



**Politecnico
di Torino**

Master of Science Program in Environmental and
Land Engineering – Climate Change

A Data-Driven Analysis of Hydrological Trends and Correlations in Aosta's Springs

Candidate:

Sara Salari
matr. 308761

Supervisor:

Prof. Paolo Dabove

Co-Supervisor:

PhD Martina Gizzi

A.Y. 2024/2025

Graduation Session October 2025

Contents

Abstract	1
1 Introduction	2
1.1 Introduction and Research Objectives	2
1.2 State of the Art	3
2 Data and Methodology	5
2.1 Case Study	5
2.1.1 Cheserod spring	6
2.1.2 Entrebin Spring	6
2.1.3 Gabiet Spring	6
2.1.4 Promiod Spring	7
2.1.5 Promise Spring	7
2.1.6 Meteorological Stations	7
2.2 Data Description	8
2.2.1 Data file format	8
2.2.2 Instrumentation and Measurement Accuracy	9
2.3 Data Filtering and Cleaning	9
2.4 Trend Analysis	10
2.5 Correlation Analysis	11
3 Results	13
3.1 Data Quality and Preprocessing	13
3.2 Hydrological Trend Analysis	14
3.2.1 Residual Analysis	16
3.3