente gestita, ponendo quindi un'archiviazione corretta ed efficiente tra le sfide importanti per i sistemi informativi di gestione. Da un punto di vista operativo, una difficoltà consiste nella trasformazione di dati grezzi provenienti dai sensori in dati utili al supporto decisionale. Senza una forma strutturata o un'elaborazione dei dati singoli, il risultato è una analisi parziale e poco utile alla gestione delle fonti. Risulta quindi fondamentale utilizzare ed elaborare i dati provenienti dalle infrastrutture di controllo per produrre indicatori significativi alla gestione e facilmente interpretabili (trend temporali, cumulate mensili, percentuali di miglioramento o peggioramento rispetto allo storico, messa a punto di anomalie ricorrenti).

Il lavoro di reportistica, che in molti contesti applicativi viene ancora realizzato tramite procedure manuali e non informatizzate, diventa un punto centrale per eliminare o ridurre al minimo tutte le criticità legate a questo tipo di sistemi di controllo. Senza adeguati strumenti di elaborazione, infatti, i dati acquisiti vengono esposti a maggiori problemi di ripetibilità, standardizzazione e aggiornamento automatico. Nasce quindi la necessità di impiegare soluzioni informatiche dedicate, capaci di automatizzare i processi di gestione, elaborazione e sintesi dei dati ambientali, migliorando l'affidabilità e creando un vero e proprio strumento di supporto decisionale efficace.

2.1 Serie temporali ambientali

Una serie temporale è una sequenza discreta di osservazioni, indicizzate rispetto al tempo, che rappresentano l'evoluzione di una determinata grandezza fisica o logica. Le serie temporali costituiscono una delle principali modalità di rappresentazione dei dati nei sistemi di monitoraggio, controllo e automazione, in quanto consentono di descrivere il comportamento dinamico di un sistema nel tempo. In ambito ingegneristico, il tempo assume il ruolo di variabile indipendente fondamentale e la risoluzione temporale, con cui le osservazioni vengono acquisite, influenza direttamente la qualità e il tipo di analisi effettuabile. La frequenza di campionamento, definita come l'intervallo temporale tra due misure successive, deve essere scelta in funzione della dinamica del fenomeno osservato e degli obiettivi di analisi, in quanto una frequenza troppo bassa può comportare una perdita di informazione, mentre una frequenza troppo elevata può generare volumi di dati difficilmente gestibili.

I dati raccolti dai sistemi di monitoraggio ambientale assumono naturalmente la forma di serie temporali, in quanto rappresentano l'evoluzione temporale di parametri fisici, chimici e biologici osservati in relazione alle risorse idriche e alle fonti di approvvigionamento. In ambito idrico, le serie temporali rappresentano uno strumento fondamentale per la rappresentazione dello stato e dell'evoluzione dei fenomeni idrologici. Esempi ricorrenti in questo contesto sono i livelli delle falde acquifere, le portate dei corsi d'acqua e le misure di precipitazioni, acquisiti con frequenze giornaliere, settimanali o mensili. La scelta della risoluzione temporale influenza direttamente la tipologia di analisi effettuabile e il livello di dettaglio con cui è possibile interpretare i fenomeni osservati.

Una caratteristica distintiva delle serie temporali è la presenza di elevata variabilità e di fenomeni non stazionari. I dati, infatti, possono essere influenzati da fattori stagionali dovuti a cicli naturali, da trend di breve o lungo periodo o comportamenti anomali legati ad eventi estremi di origine naturale o antropica. La stagionalità, in particolar modo, rappresenta un elemento chiave nell'analisi di sistemi idrici focalizzati su dati pluviometrici, poiché la distribuzione delle piogge nel corso dell'anno incide direttamente sui processi di ricarica delle falde e sull'andamento delle risorse idriche sotterranee.

Oltre ai fenomeni naturali, le serie temporali ambientali presentano ulteriori criticità legate alla qualità dei dati. Rumore di misura, valori anomali e dati mancanti sono problematiche frequenti, dovute a malfunzionamenti dei sensori, interruzioni di servizio o errori di acquisizione. Tali aspetti rendono necessarie operazioni di validazione, filtraggio e ricostruzione delle serie prima che i dati possano essere utilizzati in modo affidabile per l'analisi e la gestione delle risorse idriche. In questo contesto, un aspetto importante è rappresentato dall'eterogeneità temporale delle serie. Parametri diversi possono essere acquisiti con frequenze differenti e in intervalli di tempo non allineati. Questa frammentazione e incoerenza temporale richiede operazioni necessarie di sincronizzazione, aggregazione e

riorganizzazione che devono essere tenute in considerazione per un sistema informativo dedicato. Affinché le serie temporali siano effettivamente una risorsa utile al supporto di processi decisionali, risulta fondamentale trasformare i dati grezzi in informazioni sintetiche, efficaci e facilmente interpretabili. Sulla base di ciò, le operazioni più importanti, a livello idrico e ambientale, risiedono nell'aggregazione temporale mensile o annuale, nel calcolo di indicatori statistici dedicati e nella produzione di report e grafici in grado di evidenziare gli andamenti significativi della risorsa in analisi.

L'analisi delle serie temporali ambientali, infatti, è utilizzata per diverse finalità: il monitoraggio dell'andamento storico delle grandezze osservate; l'individuazione di trend a medio e lungo termine; la valutazione delle relazioni tra variabili diverse. La reportistica assume, quindi, un ruolo fondamentale. Essa rappresenta un punto di collegamento tra i dati grezzi acquisiti dai sistemi di monitoraggio e le attività di analisi, valutazione e supporto alle decisioni.

2.2 Sistemi informativi per la gestione delle risorse idriche

All'interno di questo contesto diventa necessario sviluppare il tema dei sistemi informativi strutturati per la gestione delle risorse idriche. Essi rappresentano la naturale evoluzione dei sistemi di controllo e archiviazione, consentendo di trasformare una gran quantità di dati grezzi acquisiti dalla rete sensoristica ambientale in informazioni utili per la gestione operativa.

Il processo di digitalizzazione di tali sistemi ha seguito un'evoluzione importante negli ultimi anni, passando da sistemi locali e isolati a piattaforme informative integrate.

Da un lato, nei sistemi isolati, ciascun ambito di monitoraggio (livello delle falde, portate, qualità delle acque, precipitazioni) e, di conseguenza, i dati provenienti da ogni ambito vengono gestiti attraverso strumenti indipendenti con un bassissimo grado di interoperabilità. I dati vengono conservati in archivi locali, con poche possibilità di integrazione o condivisione. Tutto questo descrive un approccio poco efficiente per grandi volumi di dati e rende naturalmente complesso tutto il processo di analisi delle informazioni.

Le piattaforme integrate, dall'altro lato, mirano a superare le criticità legate alla frammentazione e la gestione locale delle informazioni. Diventa fondamentale la centralizzazione dei dati all'interno di un unico sistema informativo. Le architetture informatiche e le tecnologie di storage si focalizzano sull'utilizzo di database centralizzati e/o piattaforme cloud, middleware di acquisizione dati, normalizzazione dei formati e sincronizzazione temporale e uso di protocolli standardizzati, specialmente sul lato applicazione e web services. I punti fondamentali alla base di queste soluzioni sono la scalabilità e la flessibilità [20] [21].

All'interno di sistemi informativi in questo contesto è fondamentale distinguere i livelli applicativi e funzionali che costituiscono l'intero sistema. Una adeguata suddivisione è così composta:

- Archivio dati: rappresentano il livello più basso di gestione. La loro unica e fondamentale funzione risiede nel conservare in modo strutturato i dati grezzi, garantendo tracciabilità e accesso ai dati storici.

- Sistemi di reportistica: introducono un primo livello di complessità ed elaborazione dati. Si occupano di sintetizzare le serie temporali attraverso indicatori statistici, aggregazioni temporali e confronto con valori storici di riferimento.

- Dashboard informative: rappresentano l'evoluzione dei report e consentono una visualizzazione dinamica, chiara e interattiva dei dati. Fanno parte di questo livello tutte le mappe, grafici e tutti i sistemi di esposizione funzionale dei report.

- Decision Support System: costituiscono il livello più avanzato; integrano basi di dati strutturate, modelli matematici e strumenti avanzati di analisi; permettono la valutazione di scenari alternativi, l'analisi dell'evoluzione futura e l'individuazione di possibili situazioni critiche con anticipo; garantiscono l'ultimo gradino di supporto strategico.

Esistono diverse tipologie di sistemi e di strumenti informativi che ricoprono un ruolo fondamentale nella gestione della risorsa e nell'evoluzione tecnologica a supporto di questo contesto. I principali strumenti informativi sono i Geographical Information Systems (GIS) e i Advanced Metering Infrastructures (AMI). Sono strumenti che ricoprono esigenze differenti nella gestione dell'acqua, ma sono spesso complementari in molti sistemi e progetti. I GIS sono strumenti orientati alla gestione e all'analisi di dati georeferenziati e costituiscono la base per la rappresentazione spaziale delle infrastrutture idriche e del contesto territoriale. Consentono di descrivere la distribuzione delle reti sensoristiche, localizzare i punti di monitoraggio e analizzare le relazioni tra risorsa idrica, infrastrutture e territorio. Sono fondamentali in attività di pianificazione e manutenzione del bacino idrico.

I sistemi AMI, invece, sono focalizzati sulla raccolta automatica e continua dei dati attraverso contatori intelligenti e reti di comunicazione dedicate. Producono serie temporali ad alta risoluzione che permettono di monitorare l'utilizzo della risorsa nel tempo, individuare perdite o anomalie.

Da un lato, quindi, i sistemi GIS forniscono una struttura informativa semistatica, a lento aggiornamento, basati su una dimensione spaziale; dall'altro i sistemi AMI rappresentano la fonte più dinamica che riguarda i dati, di continuo aggiornamento e orientati ad una dimensione più temporale. L'integrazione di due strumenti del genere consente di associare i dati temporali di monitoraggio continuo alla loro collocazione territoriale, migliorandone significativamente l'analisi e la gestione.

Nonostante i benefici e la naturale evoluzione di questi sistemi, ci sono ancora molte criticità da affrontare sul tema. Come prima problematica non è possibile non citare la mancanza di standard tecnici condivisi: non esistono formati di dati uniformi e protocolli comuni. L'integrazione tra sistemi sviluppati da fornitori diversi e l'interoperabilità tra enti e piattaforme risulta ancora molto complessa. Un'altra criticità è legata all'uso, ancora diffuso, di strumenti manuali o applicazioni locali che ostacolano notevolmente l'automazione dei processi e aumentano il rischio di errori, duplicazioni e/o perdita di informazione utile. Il limite principale causato risiede nell'efficienza di aggiornamento. Come punto fondamentale, inoltre, è importante stabilire che la gestione di grandi volumi di dati rappresenta una grande sfida, che deve essere fronteggiata con infrastrutture adeguate e capaci di offrire sicurezza per tutti i dati archiviati. Infine, la frammentazione delle competenze e responsabilità tra soggetti, istituzionali e non, può compromettere

o danneggiare l'efficacia complessiva di tali sistemi, anche di fronte a strumenti tecnologicamente avanzati.

Negli ultimi anni, è importante dire che le politiche e i finanziamenti nazionali ed europei stanno favorendo e accrescendo la diffusione di sistemi informativi avanzati. I processi di digitalizzazione delle reti di monitoraggio idrico vengono eletti come punti importanti dell'agenda politica locale e l'obiettivo di miglioramento per la pianificazione e gestione idrica assume un ruolo sempre più significativo nel panorama socio-economico. L'attenzione continua su questo tema porta ad uno studio intenso sulle sfide e le criticità che devono essere affrontate da parte degli sviluppatori di sistemi avanzati. Lo sviluppo tecnologico e la progettazione di piattaforme integrate e intelligenti permettono un grande spazio di crescita per la digitalizzazione di questo contesto e una futura gestione più efficiente dell'acqua.

Nel panorama internazionale stanno nascendo numerosi progetti di sistemi avanzati; uno di questi è Digital OneWater [22], proposto da Jacobs, società d'ingegneria statunitense che offre vari servizi sia di consulenza che di realizzazione di progetti tecnici in ambito industriale, commerciale e governativo. Il progetto Digital OneWater vuole rappresentare un ecosistema integrato di soluzioni digitali per la gestione dell'intero ciclo dell'acqua. La piattaforma propone l'uso di tecnologie avanzate quali digital twin e machine learning per fornire insights predittivi e migliorare la gestione delle emergenze.

In ambito nazionale, invece, un progetto di grande rilievo è dell'Istituto Superiore di Sanità, con il supporto di istituzioni nazionali. Il progetto è AnTeA (Anagrafe Territoriale Dinamica delle Acque potabili) [23], piattaforma digitale concepita per l'acquisizione, il controllo e la gestione dei dati sulle acque potabili. Nasce dal D.Lgs 23 febbraio 2023 [17] con gli obiettivi di acquisizione, elaborazione, analisi e condivisione; si concentra sulla comunicazione, integrazione e accesso pubblico delle informazioni; assicura disponibilità e accessibilità delle informazioni anche alla Commissione Europea, all'Agenzia Europea per l'Ambiente e al Centro Europeo per la prevenzione delle malattie, come indicato dalla direttiva INSPIRE (2007/2/CE) [15] per la condivisione dei dati geografici ambientali tra enti pubblici. La piattaforma risponde quindi alla necessità di standardizzazione dei dati in un sistema condiviso e interoperabile, favorendo una visione complessiva e trasparente della qualità e disponibilità di acqua sul territorio nazionale.

2.3 Contesto territoriale

La Regione Marche presenta un profilo idrico caratterizzato da una distribuzione territoriale delle risorse non omogenea, fortemente influenzato dalla dorsale appenninica. Le zone interne presentano sorgenti e corsi d'acqua a elevata portata e buona qualità dell'acqua, minimizzando la necessità di trattamenti chimici estesi per l'uso potabile. Queste fonti alimentano la rete idrica regionale, rappresentando i principali apporti stabili per l'approvvigionamento potabile. In questo territorio un aspetto importante è caratterizzato dalla sua condizione idrogeologica. La risorsa idrica nella regione è fortemente sensibile alla stagionalità delle precipitazioni, ai fenomeni di siccità e ad una domanda potenziale maggiore rispetto alle disponibilità superficiali nei periodi siccitosi.

Negli ultimi anni, la Regione Marche ha sperimentato una progressiva intensificazione degli eventi meteorologici estremi, tali da influenzare in modo diretto la disponibilità di acqua nell'intero territorio. In pratica il territorio marchigiano ha affrontato sia periodi prolungati di siccità, con conseguente riduzione delle portate di sorgenti, corsi d'acqua superficiali e captazioni sotterranee, sia eventi estremi di precipitazioni intense e alluvioni, che mettono sotto pressione la rete di drenaggio e l'intera gestione della risorsa. Un caso evidente e fondamentale è rappresentato dall'alluvione che ha colpito il territorio marchigiano nel settembre 2022, in particolare le province di Ancona e Pesaro Urbino. Le forti precipitazioni hanno condizionato l'esondazione del fiume Misa, del Cesano e del fiume Esino. Condizioni di caldo estremo dell'estate 2022 hanno compromesso il territorio, fatto prendere largo ad un temporale autorigenerante e condizionato l'incapacità del suolo di assorbire fino a 400mm di acqua in 6 ore. Dati pubblicati dalla Regione Marche, tramite i registri della protezione civile marche, sottolineano la vulnerabilità idrogeologica e la ricorrenza di lunghi periodi siccitosi, delle piene improvvise dei corsi d'acqua superficiali e di altri fenomeni estremi che diventano subito casi studio per l'impatto idrico-ambientale [24].

L'evoluzione climatica si traduce in impatti concreti sulla gestione della risorsa in questa regione:

- riduzioni temporanee delle portate delle sorgenti ai pozzi (aggravano la criticità e aumentano la pressione su grandi fonti come la sorgente di Gorgovivo)
- incremento della domanda di risorsa durante periodi di siccità (sovraccarica i sistemi di distribuzione e il consumo domestico, agricolo e industriale)
- maggiore rischio idrogeologico (aumento di interventi di emergenza e gestioni straordinarie delle acque).

Il contesto sopra descritto ha portato a interrogazioni politiche regionali e comunali inerenti alla situazione di emergenza idrica del territorio e alle interruzioni del servizio in aree gestite da enti locali, che hanno evidenziato criticità strutturali

e operative del servizio idrico regionale e reso nota la necessità di aggiornamento dei piani di gestione idrica.

L'intero sistema idrico regionale riesce comunque a mitigare gli effetti di condizioni climatiche e ambientali sfavorevoli facendo affidamento a risorse idriche sotterranee stabili e ben protette, in particolare la dipendenza dalle grandi sorgenti carsiche come la sorgente di Gorgovivo, considerata strategica ed essenziale per garantire un adeguato approvvigionamento potabile e mitigare le criticità derivanti dai fenomeni di stress idrico.

Tuttavia, la forte dipendenza del sistema idrico regionale da questa singola sorgente comporta criticità gestionali. Negli ultimi anni essa è stata oggetto di studi idrogeologici, geochimici e di progetti di monitoraggio avanzato a supporto delle decisioni operative. Le sfide di gestione sostenibile di Gorgovivo in questo contesto territoriale assumono un'importanza emblematica per molti comuni della regione.

2.4 Progetto Gorgovivo 4.0

Il Progetto Gorgovivo 4.0 nasce proprio con l'obiettivo di rispondere alle esigenze di gestione efficiente e sostenibile della sorgente. Esso si configura come un percorso di ricerca applicata e sviluppo tecnologico orientato all'applicazione di un sistema informativo evoluto, basato su dati giornalieri, modelli previsionali e strumenti digitali di supporto decisionale [25] [26] [27]. Il progetto è stato avviato nel 2021 grazie alla collaborazione tra il Consorzio di Gorgovivo [19], Viva Servizi S.p.A. [18] e l'Università Politecnica delle Marche [28].

L'idea alla base del progetto è superare un approccio puramente reattivo alla gestione della risorsa, fondato sull'osservazione ex-post delle condizioni idriche, introducendo una logica predittiva e preventiva, capace di supportare le politiche decisionali e le scelte operative di pianificazione. L'evoluzione del sistema gestionale della sorgente è resa possibile dalla disponibilità di grandi volumi di dati storici e in tempo reale, provenienti dal sistema SCADA, dalle stazioni pluviometriche e dai monitoraggi idrogeologici. Si inseriscono, inoltre, tecniche avanzate di analisi delle serie temporale e di Intelligenza Artificiale. Il progetto mira alla realizzazione di un sistema integrato di monitoraggio in grado di stimare la produttività della sorgente e il comportamento delle acque sotterranee in relazione alle condizioni meteorologiche e idrologiche. Gli obiettivi secondari del progetto riguardano l'integrazione e la standardizzazione dei dati storici e real-time, il miglioramento delle capacità previsionali dei livelli di falda e delle portate, lo sviluppo di strumenti di simulazione di scenari estremi e l'ottimizzazione energetica dei sistemi di captazione.

Da un punto di vista architetturale, il progetto si basa su una struttura a più livelli. Il livello più basso è costituito dal sistema fisico: sorgente, pozzi, opere di captazione, rete sensoristica. Il secondo livello riguarda i dati: dati storici pluridecennali e dati real-time. Il livello successivo fa riferimento alla fase di elaborazione dei dati: infrastrutture cloud, algoritmi di analisi, elaborazione e modellazione dati tramite realizzazione di report funzionali. Infine, il livello di supporto decisionale: strumenti di visualizzazione, simulazione e previsione (tramite dashboard e interfacce dedicate, modelli previsionali e Intelligenza Artificiale).

L'importanza del progetto riguarda in parte aspetti idrogeologici. La complessità della falda acquifera, caratterizzata da risposte non lineari alle precipitazioni e da tempi di ricarica variabili, rende difficile la gestione dei dati che non può affidarsi a semplici soglie o regole empiriche. Alla base del progetto si colloca quindi la piena comprensione del funzionamento del sistema idrico e della correlazione tra precipitazioni, livelli di falde e portate. Inoltre, il progetto si concentra sull'innovazione tecnologica che si mantiene costante nel tempo attraverso la centralizzazione dei dati e delle analisi che contribuisce alla conservazione del know-how tecnico e alla continuità operativa. Lo sviluppo tecnologico del progetto fa particolare riferimento all'integrazione tra sistemi SCADA, infrastrutture cloud, strumenti di data analytics e modelli predittivi di Intelligenza Artificiale.

Gorgovivo 4.0 si colloca così all'interno del paradigma dei Digital Water Systems, non solo si limita a fornire un quadro descrittivo dello stato della sorgente, ma introduce una nuova modalità di interazione tra dati, modelli e decisioni operative attraverso strumenti di ingegneria dell'informazione.

Alla luce del contesto descritto, lo sviluppo del progetto Gorgovivo 4.0 si configura come un'iniziativa strutturale e progressiva, articolata in diverse fasi di sviluppo che affiancano e supportano i vari livelli architetturali del progetto. L'evoluzione del progetto è riconducibile ad un processo modulare in cui ciascuna fase contribuisce a consolidare la successiva, aumentando progressivamente il livello di conoscenza del sistema idrico e la capacità di previsione del suo comportamento. In particolare, le prime fasi sono dedicate all'acquisizione, integrazione e standardizzazione dei dati eterogenei provenienti dalle diverse fonti di monitoraggio. Le successive si concentrano sull'elaborazione automatica delle informazioni e sullo sviluppo di strumenti di visualizzazione e di analisi predittiva e simulazione. Questo approccio consente di affrontare in modo sistematico la complessità del sistema idrico, garantendo flessibilità e scalabilità del processo di sviluppo.

Il progetto iniziale disponeva già di un'architettura definita per la raccolta giornaliera dei dati provenienti dalla rete sensoristica, insieme alla progettazione generale della pipeline di elaborazione in ambiente cloud AWS. Il lavoro di tesi si è inserito in questo contesto, concentrandosi sulla gestione, organizzazione e trasformazione dei dati raccolti al fine di produrre report Excel strutturati, coerenti e funzionali alle successive fasi del progetto.

2.5 Task di reportistica

All'interno dell'architettura tecnologica del progetto Gorgovivo 4.0, un ruolo centrale è svolto dal task di reportistica, che costituisce l'ambito applicativo su cui si concentra la presente tesi. L'elaborazione di dati strutturati deve consentire l'analisi della gestione della sorgente, considerando la situazione e lo stato di ogni stazione pluviometrica inclusa nel sistema di monitoraggio, consentendo un'analisi dettagliata e localizzata delle dinamiche meteorologiche e dell'impatto sul sistema idrico.

La produzione di report per le varie stazioni pluviometriche si occupa di trasformare il dato grezzo proveniente dai sensori pluviometrici in informazioni sintetiche e strutturate, permettendo di valorizzare l'informazione spaziale associata ai dati di precipitazione, evidenziando differenze significative tra aree del sistema e supportando l'analisi e la correlazione tra precipitazioni, ricarica delle falde e risposta della sorgente.

Il task di reportistica rappresenta l'anello di congiunzione tra l'analisi tecnica dei dati e il loro utilizzo operativo dagli strumenti di visualizzazione e modellazione predittiva. In particolare si colloca alla fine della fase di acquisizione dati. Il lavoro sviluppato nel task attinge a dati aggiornati quotidianamente provenienti dalla rete sensoristica di ogni stazione pluviometrica e si occupa di elaborare e archiviare correttamente ed efficientemente i dati analizzati ed elaborati. Operativamente, si traduce in un processo strutturato di sintesi, interpretazione e restituzione delle informazioni; utilizza i dati pluviometrici per ricavare indicatori statistici e rappresentazioni grafiche sui trend stagionali, aggregazioni annuali o mensili, pattern ricorrenti o anomalie per ogni stazione.

Un requisito fondamentale del task risiede nella fruibilità delle informazioni. La struttura di ogni stazione aiuta la standardizzazione del formato dei dati, consentendo di produrre report omogenei tra diverse aree e mantenere un rigore tecnico nell'elaborazione dei dati, contribuendo anche all'uniformità delle modalità di analisi e di valutazione dei dati per una continuità operativa e gestionale nel tempo. L'aspetto più importante per una gestione sostenibile ed efficiente dei dati in analisi è l'automazione dei processi di creazione e aggiornamento dei report.

La presente tesi si concentra quindi sullo sviluppo e sull'implementazione di script python che possa generare automaticamente dei report di analisi e gestione dei dati per ogni singola stazione pluviometrica e possa integrare efficientemente i report all'interno dell'architettura software del progetto.

Capitolo 3

Caso studio: Sorgente di Gorgovivo

La sorgente di Gorgovivo è la principale infrastruttura di approvvigionamento idrico potabile per una vasta porzione della Regione Marche. La sorgente emerge da formazione calcarea fratturata ed è alimentata da un bacino di circa 200 km^2, collocato a ridosso del Monte San Vicino, e le portate misurate oscillano tra circa 1.800 e 4.000 litri al secondo, favorendo un'elevata produttività. Alimenta un bacino di utenza estremamente ampio, fungendo da fonte primaria per la distribuzione di acqua potabile per decine di comuni di Ancona e Macerata.

La captazione di Gorgovivo, di tipo sotteraneo, è basata su un articolato sistema di pozzi profondi gestito da Viva Servizi S.p.A. [18]. Ogni pozzo funge da punto di estrazione primaria, la cui gestione è demandata alle stazioni di monitoraggio locali. Esse acquisiscono e forniscono quasi in tempo reale parametri fondamentali di portata e qualità dell'acqua per ciascun pozzo.

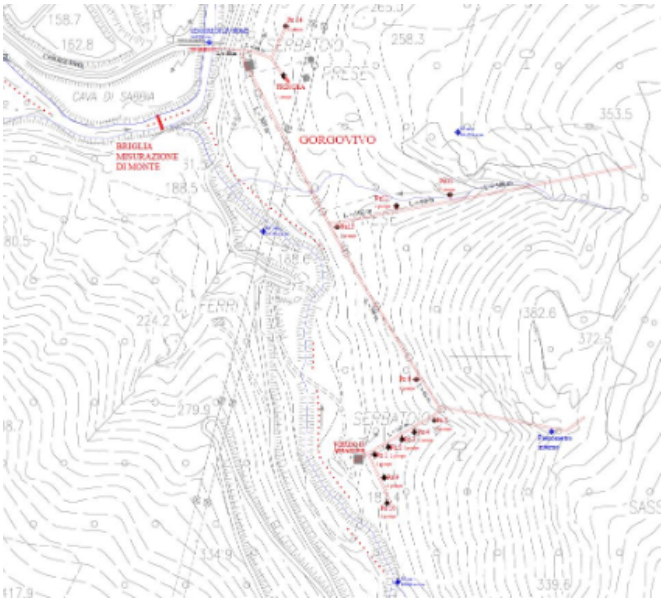

Figura 3.1: Localizzazione dei pozzi piezometrici.

La seconda importante rete che fornisce dati utili al progetto fa riferimento alle stazioni pluviometriche della Protezione Civile presenti nei comuni di tutto il bacino di riferimento per la sorgente di Gorgovivo. Le stazioni consentono di monitorare i dati di precipitazione nel territorio interessato, al fine di fornire una notevole quantità di informazione di regime di acqua precipitata nel sottosuolo in giorni, mesi e anni, favorendo la precisione del bilancio idrico dell'area interessata dalla sorgente.

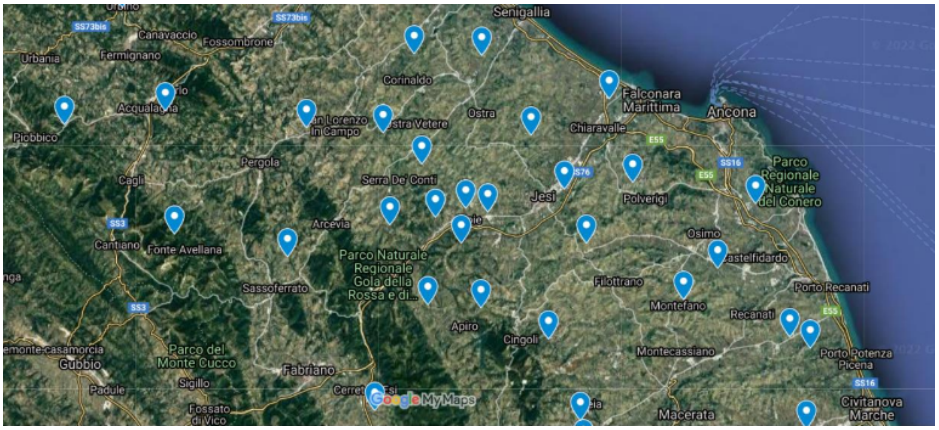

Figura 3.2: Stazioni pluviometriche di riferimento nella regione Marche.

L'importanza di Gorgovivo riguarda anche la qualità intrinseca dell'acqua captata. La sorgente presenta caratteristiche chimico-fisiche che rispettano numerosi standard di potabilità delle acque tramite la naturale filtrazione delle rocce.

Il complesso sorgentizio di Gorgovivo rientra tra i beni amministrati dal Consorzio Gorgovivo [19], mentre Viva Servizi S.p.A. [18] svolge il ruolo di gestore operativo del servizio idrico. Il Consorzio, composto da venti Comuni di cui diciotto in provincia di Ancona e due in provincia di Macerata, è l'Ente proprietario dei beni del complesso sorgentizio di Gorgovivo e quindi ne cura l'esercizio e l'amministrazione. Ha come oggetto anche l'amministrazione degli impianti e delle reti di distribuzione dell'acqua, del gas metano, delle reti fognarie e degli impianti di depurazione delle acque reflue e di tutti gli impianti connessi facenti parte del patrimonio assegnato in proprietà indivisa dal Consorzio ai Comuni consorziati e conferenti.

Il Consorzio Acquedotto Valle dell'Esino (C.A.V.E.) fu costituito nel 1963 dai comuni di Ancona, Chiaravalle, Falconara Marittima, Jesi, Monsano, Monte San Vito, Montemarciano e Senigallia con lo scopo di costruire l'acquedotto che avrebbe trasportato l'acqua di sorgente ai comuni. Divenne poi gestore del ciclo completo dell'acqua, comprendente la captazione e la distribuzione dell'acqua potabile, la raccolta dell'acqua usata in fognatura e della depurazione nel territorio dei comuni associati. Le evoluzioni legislative in materia di servizi pubblici locali

hanno poi portato nel 2001 a scindere dal Consorzio la gestione del servizio idrico per affidarla alla società Viva Servizi S.p.A.[18].

Il lavoro di monitoraggio portato avanti da Viva Servizi S.p.A. [18] e dal Consorzio Gorgovivo [19] e l'intera rete sensoristica offrono i dati fondamentali per l'analisi e lo sviluppo del progetto. Riguardano essenzialmente i dati dei pozzi acquisiti tramite la rete sensoristica di piezometri e i dati recuperati dalle stazioni pluviometriche per valutare l'influenza delle precipitazioni.

Capitolo 4

Tecnologie utilizzate e architettura software

Il progetto Gorgovivo 4.0 e il task di reportistica si collocano in un contesto applicativo caratterizzato da un'elevata complessità informativa e operativa, in cui è richiesto l'utilizzo di strumenti tecnologici in grado di supportare l'intero processo di elaborazione del dato e di rispondere ad alcuni requisiti funzionali fondamentali.

Il requisito più importante è quello di poter automatizzare tutta la procedura di generazione dei report e renderla flessibile per ogni stazione. Indipendentemente dalla stazione di riferimento, le procedure devono poter gestire i dati nello stesso identico modo. L'automazione del processo deve garantire continuità operativa e risposta a errori o anomalie, garantendo un'analisi sostenibile nel tempo.

Da un punto di vista di mantenimento dei dati, l'accesso ai report deve essere flessibile, sempre disponibile, affidabile e sicuro. Il task deve consentire una corretta archiviazione delle differenti tipologie di informazione, riducendo il più possibile i rischi di lettura e accesso ai dati.

Infine, l'architettura dell'intero task deve essere organizzata in modo tale da separare logicamente i livelli di operatività del sistema: acquisizione dei dati, elaborazione in differenti indicatori e grafici, archiviazione ed integrazione con strumenti di visualizzazione e analisi predittiva del progetto generale.

4.1 Python per l'elaborazione e l'analisi dei dati

Python è un linguaggio interpretato, multiparadigma e caratterizzato da una sintassi chiara e leggibile che favorisce uno sviluppo rapido e una manutenzione efficiente del codice.

Python è stato adottato come linguaggio principale per l'elaborazione e l'analisi dei dati. Tale scelta risiede nelle necessità del task. Le esigenze di versatilità, affidabilità e manutenzione del codice hanno orientato la scelta su un linguaggio che offre una serie di librerie adatte, mature e specializzate nel data processing (gestione di grandi quantità di dati e integrazione con report excel), time series analysis (gestione di dati e file csv) e comunicazione con servizi cloud (integrazione con i servizi di Amazon Web Services). In questo contesto, inoltre, consente un'ottima continuità operativa per evoluzioni future e riutilizzo del codice.

L'implementazione del task di reportistica si basa sull'uso integrato di diverse librerie Python, per rispondere a specifiche esigenze funzionali della pipeline di elaborazione:

- **pandas:** consente di importare dati in formato CSV, strutturarli in DataFrame e manipolarli efficacemente per fogli di calcolo. Utilizzata nel task per il parsing dei file CSV contenenti i dati di precipitazione, per la conversione delle stringhe temporali in oggetti datetime, per la gestione dei dati mancanti o anomalie e per l'aggregazione dei valori su base mensile o annuale [29].

- **openpyxl:** permette di controllare ogni singolo aspetto di file .xlsx; consente di leggere, modificare e salvare file Excel mantenendo intatta la struttura dei fogli, delle formule e dei grafici. Utilizzata nel task per aggiornare i fogli dei dati giornalieri, compilare le celle e le tabelle mensili e annuali, inserire o estendere formule specifiche, garantire continuità storica aggiungendo nuove celle in specifici punti del file per l'aggiunta di mesi o anni [30].

- **BytesIO:** il modulo **io** fornisce un'interfaccia uniforme per gestire stream di dati. La classe **BytesIO**, fornita dal modulo **io**, consente, nello specifico, di trattare una sequenza di byte come fosse un file fisico che risiede interamente in RAM (creazione di un file virtuale in RAM senza scrivere file temporanei su disco). Nel task migliora l'efficienza del flusso di elaborazione, riducendo la latenza e garantendo un'adeguata portabilità del codice in ambienti cloud [31].

- **boto3:** è un sdk (software development kit) ufficale di AWS per Python. Strumento utile per permettere al codice di creare server, gestire database o caricare/scaricare file da S3 (Simple Storage Service) di Amazon. Nel task è utilizzata per accedere al bucket S3, scaricare i file CSV e caricare i report Excel e gestire in modo automatizzato il flusso di input e output

del processo. **boto3** è, molto probabilmente, la libreria più influente per quanto riguarda la sicurezza nella gestione dei file elaborati e la scalabilità dell'intero progetto [32].

- **datetime:** libreria standard per gestire i formati temporali, legati, principalmente, ai dati giornalieri raccolti.

La combinazione di queste librerie fornisce un punto centrale per gli strumenti necessari alla realizzazione di un processo di reportistica automatizzato, robusto e coerente con le esigenze operative del progetto.

4.2 Servizi cloud AWS per archiviazione e automazione

Il task di reportistica automatizzata richiede un'infrastruttura di gestione e archiviazione dati efficiente e disponibile. Utilizzare un servizio di cloud computing rappresenta, quindi, un'ottima scelta per la scalabilità e la continuità operativa nel tempo del codice e dei dati.

Il cloud computing introduce un paradigma di utilizzo delle risorse informatiche basato sull'erogazione di servizi on-demand attraverso una rete di risorse di calcolo scalabili e flessibili che consentono di separare la logica applicativa dalla gestione dell'hardware sottostante. Esistono vari modelli di servizio che caratterizzano il differente livello di astrazione dell'architettura del servizio cloud:

- Iaas (Infrastructure as a Service): i servizi IaaS rappresentano il livello più fondamentale di servizi cloud, offrendo servizi base come server virtuali e storage. Con questi servizi, l'utente ha pieno controllo sull'ambiente virtuale ed è capace di gestire e configurare le risorse a disposizione. Permette una grande flessibilità nel personalizzare l'infrastruttura sottostante. Esempi importanti sono Google Compute Engine, Amazon Elastic Compute Cloud e Amazon Simple Storage Service.

- PaaS (Platform as a Service): offre un ambiente di sviluppo completo e gestito per la creazione, il test e la distribuzione di intere applicazioni; gli sviluppatori possono concentrarsi esclusivamente sullo sviluppo del codice. Alcuni servizi Paas di esempio sono Google App Engine e Microsoft Azure App Service.

- FaaS (Function as a Service): permette l'esecuzione di un singolo frammento di codice; è un altissimo livello di astrazione per sviluppatori, in cui è possibile inserire la logica di business all'interno di funzioni discrete che vengono eseguite in risposta a eventi specifici. Una differenza con altri modelli di astrazione risiede anche nel costo del servizio che, in questo caso, è Pay-per-execution, ovvero si paga il servizio solo per il tempo di esecuzione del codice. Esempi importanti sono Google Cloud Functions e AWS Lambda.

- SaaS (software as a Service): è il livello più alto di astrazione; permette l'utilizzo di applicazioni complete gestite dal provider cloud; riguarda praticamente un'interazione con un'interfaccia che riduce al minimo il controllo tecnico del sistema, solitamente usato da utenti finali e non da sviluppatori. Utile per accedere a servizi pronti all'uso come Microsoft365, Microsoft Power BI e Google Workspace.

Amazon Web Services (AWS) è il principale fornitore di servizi di Cloud Computing al mondo. Rappresenta una filosofia e un insieme di tecnologie fondamentali per un nuovo paradigma che ha cambiato il modo di concepire e implementare le infrastrutture IT. AWS offre un'ampia gamma di servizi modulari che coprono diversi ambiti applicativi, dallo storage all'analisi dei dati o all'automazione di processi, tutto volto all'elaborazione in modalità serverless delle applicazioni. AWS favorisce numerosi vantaggi nel suo utilizzo; i principali e utili all'implementazione del task di reportistica risiedono nella maturità e supporto della piattaforma, nell'elevata affidabilità dei servizi offerti, nella forte integrazione con Python e nella disponibilità di soluzioni specifiche per l'archiviazione dei dati e l'esecuzione automatica di processi batch.

- **Amazon S3 - Simple Storage Service:** S3 è un servizio di cloud storage affidabile, semplice e flessibile. La metodologia di archiviazione è basata sui *bucket*, generici contenitori, in cui è possibile caricare oggetti contenenti i dati archiviati in diverse unità del disco fisiche all'interno di un data center. S3 funge da *repository* centrale, consentendo un accesso rapido e sicuro a dati importanti per l'analisi temporale. Nel progetto viene utilizzato per salvare i file CSV dei dati pluviometrici e i report Excel generati automaticamente.

Figura 4.1: AWS S3.

- **AWS Lambda:** Lambda è una piattaforma che permette la creazione di applicazioni *serverless* (senza necessità di gestire risorse hardware per eseguire determinate funzioni). Lambda, inoltre, permette di eseguire funzioni in risposta a eventi o secondo pianificazioni temporali, adattandosi a processi batch come elaborazioni periodiche. Nel progetto viene utilizzata per eseguire lo script Python periodicamente, una volta al giorno. L'esecuzione avviene in questo ambiente isolato e controllato di AWS, dove la funzione Lambda scarica i file da S3, elabora tutte le principali informazioni per ogni stazione e ricarica i report all'interno dello specifico bucket su S3 [33].

Figura 4.2: AWS Lambda.

4.3 Strumenti di visualizzazione e supporto decisionale

L'efficacia del task di reportistica e dell'intero progetto risiede anche nelle modalità di presentazione e restituzione delle informazioni agli stakeholders. Una particolare attenzione va infatti rivolta agli strumenti di Data Visualization. La loro funzione principale risiede nella trasformazione di dati grezzi in informazioni significative e renderle facilmente interpretabili all'esterno. Un requisito importante è di rendere accessibili i risultati dell'analisi a utenti non specialisti in ambito informatico, mantenendo però coerenza e affidabilità delle informazioni ricavate.

In tal senso lo strumento fondamentale utilizzato nel lavoro di tesi è Excel.

- **Microsoft Excel** rappresenta lo standard per strumenti di calcolo, analisi rapide e consultazione dei dati. I report Excel generati all'interno del task di reportistica costituiscono una sintesi strutturata delle informazioni. Vengono utilizzate rappresentazioni tabellari e grafiche per tenere traccia dell'andamento temporale dei dati e degli indicatori statistici di sintesi. Le funzioni più interessanti di Excel utilizzate nei report sono *STDEV*, deviazione standard per comprendere la stabilità del regime pluviometrico, e *PERCENTILE*, per il calcolo del ventesimo percentile e ricavare facilmente livelli di soglia ottimale e casi di anomalia.

Figura 4.3: Microsoft Excel.

Capitolo 5

Progettazione e sviluppo

Il presente capitolo descrive in modo dettagliato il processo di sviluppo e di implementazione del sistema realizzato per il task di elaborazione e reportistica dei dati pluviometrici. L'obiettivo operativo del task è la realizzazione di uno script Python in grado di acquisire ed elaborare i dati provenienti dalle stazioni pluviometriche di monitoraggio in maniera automatica ed autonoma. Lo script Python si occupa, quindi, della creazione, mantenimento e aggiornamento dei report strutturati in file excel, come base storica informativa del progetto Gorgovivo 4.0.

Lo script sviluppato segue pienamente il paradigma delle *data processing serverless pipelines*, modello architetturale per la gestione e il processamento di grandi volumi di dati che ha modernizzato il settore riducendo la complessità infrastrutturale per lo sviluppatore. Nel paradigma serverless, la pipeline di un tale sistema è di natura *event-driven*, cioè il sistema reagisce a dei trigger che invocano componenti di elaborazione e l'implementazione viene eseguita su richiesta in risposta a specifici eventi; inoltre, viene ridotto il costo attraverso una logica pay-per-use, che non addebita il tempo di inattività, favorendo una grande elasticità dinamica del sistema.

Le *data processing serverless piepelines* utilizzano servizi cloud per favorire flessibilità e scalabilità. Si fondano sull'integrazione di diversi servizi, ognuno dei quali svolge uno specifico ruolo all'interno del flusso di elaborazione. Nel caso particolare del task di reportistica in analisi, il sistema è concettualmente suddiviso in una serie di fasi logiche:

- ingestione: fase di intercettazione dei trigger (eventi che avviano l'esecuzione) e identificazione della sorgente dei dati.
- elaborazione: fase di trasformazione dei dati; parsing, validazione, aggregazione temporale e calcolo indicatori statistici.
- destinazione: scrittura dei risultati nei report strutturati.
- storage: conservazione dei dati nei sistemi di archiviazione persistente.

- governance e sicurezza: fasi di controllo accessi e operazioni effettuate sui dati; gestiti direttamente dai servizi cloud per tenere la tracciabilità delle esecuzioni del codice.

Questa impostazione definisce perfettamente la logica dell'architettura e del codice implementato nel task e verrà descritto da tre sezioni principali presenti nel capitolo. Nella prima viene descritta l'architettura generale del sistema, con riferimento all'organizzazione dei componenti software e al flusso logico delle informazioni (come mostrato in figura 5.1). La seconda sezione è dedicata alla pipeline di elaborazione dei dati, analizzando nel dettaglio le fasi di acquisizione dati in input, pre-elaborazione, aggregazione temporale ed elaborazione indicatori statistici e aggiornamento dei report (come mostrato in figura 5.2). Infine, nell'ultima sezione, vengono approfonditi gli aspetti legati all'automazione del sistema e le soluzioni operative tecniche vantaggiose per il sistema.

Figura 5.1: Flusso generale implementato per l'aggiornamento automatico del report Excel tramite AWS Lambda e S3.

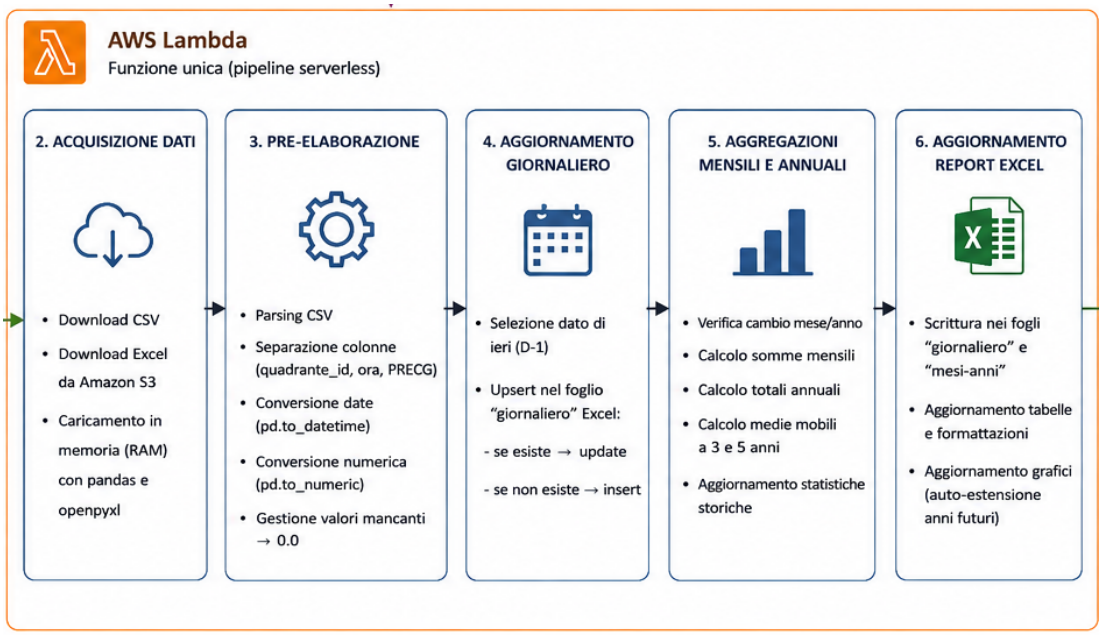

Figura 5.2: Pipeline di elaborazione dati.

5.1 Architettura generale del sistema

Da un punto di vista metodologico e architetturale ci si è concentrati sulla robustezza e solidità del codice di fronte a un'elevata variabilità temporale con frequenti aggiornamenti, possibili anomalie e future estensioni del sistema. La gestione delle serie temporali ha identificato la necessità di funzioni e operazioni che agissero su diverse scale temporali, mantenendo una solida coerenza del dato.

Inoltre l'automazione dell'intero script attraverso il servizi di AWS Lambda ha indicato la necessità di inserire l'intera pipeline in un'unica unità logica, nonostante le best-practises del cloud computing suggeriscano la scomposizione dei sistemi in microservizi indipendenti. L'adozione di una sola funzione contenitore favorisce la riduzione di latenza nell'esecuzione dell'intero script, riducendo la presenza di overhead. Si facilita inoltre la gestione dei vari stati dei file utilizzati durante l'elaborazione e permette di controllare in modo centralizzato l'intero ciclo di vita dei dati. Da un punto di vista dell'affidabilità viene favorita e migliorata la gestione di eccezioni ed errori, eliminando il rischio di inconsistenza nello storico informativo.

I componenti software del sistema possono essere descritti secondo una suddivisione logica che riflette le diverse responsabilità dello script. Un primo componente risiede nel modulo di accesso ai dati di ingresso, responsabile dell'individuazione e del caricamento dei file contenenti le misure pluviometriche delle varie stazioni (Fasi di ACQUISIZIONE DATI e PRE-ELABORAZIONE in figura 5.2). Questo modulo costituisce il punto di contatto tra l'elaborazione dei report e le sorgenti dei dati, assicurandosi che le informazioni siano correttamente interpretate e rese, perciò, efficientemente disponibili per le fasi successive. Un secondo componente è dedicato alla gestione e all'elaborazione delle serie temporali (Fasi di AGGIORNAMENTO GIORNALIERO e AGGREGAZIONI MENSILI E ANNUALI in figura 5.2). Il sistema si occupa delle funzioni necessarie a trattare i dati in aggregazioni temporali mensili e annuali. In ultima analisi, il sistema si occupa dell'aggiornamento dei report, operando in modalità incrementale, aggiornando le nuove osservazioni e preservando lo storico informativo.

5.2 Pipeline elaborazione dati pluviometrici

La pipeline di elaborazione dei dati pluviometrici in figura 5.2 rappresenta l'essenza del task di reportistica sviluppato. Il workflow seguito nell'implementazione rispecchia le caratteristiche di automazione e di aggiornamento incrementale dei dati. Lo script python implementato segue una sequenza ordinata di funzioni corrispondenti alle varie fasi di elaborazione, facilitando la leggibilità e la manutenibilità pur mantenendo l'intera logica all'interno di un'unica unità logica funzionale. La sequenza di operazioni è univoca per ogni stazione e, quindi, il codice Python itera sull'insieme delle stazioni di riferimento per un'elaborazione sistematica e ripetibile; rappresenta inoltre un punto importante per la scalabilità del sistema permettendo l'aggiunta di nuove stazioni senza modifiche sostanziali allo script.

5.2.1 Dati in ingresso

I dati in ingresso, ricavati dalla rete di monitoraggio pluviometrico, sono costituiti da due tipologie di file per ogni stazione, archiviate su un bucket di AWS S3:

- file CSV, contenenti le osservazioni giornaliere di precipitazione. Forniscono indicazioni sulla stazione da cui provengono, sul giorno di acquisizione e sul livello di acqua registrato.

  [notazione: *APC_ST{codice_stazione}_{nome_stazione}.csv*]

	A
8402	APC_ST1263;01/01/2024;1.6
8403	APC_ST1263;02/01/2024;0.2
8404	APC_ST1263;03/01/2024;0
8405	APC_ST1263;04/01/2024;0
8406	APC_ST1263;05/01/2024;0.4
8407	APC_ST1263;06/01/2024;4.6
8408	APC_ST1263;07/01/2024;10
8409	APC_ST1263;08/01/2024;1.6
8410	APC_ST1263;09/01/2024;0.6

Figura 5.3: dati pluviometrici della stazione di Cupramontana; gennaio 2024.

- report Excel, strutturati in più fogli di lavoro e contenenti le elaborazioni e le sintesi statistiche prodotte.

  [notazione: *{nome_stazione}_cod{codice_stazione}_CG*.xlsx]

Sono composti da tre fogli principali: 'giornaliero', contenente le informazioni dei dati pluviometrici giornalieri; 'mesi-anni', contenente la sommatoria di ciascun mese per ogni anno al fine di valutare e analizzare l'andamento temporale nel corso degli anni; 'cumulate mensili' che si concentra sulle cumulate di più mesi per ogni anno, al fine di valutare l'influenza della stagionalità nei dati raccolti.

Il primo passo per l'acquisizione dei file in input è, quindi, costituito dalla configurazione delle credenziali di S3 e creazione di una sessione sicura di accesso al bucket *gorgovivo-s3*.

Come mostra la figura 5.4, viene poi scaricato il report della stazione di riferimento; attraverso le librerie *BytesIO*, il flusso di byte (stream) proveniente dai file su S3 viene caricato in un buffer di memoria volatile (RAM); questo approccio evita operazioni di I/O su disco per lettura/scrittura e riduce la latenza nelle operazioni portate avanti dalla libreria `openpyxl`.

```python
# Funzione per scaricare il file Excel da S3
def download_excel_from_s3(bucket, key):
    try:
        obj = sess.get_object(Bucket=bucket, Key=key)
        return obj['Body'].read()
    except Exception as e:
        print(f"Errore durante il download del file da S3: {e}")
    return None

# Scarica il file Excel
file_content = download_excel_from_s3(bucket_name, file_key)
if not file_content:
    print(f"[ERRORE] Download fallito per il file: {file_key}")
    continue

# Carica il file Excel in openpyxl
try:
    wb = openpyxl.load_workbook(BytesIO(file_content))
except Exception as e:
    return {
        'statusCode': 500,
        'body': f"Errore durante il caricamento del file Excel: {str(e)}"
    }
```

Figura 5.4: Funzione per il download del file excel dal bucket di S3.

Allo stesso modo e con l'aiuto della libreria *Pandas* vengono estratti i dati del file `.csv` in una tabella virtuale su cui il codice potrà effettuare operazioni avanzate, come mostrato in figura 5.5.

```python
# Cerca un file che inizi con 'APC_ST' e contiene il codice nel nome
target_file_key = None
files = sess.list_objects_v2(Bucket=bucket_name, Prefix='APC_ST')
for file in files.get('Contents', []):
    if f'APC_ST{code}' in file['Key']:
        target_file_key = file['Key']
        break

if not target_file_key:
        print(f"[ERRORE] File CSV non trovato per la stazione di {nome_base}")
        return{
            'statusCode': 404,
            'body': f"File CSV non trovato per la stazione di {nome_base}"
        }

# Scarica il file corrispondente che è un CSV
data_target = download_excel_from_s3(bucket_name, target_file_key)
if not data_target:
    return {
        'statusCode': 500,
        'body': 'Errore durante il download del file CSV da S3.'
    }

# Carica il file CSV in pandas
df_target = pd.read_csv(BytesIO(data_target))
```

Figura 5.5: download file csv.

5.2.2 Pre-elaborazione

La fase di pre-elaborazione è cruciale per uniformare in maniera strutturata e coerente i dati grezzi provenienti dai file csv. I dati in ingresso presentano una struttura *Comma-Separated Values* con attributi suddivisi da un punto e virgola. Il dataset è, quindi, definito e strutturato attorno a tre variabili fondamentali che ne definiscono la dimensione spaziale, temporale e quantitativa. Gli attributi in questione sono:

- quadrante_id : rappresenta l'identificativo univoco della stazione di monitoraggio all'interno della rete pluviometrica; funge da chiave di ricerca per la tracciabilità della sorgente; consente di associare correttamente i dati grezzi al rispettivo report excel della stazione.
- ora : rappresenta il riferimento temporale a cui si riferisce la misurazione; indica la data esatta della precipitazione ed è fondamentale per tutte le aggregazioni temporali da sostenere e calcolare.
- PRECG : rappresenta la PREcipitazione Cumulata Giornaliera, espressa in mm; è il dato fisico misurato dai pluviometri e costituisce l'input numerico per le somme e i trend idrologici dei quantitativi di acqua che influiscono sulla stazione di riferimento.

L'intero processo di pre-elaborazione si articola in tre momenti e azioni principali:

1. parsing dell'informazione:

   ```
   df_target[[‘quadrante_id’, ‘ora’, ‘PRECG’]] =
   df_target[‘quadrante_id;ora;PRECG’].str.split(‘;’, expand=True)
   ```

   Lo script utilizza questa riga per eseguire una tokenizzazione basata sul separatore ‘;’. Tramite il metodo della libreria Pandas `.str.split()` lo script recupera i dati contenuti nell'unica stringa testuale dei file csv e distingue le informazioni in tre campi distinti, rendendo la data e il valore di livello di acqua caduta in entità atomiche e manipolabili singolarmente; così facendo ciascun attributo risulta associato a una colonna dedicata.

2. normalizzazione temporale:

   ```
   df_target[‘ora’] = pd.to_datetime(df_target[‘ora’],
   errors=‘coerce’, dayfirst=True)
   ```

   Per la gestione temporale, il sistema deve convertire le stringhe temporali in oggetti Python di tipo `datetime`. Viene invocata, perciò, la funzione `pd.to_datetime` con il parametro `dayfirst=True` per interpretare correttamente il formato data con il primo numero che rappresenta il giorno ed evitare ambiguità tra giorni e mesi e non tradurre la stringa in una data diversa da quella che indica effettivamente. Inoltre, l'uso del parametro `errors='coerce'` irrobustisce il sistema che elabora eventuali stringhe corrotte in valori `Null (NaT)` così da impedire il crash della pipeline, accettando una perdita di informazione corrotta.

3. validazione numerica:

   ```
   df_target[‘PRECG’] = pd.to_numeric(df_target[‘PRECG’],
   errors=‘coerce’).fillna(0.0)
   ```

   Lo script, infine, elabora il formato numerico dei dati fisici pluviometrici. Il dato viene convertito in formato `numeric` per permettere le successive operazioni aritmetiche. In più viene controllata la continuità delle informazioni grazie all'istruzione `.fillna(0.0)` che sostituisce eventuali valori mancanti con valori nulli. La scelta di tale istruzione risiede nel non compromettere le cumulate mensili o annuali da eventuali buchi nel dataset, assicurando una produzione di un risultato, anche in presenza di brevi interruzioni nel flusso di dati.

Nel complesso la pre-elaborazione si occupa di effettuare una pulizia e preparazione del dato che sia il più facile da gestire per le operazioni future, anche agendo su eventuali dati mancanti o corrotti che è possibile gestire facilmente in questa fase, mentre diventerebbe complesso nelle fasi successive.

	A	B	C
1	quadrante	ora	PRECG
8402	APC_ST12(	1/1/2024	1.6
8403	APC_ST12(	1/2/2024	0.2
8404	APC_ST12(	1/3/2024	0.0
8405	APC_ST12(	1/4/2024	0.0
8406	APC_ST12(	1/5/2024	0.4
8407	APC_ST12(	1/6/2024	4.6
8408	APC_ST12(	1/7/2024	10.0
8409	APC_ST12(	1/8/2024	1.6
8410	APC_ST12(	1/9/2024	0.6

Figura 5.6: estratto dei dati pluviometrici giornalieri pre-elaborati per la stazione di Cupramontana.

5.2.3 Aggregazione giornaliera

L'obiettivo di questa fase è l'aggiornamento incrementale del registro quotidiano delle precipitazioni. Lo script, infatti, non sovrascrive il dataset, ma effettua un append mirato e garantisce che ogni nuova misura venga collocata correttamente nella serie storica.

Per rendere il sistema autonomo nell'effettuare un update corretto, lo script deve individuare la data del giorno per il quale inserire il valore corretto. Andando in esecuzione ogni giorno, il sistema si occupa di individuare il giorno precedente al giorno corrente e definisce un oggetto `yesterday` così da poter richiamare metodi utili all'elaborazione (es. `.date()`) e usarlo come chiave di ricerca per ricavare il valore corretto:

```
yesterday = datetime.now() - timedelta(days=1)
```

Una volta identificata la data di interesse, il sistema si occupa di estrapolare, dal dataset pre-elaborato, il valore corrispondente: viene scansionato il file csv nella sezione 'ora' per trovare la data che coincide con `yesterday.date()`; a questo punto si estrapola il valore di precipitazione corrispondente.

```
value_row = df_target[df_target['ora'].dt.date == yesterday.date()]
if value_row.empty:
    value_to_insert = 0
else:
    value_to_insert = value_row['PRECG'].values[0]
```

Avendo a disposizione il valore numerico per la data del giorno precedente, è possibile interagire con il report excel, selezionando il foglio 'giornaliero' e aggiornando il valore corretto. La logica implementata segue il pattern Upsert (Update/Insert): il sistema verifica l'esistenza della data; se la riga esiste, il valore viene aggiornato; se assente, viene creata una nuova entry, garantendo la continuità della serie temporale senza duplicazioni.

```
for row in sheet.iter_rows(min_row=2):
    cell_value = row[0].value
    if isinstance(cell_value, (datetime, date)) and cell_value.date() ==
yesterday.date():
        date_row = row[0].row
        break
```

Inoltre, se la data non è già presente nel report excel, il sistema si occupa di creare un nuova riga con la cella per la data corretta e la cella per il valore estrapolato.

```
if date_row is None:
    new_row = sheet.max_row + 1
    sheet.cell(row=new_row, column=1).value = yesterday.date()
    sheet.cell(row=new_row, column=2).value = value_to_insert
else:
    sheet.cell(row=date_row, column=2).value = value_to_insert
```

In questa fase, lo script si concentra sulla dinamicità dell'aggiornamento dei dati giornalieri e su una buona gestione degli errori o di dati mancanti, garantendo continuità operativa nell'esecuzione del file.

5.2.4 Aggregazione annuale

Dopo aver aggiornato il foglio 'giornaliero', il sistema ha dunque a disposizione tutto lo storico aggiornato direttamente nel file excel, senza dover recuperare altri dati all'esterno e potendo quindi facilmente lavorare sulle aggregazioni e le analisi statistiche.

Lo script, nel foglio 'mesi-anni', si occupa di gestire i dati ed elaborarli in aggregazioni mensili e annuali, per osservarne l'andamento su lungo periodo; il foglio è costituito da:

- una tabella contenente i dati di ogni mese per ciascun anno presente nello storico della stazione
- valori di media mobile calcolati sui precedenti 3 o 5 anni
- percentuali indicative di minimo, massimo, media storica, variazione, deviazione standard e ventesimo percentile per i vari mesi e l'anno corrente

- grafico che rappresenta l'andamento dei valori di somma totale annua, media mobile a 3 anni e media mobile a 5 anni.

Da un punto di vista operativo è necessario effettuare controlli sulla data di esecuzione dello script e verificare se essa coincide con l'inizio di un nuovo anno o di un nuovo mese per poter aggiornare la tabella e il grafico corrispondente e poter consolidare i dati del mese o anno precedente.

```
if(today.month == 1 and today.day == 1):
    row_found = None
```

Risulta, quindi, utile per la costruzione di nuove righe e celle grazie all'uso di `row_found+1` per la corretta gestione del foglio di calcolo.

Questo fattore è fondamentale per la solidità dei report excel; il database diventa, in questo senso, auto-estendibile: lo script si assicura che il foglio di calcolo possa aggiungere valori ed espandersi negli anni futuri in maniera completamente autonoma, anche per quanto riguarda la formattazione degli stili e formati delle tabelle e dei grafici.

Inoltre, se il giorno coincide con quello di un nuovo mese risulta fondamentale per il sistema iterare su tutte le osservazioni giornaliere del mese precedente, eseguire una sommatoria dei valori di interesse sul campo `PRECG` e inserire tale valore nella cella di competenza per l'anno in corso (processo semplificato tramite una mappatura del numero del mese alla colonna corrispondente nel foglio di calcolo)

```
for row in sheet.iter_rows(min_row=2):
    cell_data = row[0].value
    dato_cell = row[1].value
    if isinstance(cell_data, datetime):
        if last_month_start <= cell_data <= last_month:
        total_sum += float(dato_cell) if dato_cell is not None else 0
```

La logica appena descritta di aggiornamento annuale, utilizzata anche nelle cumulate mensili, impone, quindi, la produzione e aggiornamento di un database storico coerente, stabile e strutturato. Le informazioni annuali rappresentano, infatti, dati fondamentali per verificare l'andamento lento nel tempo della sorgente e la sua capacità su lungo periodo di autoricaricarsi dalle piogge.

Un altro aspetto fondamentale delle cumulate annuali calcolate mese per mese è quello di creare una tabella di partenza per le operazioni successive; da questa tabella, infatti, vengono recuperate le informazioni sulle quali si costruiscono i grafici e le funzioni statistiche che sottolineano i trend mensili che si ripetono ogni anno o le anomalie che interessano gli stessi mesi nel corso degli anni.

5.2.5 Cumulate mensili

L'ultimo stadio finale dell'elaborazione dati è costituito dal foglio 'cumulate mensili'. In questo modulo, il sistema calcola le precipitazioni cumulative su diverse finestre temporali (3, 4, 5, 6 e 7 mesi); le cumulate su più mesi forniscono una visione d'insieme dei trend stagionali pluviometrici. A supporto di un'analisi stagionale, inoltre, vengono calcolate le cumulate da ottobre, verificando l'effetto dell'inizio della stagione delle piogge sui mesi successivi. In ultima analisi, vengono presi in considerazione fattori statistici importanti e grafici di visualizzazione.

- Cumulate di più mesi:

  La gestione delle cumulate mobili rappresenta una sfida nel codice per il calcolo temporale, soprattutto nel caso di un nuovo anno, in quanto deve essere garantita una certa precisione nella somma dei contributi mensili precedenti al mese corrente. Ciò viene garantito grazie alla mappatura sulle colonne dei vari mesi dell'anno effettuata nel foglio 'mesi-anni':

  ```
  month_to_column = 1: 'B', 2: 'C', 3: 'D', 4: 'E', 5: 'F', 6:
  'G', 7: 'H', 8: 'I', 9: 'J', 10: 'K', 11: 'L', 12: 'M'
  ```

  Consentendo perciò l'utilizzo di formule di somma dinamiche per il mese corrente. L'operazione viene eseguita il primo giorno del mese ed è mostrata interamente nella figura 5.7: all'interno di un ciclo iterativo sulle colonne dei mesi del foglio 'mesi-anni' si identifica il mese precedente come mese di partenza per l'iterazione nei mesi precedenti: `mese = today.month-1-i` e si ricavano i valori cumulativi di ogni mese. I singoli valori cumulati vengono, così, aggregati tramite una funzione di somma che calcola i valori nell'ordine temporale corretto:

  ```
  formula = f"=SUM('','.join(reversed(formula_parts)))"
  ```

  Una particolare attenzione viene fatta per l'inizio di un nuovo anno; viene infatti posta attenzione a definire la coppia mese e anno corretta, in modo tale da poter sommare Gennaio 2026 e Dicembre 2025 all'interno della stessa formula. All'interno del ciclo `for`, il valore del mese può diventare negativo indicando che precede Gennaio; il codice si pone di riportare il mese nell'intervallo corretto di valori, diminuendo di 1 il valore dell'anno di riferimento:

  ```
  if mese <= 0:
      mese += 12
      anno_corrente -= 1
  ```

```
for k in range(3, 8):  # Cumulate da 3 a 7 mesi
    formula_parts = []
    for i in range(k):
        mese = today.month - 1 - i  # Partire dal mese precedente
        anno_corrente = anno
        if mese <= 0:
            mese += 12
            anno_corrente -= 1
        #Trova la riga corrispondente all'anno nel foglio 'mesi - anni'
        riga_mesi = None
        for riga_m in range(3, sheet_mesi.max_row + 1):
            anno_cell = sheet_mesi.cell(row=riga_m, column=1).value
            if anno_cell == anno_corrente:
                riga_mesi = riga_m
                break
        #Somma 0 se il dato è mancante
        if riga_mesi is None:
            formula_parts.append("0")
        else:
            col_letter = get_column_letter(mese + 1)
            valore = sheet_mesi[f"{col_letter}{riga_mesi}"].value
            if valore is None:
                formula_parts.append("0")
            else:
                formula_parts.append(f"'mesi - anni'!{col_letter}{riga_mesi}")
    #Restituisce la formula con i mesi in ordine corretto
    formula = f"=SUM({','.join(reversed(formula_parts))})"
    cell = sheet_cumulate.cell(row=anno_riga, column=k - 3 + 2)
    cell.value = formula
```

Figura 5.7: Calcolo delle precipitazioni cumulate su finestre temporali di 3-7 mesi.

- Cumulate da ottobre:

  Le somme dei dati da ottobre sono un indicatore importante per definire l'andamento dell'anno idrologico corrente. L'anno idrologico è il periodo di 12 mesi che inizia il 1 ottobre e termina il 30 settembre; fondamentale perché dopo la stagione secca estiva si ha un azzeramento sostanziale delle piogge, perciò a ottobre si iniziano a ricaricare le falde e i bacini idrologici.

  Su questa indicazione, lo script Python si occupa di calcolare le cumulate da ottobre fino a maggio, giugno, luglio, agosto, settembre e ottobre successivo, ovvero la valutazione di come il periodo delle piogge autunnali e invernali favorisce una buona base per l'arrivo dei mesi estivi e siccitosi. Da un punto di vista operativo, la cumulata si divide nella somma di ottobre, novembre e dicembre dell'anno corrente e la somma dei mesi dell'anno successivo:

  ```
  sheet_cumulate.cell(row_found+1, column=8).value = f'=SUM(m̃esi -
  anni!K{row_found}:M{row_found})+SUM(m̃esi -
  anni!B{row_found+1}:i{row_found+1})'
  ```

`i=(F,G,H,I,J,K)`

`[F=maggio, G=giugno, H=luglio, I=agosto, J=settembre, K=ottobre]`

Il risultato di questa elaborazione permette di osservare i trend stagionali nei diversi anni, permettendo analisi di surplus o deficit stagionali; confrontando la cumulata da ottobre corrente con la media storica dello stesso periodo, il sistema è in grado di fornire un alert immediato sullo stato di salute delle riserve idriche.

- Indicatori statistici: Oltre alle tabelle cumulative, il sistema integra una serie di parametri statistici fondamentali che rispecchiano i valori di massimo, minimo, media storica, variazione, deviazione standard e ventesimo percentile per le cumulate mobili dei mesi e le cumulate da ottobre.

  Le formule e i valori più importanti sono l'`ESITO` (confronta la variazione dei valori di pioggia correnti con la media storica e visualizza se la differenza è sopra o sotto soglia, ovvero ha un delta maggiore di 0,05 rispetto alla media), la `STDEV` (deviazione standard; misura la dispersione dei dati intorno alla media; utile per valutare la stabilità della sorgente) e il `PERCENTILE` (ventesimo percentile; rappresenta un livello di soglia al di sotto del quale un evento è considerato critico).

- Visualizzazione grafica: L'ultimo passaggio è la trasposizione delle cumulate su più mesi nel corso degli anni in un *line chart*, in cui le linee rappresentano le cumulate degli ultimi y mesi per tutti gli anni in questione (con y= 3,4,5,6,7).

  L'aggiornamento avviene in maniera autonoma grazie all'uso della classe `Reference` di `openpyxl`. Questo permette un confronto visivo immediato tra periodi di siccità di breve periodo e quelli di lungo periodo; quindi viene analizzata e visualizzata efficacemente la capacità di ricarica delle falde.

```
#Aggiorno il grafico in cumulate mensili
try:
    chart = sheet_cumulate._charts[0]
    # aggiorno le categorie (assi X)
    chart.categories = Reference(sheet_cumulate, min_col=1, min_row=3, max_col=1, max_row=row_found + 1)
    # aggiorno i valori (assi Y)
    chart.series[0].values = Reference(sheet_cumulate, min_col=2, min_row=3, max_col=2, max_row=row_found + 1)
    chart.series[1].values = Reference(sheet_cumulate, min_col=3, min_row=3, max_col=3, max_row=row_found + 1)
    chart.series[2].values = Reference(sheet_cumulate, min_col=4, min_row=3, max_col=4, max_row=row_found + 1)
    chart.series[3].values = Reference(sheet_cumulate, min_col=5, min_row=3, max_col=5, max_row=row_found + 1)
    chart.series[4].values = Reference(sheet_cumulate, min_col=6, min_row=3, max_col=6, max_row=row_found + 1)
    chart.x_axis.scaling.max = None
except Exception as e:
    print(f"Errore nell'aggiornamento del grafico: {e}")
```

Figura 5.8: funzione aggiornamento grafico cumulate mensili.

Con l'aggiornamento del foglio 'cumulate mensili' si concludono le operazioni di elaborazione delle informazioni. Lo script pone attenzione all'automazione nei

processi di aggiornamento, alla dinamicità della gestione dei vari anni idrologici e alla presentazione di un numero efficiente ed esaustivo di tabelle, formule e grafici per una corretta gestione e supporto nei processi decisionali.

5.3 Automazione del sistema e processi di aggiornamento

L'ultima fase dell'implementazione riguarda l'automazione completa dello script; la sua esecuzione viene affidata all'ecosistema di Amazon Web Services: lo script viene, infatti, caricato nel bucket di S3 e gestito da AWS Lambda, che consente un'esecuzione giornaliera del codice, tramite Amazon EventBridge ed espressioni `cron`.

Sulla base di questo ambiente di sviluppo, lo script implementa una logica di aggiornamento incrementale, non riscrivendo l'intero database a ogni iterazione ma focalizzandosi sugli ultimi dati inseriti (quali ultima data presente e ultima riga contenente dati e valori utili). Questo garantisce la ripetibilità dei risultati e la fruizione di output sempre aggiornati. Inoltre, l'utilizzo della libreria `BytesIO` permette una gestione efficiente dei report, diminuendo i tempi di esecuzione, l'utilizzo di banda necessario e i costi del servizio cloud.

L'implementazione dello script si pone alcuni obiettivi principali per favorire l'automazione dell'intero sistema:

- Robustezza (Fault Tolerance): tramite l'incapsulamento delle operazioni in blocchi `try ... except`, il sistema è in grado di individuare errori senza fermare l'esecuzione, riducendo interventi manuali di gestione, che risultano contrastare la logica di una pipeline `cloud-native`.

- Scalabilità orizzontale: l'architettura serverless permette di inserire o eliminare stazioni pluviometriche o di definire intervalli di tempo su cui agire, senza interferire con la logica di elaborazione e senza la necessità di ridimensionamento del server.

- Integrità delle informazioni: tramite la gestione di funzioni di formattazione e calcolo che forniscono una linea guida rigida e solida per l'elaborazione dei dati.

L'intero workflow dedicato alla gestione dei dati pluviometrici diventa un vero e proprio asset tecnologico autonomo a disposizione delle operazioni decisionali di gestione della risorsa idrica proveniente dalla sorgente di Gorgovivo.

Capitolo 6

Risultati

Il sistema sviluppato consente di trasformare i dati grezzi ed eterogenei in informazioni organizzate e coerenti, attraverso diverse fasi di acquisizione, elaborazione e aggiornamento incrementale dei report. Su questa base procede l'analisi dei risultati del task di reportistica. Nelle seguenti sezioni verranno analizzati i risultati ottenuti dal report generato per la stazione di Cupramontana, suddividendo l'analisi dei risultati ottenuti in dati giornalieri, annuali e cumulativi mensili. La stazione di Cupramontana, in analisi, funge solo come esempio in quanto i report presentano la stessa struttura e l'analisi dei risultati segue lo stesso schema per tutte le stazioni considerate.

6.1 Acquisizione e processamento dati giornalieri

Nel foglio giornaliero vengono riportati i dati ottenuti dalla preelaborazione dei file CSV. Vengono aggiornati e inseriti i dati giornalieri in un tabella contenente la data e il valore del dato corrispondente, come mostrato in figura 6.1. Questa tabella nel foglio giornaliero rappresenta il dataset principale da cui vengono poi presi i dati per l'analisi nei fogli successivi del report.

	A	B
1	Data	Dato
26664	1/1/2024	1.6
26665	2/1/2024	0.2
26666	3/1/2024	0.0
26667	4/1/2024	0.0
26668	5/1/2024	0.4
26669	6/1/2024	4.6
26670	7/1/2024	10.0
26671	8/1/2024	1.6
26672	9/1/2024	0.6
26673	10/1/2024	0.4

Figura 6.1: estratto dei dati pluviometrici trasferiti nel foglio 'giornaliero' della stazione di Cupramontana.

6.2 Processamento dati annuali

In figura 6.2 è rappresentata la tabella delle precipitazioni su base mensile e annuale, con una sommatoria totale per ogni anno, permettendo l'osservazione della variabilità interannuale della singola stazione di riferimento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		pioggia mensile - Cupramontana												
2	anno	gennaio	febbraio	marzo	aprile	maggio	giugno	luglio	agosto	settembre	ottobre	novembre	dicembre	tot annuo
62	2010	15.0	75.4	52.0	131.4	56.2	28.0	45.4	31.0	37.6	40.8	43.6	46.0	602
63	2011	78.0	30.0	47.2	9.4	48.2	43.2	49.4	0.0	24.0	47.2	24.2	60.8	462
64	2012	24.4	173.0	17.0	73.2	37.6	2.4	8.6	5.8	261.0	55.6	142.2	53.6	854
65	2013	49.0	72.2	52.8	32.2	81.2	69.8	46.2	47.0	86.6	89.8	223.0	74.0	924
66	2014	53.8	53.0	94.4	93.6	98.0	71.4	105.4	10.6	118.2	47.2	85.0	66.6	897
67	2015	54.6	147.4	163.6	69.4	175.2	55.2	0.0	82.2	36.6	209.6	47.2	1.8	1043
68	2016	64.6	74.2	107.8	70.6	113.4	104.0	82.6	93.8	27.2	103.8	56.2	3.2	901
69	2017	107.8	125.6	57.8	76.8	64.2	62.8	60.8	19.6	116.8	47.6	182.6	76.4	999
70	2018	25.6	172.4	130.2	28.8	103.2	110.2	32.8	32.6	52.8	49.4	44.8	64.2	847
71	2019	72.0	24.2	30.8	112.8	300.2	0.4	84.8	24.4	118.4	80.8	89.6	58.8	997
72	2020	14.0	31.8	104.8	53.2	91.0	76.4	51.4	48.4	68.4	47.0	59.2	106.8	752
73	2021	68.4	29.8	35.8	76.2	32.2	14.2	31.2	62.0	65.8	162.4	211.8	118.2	908
74	2022	35.8	69.4	33.2	54.0	35.4	40.2	35.4	42.0	201.8	0.4	127.8	87.0	762
75	2023	124.0	57.8	91.8	135.2	200.0	154.4	9.4	59.4	38.2	19.4	142.0	0.0	1032
76	2024	34.4	18.4	74.2	39.0	60.6	56.0	7.8	5.0	132.8	168.6	28.6	0.0	625

Figura 6.2: tabella mesi-anni - stazione di Cupramontana.

Il report mostra anche le medie mobili a 3 e 5 anni (figura 6.3) e i parametri statistici per ciascun mese (figura 6.4) descritti precedentemente. Queste indicazioni sono strumenti essenziali per evidenziare le oscillazioni annuali, mensili e potenziali anomalie pluviometriche che potrebbero influenzare la ricarica delle falde in determinati periodi.

L'ultima parte dell'analisi annuale delle precipitazioni è descritta dal grafico in figura 6.5. Nel grafico vengono rappresentati i valori di pioggia totale annua e i valori di media mobile a 3 e 5 anni. Il grafico mostra come l'andamento generale dei valori pluviometrici dal 1951 al 2025 calcolati per la stazione di Cupramontana sono in continua descrescita a conferma del cambiamento climatico vissuto in una regione particolarmente attenzionata come le Marche.

O	P
media mobile a 3 anni	media mobile a 5 anni
566	697
639	695
747	748
892	836
955	924
947	953
981	937
916	957
948	899
866	901
886	853
808	890
901	816
806	781
787	787

Figura 6.3: medie mobili - stazione di Cupramontana.

	gennaio	febbraio	marzo	aprile	maggio	giugno	luglio	agosto	settembre	ottobre	novembre	dicembre	tot annuo
min ass	1.3	4.6	2.8	5.6	0.0	0.4	0.0	0.0	7.1	0.0	15.4	0.0	454.0
max ass	300.0	220.4	248.5	176.8	300.2	188.3	258.8	189.8	274.4	259.8	239.9	263.3	1529.1
media storica	70.0	69.2	78.5	75.4	74.8	64.8	50.6	58.8	83.6	87.6	101.5	84.7	897.0
valore anno in corso	23.8	71.2	163.2	51.4	103.8	37.6	88.4	71.6	67.2	26.0	0.0	0.0	704.2
differenza (mm)	-46.2	2.0	84.7	-24.0	29.0	-27.2	37.8	12.8	-16.4	-61.6	-101.5	-84.7	-192.8
variazione %	-66%	3%	108%	-32%	39%	-42%	75%	22%	-20%	-70%	-100%	-100%	-21%
dev stand	54.59	45.82	53.00	40.36	55.38	45.42	46.54	45.66	56.56	60.34	59.53	60.61	217.78
20%	28.24	31.52	33.50	38.36	30.12	32.88	17.38	19.44	37.36	40.16	45.76	39.28	713.12

Figura 6.4: statistiche mesi-anni - stazione di Cupramontana.

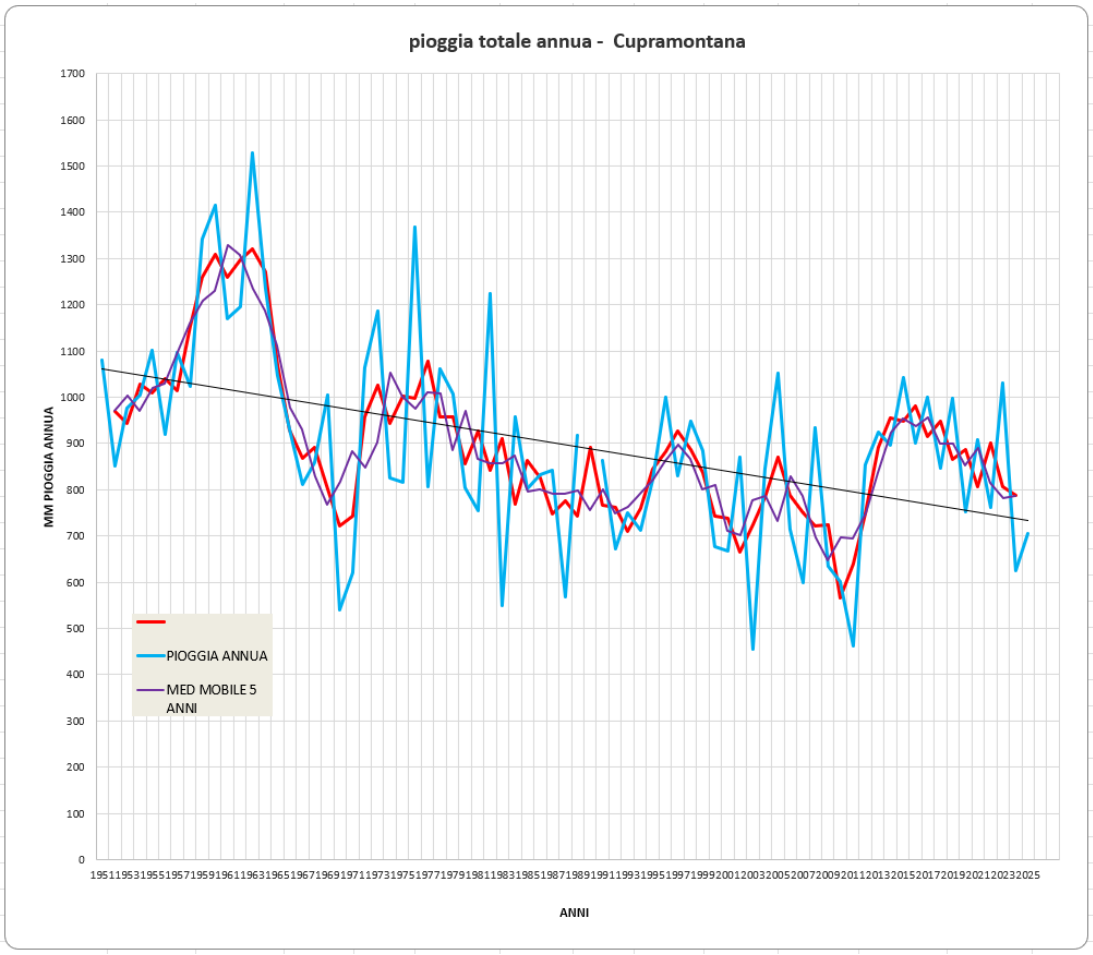

Figura 6.5: grafico mesi-anni - stazione di Cupramontana.

6.3 Processamento dati cumulativi mensili

Nel foglio delle cumulate mensili del report i risultati vengono suddivisi nelle cumulate di più mesi, in base al mese corrente di visualizzazione, (figura 6.6) e nelle cumulate da ottobre (figura 6.7). I risultati mostrati in questa tabella sono fondamentali per un'analisi temporale di breve e lungo periodo in base al giorno e mese di riferimento. Eventi climatici anomali possono avere un'influenza temporali su più mesi durante l'anno ed è fondamentale osservarne l'influenza su pochi o più mesi.

Inoltre l'analisi delle cumulate da ottobre favorisce lo studio degli effetti della stagione invernale delle piogge sui livelli pluviometrici. Livelli bassi in questa sezione della tabella possono rappresentare un grande rischio per la stazione in riferimento in quanto le cumulate da ottobre sono il più grande contributo pluviometrico durante l'anno.

	A	B	C	D	E	F	G
1	CUMULATE DI PIU' MESI						
2	anno	tot pioggia ultimi 3 mesi	totale ultimi 4 mesi	totale ultimi 5 mesi	totale ultimi 6 mesi	totale ultimi 7 mesi	DA INIZIO ANNO
59	2007	205	231	272	311	337	495
60	2008	245	279	326	474	582	666
61	2009	185	210	271	353	404	554
62	2010	109	155	183	239	370	513
63	2011	71	121	164	212	221	377
64	2012	322	331	333	371	444	659
65	2013	223	270	339	421	453	627
66	2014	176	281	353	451	544	746
67	2015	328	328	384	559	628	994
68	2016	225	307	411	525	595	842
69	2017	184	245	308	372	449	740
70	2018	135	168	278	381	410	738
71	2019	224	308	309	609	722	849
72	2020	164	215	292	383	436	586
73	2021	290	321	336	368	444	578
74	2022	244	280	320	355	409	548
75	2023	117	126	281	481	616	890
76	2024	306	314	370	431	470	597

Figura 6.6: Cumulate di più mesi - stazione di Cupramontana.

H	I	J	K	L	M
CUMULATE DA OTTOBRE					
OTT - MAG	OTT - GIU	OTT - LUGL	OTT - AGO	OTT - SETT	OTT - OTT
304	345	371	422	477.4	575.6
542	589	622	672	731.2	867.4
688	749	774	803	848.8	959.2
521	549	594	625	662.6	703.4
343	386	436	436	459.8	507.0
457	460	468	474	735.2	790.8
539	609	655	702	788.4	878.2
780	851	956	967	1085.2	1132.4
809	864	864	946	983.0	1192.6
689	793	876	970	996.8	1100.6
595	658	719	739	855.4	903.0
767	877	910	942	995.2	1044.6
698	699	784	808	926.4	1007.2
524	600	652	700	768.6	815.6
455	470	501	563	628.6	791.0
720	760	796	838	1039.6	1040.0
824	978	988	1047	1085.4	1104.8
388	444	452	457	589.6	758.2

Figura 6.7: Cumulate da ottobre - stazione di Cupramontana.

La seconda parte del report riguardo le cumulate mensili fa riferimento ai parametri statistici calcolati sulla base dei dati descritti precedentemente, sia per quanto riguarda le cumulate di più mesi consecutivi (figura 6.8) sia per le cumulate mensili da ottobre (figura 6.9). Oltre a informazioni tecniche di minimo, massimo e percentuali di variazioni indicative, viene visualizzato un parametro "ESITO" che descrive immediatamente la situazione del periodo di riferimento per facilitare l'analisi e la lettura dei parametri calcolati.

	CUMULATE DI PIU' MESI					
	tot pioggia ultimi 3 mesi	totale ultimi 4 mesi	totale ultimi 5 mesi	totale ultimi 6 mesi	totale ultimi 7 mesi	DA INIZIO ANNO
min ass	0.0	0.0	0.0	0.0	0.0	0.0
max ass	468.1	578.0	741.5	764.0	826.7	1310.5
media storica	226.9	276.8	340.8	414.6	489.1	703.8
valore anno in corso	164.8	253.2	290.8	394.6	446.0	704.2
ESITO	sotto media	sotto media	sotto media	nella media	sotto media	nella media
differenza (mm)	-62	-24	-50	-20	-43	0
variazione %	-27.4%	-8.5%	-14.7%	-4.8%	-8.8%	0.1%
dev stand	94.16	111.13	125.60	138.64	150.85	203.60
20%	139.40	203.56	266.76	310.96	376.00	552.44

Figura 6.8: Parametri statistici cumulate di più mesi - stazione di Cupramontana.

	CUMULATE DA OTTOBRE					
	OTT - MAG	OTT - GIU	OTT - LUG	OTT - AGO	OTT - SETT	OTT - OTT
min ass	304.0	344.6	371.0	421.6	459.8	507.0
max ass	1208.4	1291.1	1340.1	1421.7	1549.1	1804.4
media storica	641.4	707.2	758.2	818.1	901.9	987.9
valore anno in corso	610.6	648.2	736.6	808.2	875.4	901.4
ESITO	nella media	sotto media	nella media	nella media	nella media	sotto media
differenza (mm)	-31	-59	-22	-10	-27	-87
variazione %	-4.8%	-8.3%	-2.8%	-1.2%	-2.9%	-8.8%
dev stand	179.50	193.44	199.74	210.28	209.56	232.08
20%	474.12	536.92	584.72	637.76	732.00	791.64

Figura 6.9: Parametri statistici cumulate da ottobre - stazione di Cupramontana.

Il grafico finale del report, in figura 6.10, rappresenta l'andamento nei vari anni dei valori di cumulate negli ultimi 3, 4, 5, 6 e 7 mesi dal mese corrente di riferimento. Graficamente si nota la variazione nei vari anni dello stesso periodo di riferimento per quanto riguarda i dati pluviometrici, consentendo una valutazione complessiva dello stesso riferimento temporale preso in considerazione.

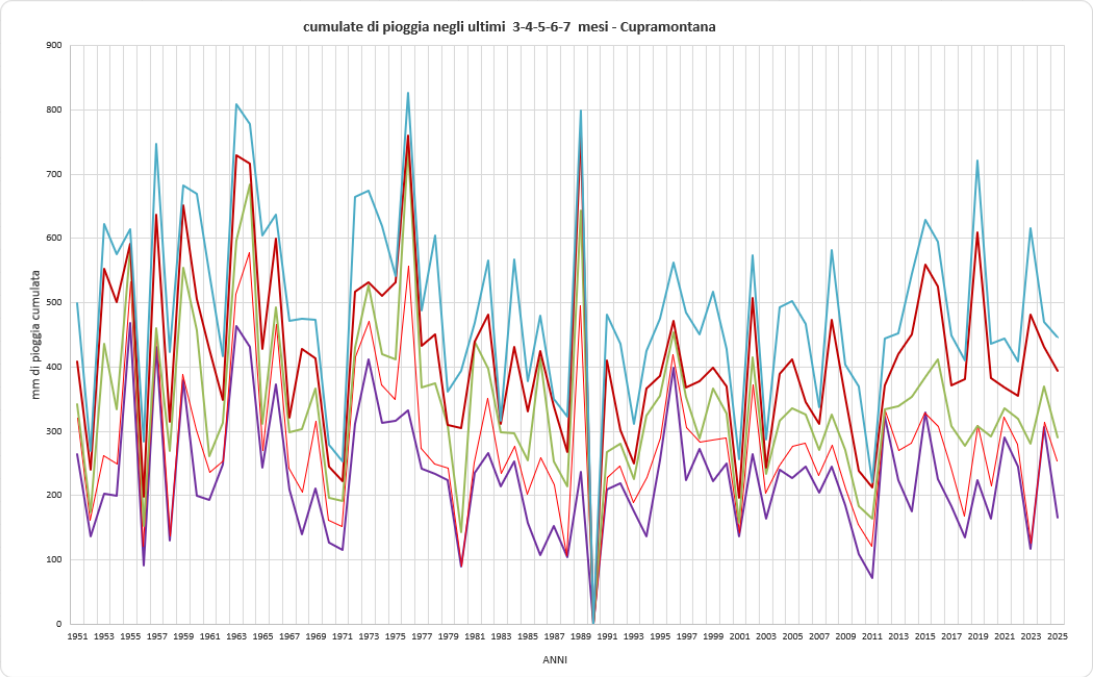

Figura 6.10: Grafico cumulate mensili - stazione di Cupramontana.

Capitolo 7

Conclusioni

Il presente lavoro di tesi ha riguardato l'integrazione di un sistema di reportistica automatizzata nell'architettura cloud del progetto Gorgovivo 4.0, con l'obiettivo di supportare la gestione, l'organizzazione e la restituzione strutturata dei dati raccolti. La tesi si è concentrata sul lavoro di progettazione di script in linguaggio Python, capaci di dialogare con l'infrastruttura AWS preesistente per garantire continuità, precisione e scalabilità.

Il lavoro svolto ha permesso di confrontarmi con le sfide reali di un progetto infrastrutturale complesso, consolidando le mie competenze nell'ingegneria dei dati e nell'adozione di architetture cloud tramite i servizi di AWS. Inoltre, il contributo descritto nella presente tesi ha permesso di raggiungere gli obiettivi prefissati, fornendo un contributo significativo all'automazione dei processi di elaborazione dei dati.

Attraverso l'integrazione dello script Python, ho integrato l'infrastruttura di acquisizione dati esistente con un sistema di reportistica automatizzata. Lo script descritto nella tesi consente di facilitare un'analisi accurata delle informazioni. I report generati diventano uno strumento di supporto alle decisioni e alle analisi future del bacino idrico della sorgente di Gorgovivo. I requisiti raggiunti tramite l'elaborazione del task di reportistica sono descritti in quattro principali aspetti funzionali:

- Automazione del processo:

  il sistema riduce significamente l'intervento manuale richiesto per l'analisi dei dati, garantendo ripetibilità, standardizzazione e continuità operativa. Il lavoro svolto consente di eseguire l'intera pipeline di elaborazione dei dati in maniera autonoma, a partire dall'acquisizione dei dati fino all'elaborazione e aggiornamento dei report strutturati. L'aggiornamento incrementale tramite Python consente di gestire efficacemente e coerentemente le serie temporali di grandi dimensioni, riducendo il rischio di errori operativi e garantendo integrità nello storico informativo. L'adozione di un'architettura serverless basata sui servizi cloud di AWS permette l'esecuzione pe-

riodica dello script e assicura continuità operativa e scalabilità del sistema, migliorandone l'affidabilità e l'efficienza.

- Qualità e integrità dei dati:

  la pipeline di elaborazione consente una corretta gestione di dati eterogenei, raffinando i formati di dato utilizzati e garantendo uniformità tra le informazioni. Le operazioni di pre-elaborazione e validazione, tramite l'uso di librerie Python strutturate, definiscono le azioni per individuare e trattare dati mancanti, anomalie di misura e discontinuità nelle serie temporali. Questo processo contribuisce a incrementare l'affidabilità complessiva delle informazioni. Lo script assicura la produzione di dati consistenti, confrontabili e integri per consolidare le informazioni storiche e facilitare i processi di lettura e analisi dei report.

- Ottimizzazione delle risorse:

  l'intero processo gestisce correttamente ed efficacemente grandi serie temporali storiche tramite un'adeguata architettura cloud, centrata sul paradigma dei data processing pipeline serverless systems. L'adozione di AWS rappresenta una scelta ottimizzata sotto questo aspetto. Dal punto di vista computazionale, infatti, l'esecuzione dello script e l'aggiornamento dei report avviene esclusivamente quando necessario, riducendo i costi di elaborazione e di mantenimento di infrastrutture di calcolo dedicate e i tempi di elaborazione tramite un elaborazione incrementale. Per quanto riguarda l'archiviazione, l'unità cloud di storage S3 consente la gestione di grandi quantitativi di dati indipendente dall'ambiente di sviluppo e senza inficiare sulle prestazioni e velocità di calcolo per l'acquisizione e l'aggiornamento dei report prodotti. La standardizzazione dei report e l'automazione generale del processo riducono il carico di lavoro manuale e il tempo richiesto agli operatori, favorendo e orientando l'attenzione sulle attività di analisi a livello decisionale e gestionale.

- Valorizzazione delle informazioni:

  in ultimo, il sistema rappresenta e implementa perfettamente la logica dei Decision Support Systems. Risulta significativa la strutturazione dei report per garantire l'efficacia di questo strumento di supporto e mira a rendere il patrimonio informativo più accessibile, interpretabile e funzionale. Le diverse aggregazioni temporali, l'organizzazione delle informazioni in grafici e tabelle più leggibili e facilmente confrontabili rende i report strumenti efficaci in fase decisionale operativa. La produzione di una tale fase reportistica consente una facilità di integrazione delle informazioni con altri sistemi di analisi e visualizzazione e incrementa, così, il suo valore applicativo.

7.1 Limitazioni

Il sistema sviluppato ha permesso di conseguire gli obiettivi definiti e discussi nel capitolo precedente. Nonostante i risultati ottenuti, la soluzione proposta presenta alcune limitazioni, riconducibili alle scelte progettuali e implementative affrontate nel presente lavoro di tesi.

- Dipendenza dai dati di input: per quanto riguarda i dati in ingresso, il sistema opera su una base incrementale giornaliera di dati forniti tramite file CSV, disponibili presso un bucket S3. Su questo funzionamento possiamo introdurre una possibile limitazione data da eventuali interruzioni o malfunzionamenti di disponibilità dei dati. Interruzioni o malfunzionamenti sporadici non intaccano l'integrità dei report e una loro corretta analisi, ma interruzioni prolungate potrebbero portare un'assenza significativa nei dati e nelle tabelle mostrate nei report tanto da influenzare le medie annuali o mensili prese in analisi e compromettere lo storico delle informazioni.

- Complessità degli automatismi: il sistema prevede un'accurata logica di preelaborazione per gestire correttamente i dati in analisi. Tuttavia una modifica strutturale dei dati presenti nei file CSV e del sistema di raccoglimento dati dalla rete sensoristica fino ai file CSV potrebbe presentare problemi di correttezza nell'elaborazione delle informazioni andando a compromettere tutta la logica successiva dell'aggiornamento dei report considerando eventuali modifiche nel formato come dati nulli. Una particolare limitazione, che facilmente può essere superata, riguarda anche la frequenza di campionamento. Un frequenza diversa da quella giornaliera richiederebbe una ristrutturazione di alcune parti fondamentali dello script in analisi.

- Scalabilità del modello: attualmente il sistema si occupa della gestione di dati pluviometrici ricavati nelle varie stazioni in analisi. Il sistema può facilmente incrementare o diminuire il numero di stazioni prese in considerazione e gestire la scalabilità sotto questo aspetto. Se il sistema prevede un'estensione del tipo di dato in analisi (considerando dati relativi non solo alla quantità di acqua, ma anche alle qualità di essa), sarebbe necessaria una ristrutturazione generale dello script in analisi per gestire una diversa natura del dato.

7.2 Sviluppi futuri

All'interno dell'ecosistema Gorgovivo 4.0, il sistema di reportistica pluviometrica realizzato in questo lavoro di tesi costituisce una base operativa scalabile per l'evoluzione delle funzionalità di supporto decisionale.

I report generati attraverso lo script sviluppato verranno, infatti, integrati all'interno di dashboard interattive, realizzate tramite Microsoft Power BI, consentendo una visualizzazione dinamica e intuitiva degli indicatori pluviometrici e facilitando l'accesso alle informazioni da parte di tutti gli stakeholders. Un ulteriore sviluppo riguarda l'estensione o la comunicazione della pipeline di elaborazione con altre tipologie di dati idrologici (livelli di falde, portate della sorgente) al fine di ottenere una visione integrata e multi-variabile dell'intero sistema idrico. La disponibilità di informazioni più complete e tempestive può rappresentare un elemento rilevante per la gestione del sistema, soprattutto in presenza di condizioni critiche. In tal senso, uno sviluppo futuro significativo potrebbe consistere nell'implementazione di un sistema di alerting, basato su soglie critiche o sull'individuazione automatica di anomalie nei dati registrati, così da supportare interventi più rapidi ed efficaci.

L'architettura adottata definisce e sottolinea la necessità dell'ingegneria informatica come elemento fondamentale e abilitante per una gestione sostenibile delle risorse idriche e per lo sviluppo di strumenti informativi evoluti a supporto dei processi e delle politiche decisionali.

Questa tesi conferma che l'ingegneria informatica e automatica moderna assume grande valore quando è messa al servizio della società e dell'ambiente. Essa può diventare garante di una convivenza più equa e sostenibile con l'ecosistema, oltre ad essere un ottimo alleato nella protezione delle risorse vitali del nostro territorio.

Bibliografia

[1] World Bank. *High and Dry: Climate Change, Water, and the Economy.* World Bank, 2016.

[2] WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP). Progress on household drinking water, sanitation and hygiene 2000-2022: Special focus on gender. Technical report, United Nations Children's Fund (UNICEF) and World Health Organization (WHO), 2023.

[3] United Nations. Transforming our world: the 2030 agenda for sustainable development. Technical Report A/RES/70/1, United Nations General Assembly, 2015.

[4] World Meteorological Organization. 2021 state of climate services: Water. Technical Report 1278, WMO, Geneva, Switzerland, 2021.

[5] M. O. Cuthbert, R. G. Taylor, T. Gleeson, et al. Global patterns and dynamics of climate–groundwater interactions. *Nature Climate Change*, 9, 2019.

[6] FAO. Water for sustainable food and agriculture. Technical report, Food and Agriculture Organization of the United Nations, 2017.

[7] United Nations. World water development report 2024. Technical report, UNESCO, 2024.

[8] Legambiente. Osservatorio città clima. `https://cittaclima.it/`, 2023.

[9] skytg24. ciclone harry: Sicilia, calabria e sardegna. `https://tg24.sky.it/cronaca/ciclone-harry`, 2026.

[10] ENEA. Rapporto sulla gestione delle risorse idriche. Technical report, Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile, 2023.

[11] ISTAT. Le statistiche sull'acqua | anni 2020-2024. Technical report, Istituto Nazionale di Statistica, Marzo 2025. Pubblicato in occasione della Giornata mondiale dell'acqua.

[12] ENEA. Gestione delle risorse idriche: criticità e scenari. *Energia, Ambiente e Innovazione*, 2023.

[13] Parlamento Europeo. Direttiva 2000/60/ce del parlamento europeo e del consiglio, del 23 ottobre 2000, che istituisce un quadro per l'azione comunitaria in materia di acque. Technical report, Unione Europea, 2000.

[14] Parlamento Europeo. Direttiva 2006/118/ce sulla protezione delle acque sotterranee. Technical report, Unione Europea, 2006.

[15] Parlamento Europeo. Direttiva 2007/2/ce del parlamento europeo e del consiglio, del 14 marzo 2007, che istituisce un'infrastruttura per l'informazione territoriale nella comunità europea (inspire). Technical report, Unione Europea, 2007.

[16] Repubblica Italiana. *Decreto Legislativo 3 aprile 2006, n. 152 - Norme in materia ambientale (Testo Unico Ambientale)*, 2006.

[17] Repubblica Italiana. *Decreto Legislativo 23 febbraio 2023, n. 18 - Attuazione della direttiva (UE) 2020/2184 concernente la qualità delle acque destinate al consumo umano*, 2023.

[18] Viva Servizi. Viva Servizi S.p.A. `https://www.vivaservizi.it/index.php?`

[19] Consorzio Gorgovivo. Consorzio Gorgovivo. `https://www.gorgovivo.it/`.

[20] A. Galdelli, A. Mancini, E. Frontoni, and A. N. Tassetti. A Feature Encoding Approach and a Cloud Computing Architecture to Map Fishing Activities. volume Volume 7: 17th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA) of *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, page V007T07A003, 08 2021.

[21] A. Galdelli, A. Mancini, A. N. Tassetti, C. Ferrà Vega, E. Armelloni, G. Scarcella, G. Fabi, and P. Zingaretti. A Cloud Computing Architecture to Map Trawling Activities Using Positioning Data. volume Volume 9: 15th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications of *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, page V009T12A035, 08 2019.

[22] Jacobs. Digital onewater - integrated water management solutions. https://www.jacobs.com/.

[23] Istituto Superiore di Sanità. Antea - anagrafe territoriale dinamica delle acque potabili. https://www.iss.it/, 2023.

[24] Protezione Civile Regione Marche. Delibere maltempo settembre 2022. https://www.regione.marche.it/Regione-Utile/Protezione-Civile/Maltempo-Settembre-2022, 2022.

[25] Simona Epasto and Alessandro Galdelli. L'intelligenza artificiale per la gestione sostenibile delle risorse idriche. Caso di studio: Gorgovivo, Ancona. *Documenti geografici*, 1(1):271–302, 2024. https://doi.org/10.19246/DOCUGEO2281-7549/202401_14.

[26] Davide Fronzi, Gagan Narang, Alessandro Galdelli, Alessandro Pepi, Adriano Mancini, and Alberto Tazioli. Towards Groundwater-Level Prediction Using Prophet Forecasting Method by Exploiting a High-Resolution Hydrogeological Monitoring System. *Water*, 16(1), 2024.

[27] Alessandro Galdelli, Gagan Narang, Lucia Migliorelli, Antonio Domenico Izzo, Adriano Mancini, and Primo Zingaretti. An AI-Driven Prototype for Groundwater Level Prediction: Exploring the Gorgovivo Spring Case Study. In *Image Analysis and Processing – ICIAP 2023*, pages 418–429, Cham, 2023. Springer Nature Switzerland.

[28] Univpm. Università politecnica delle marche. https://www.univpm.it/Entra/.

[29] Wes McKinney et al. pandas: a foundational python library for data analysis and statistics. *Python for high performance and scientific computing*, 14(9):1–9, 2011.

[30] Eric Clark et al. openpyxl - a python library to read/write excel 2010 xlsx/xlsm files, 2026.

[31] Python Software Foundation. *The Python Standard Library : io — Core tools for working with streams*, 2026.

[32] AWS development team. *AWS SDK for Python (Boto3) Documentation.* Amazon Web Services, 2026.

[33] Amazon Web Services. Aws lambda - serverless compute service. https://aws.amazon.com/lambda/, 2024.

[34] ISTAT. Statistiche sulle risorse idriche. Technical report, Istituto Nazionale di Statistica, 2023. Dati relativi alle perdite della rete idrica nazionale.

[35] Green School Varese. Sdg 6: L'accesso all'acqua nel mondo. `https://www.green-school.it/uploads/files/419.pdf`, Gennaio 2024.

[36] Claudia Brunori. L'approccio enea alla gestione sostenibile e circolare dell'acqua. *Energia, Ambiente e Innovazione*, 2023.

[37] Luigi Petta, Gianpaolo Sabia, Davide Mattioli, Girolamo Di Francia, and Saverio De Vito. Tecnologie e sistemi intelligenti per la gestione sostenibile della risorsa idrica. *EAI - Energia, Ambiente e Innovazione*, 2024.

[38] Sergio Cappucci, Luca Maria Falconi, Francesco Pasanisi, Carlo Tebano, and Marco Proposito. Valutazione della risorsa idrica su base territoriale. *EAI - Energia, Ambiente e Innovazione*, 2023.

[39] Rosa Padrevita. Il monitoraggio dei corpi idrici in italia dopo il recepimento della normativa 2000/60/ce. Rapporto tecnico / tesi di stage, ISPRA - Istituto Superiore per la Protezione e la Ricerca Ambientale, 2011.

[40] Primo Zingaretti. Gorgovivo 4.0: Acqua, dati e intelligenza artificiale, Dicembre 2025. Presentazione del volume "Dove nasce l'acqua", Auditorium Viva Servizi, Ancona.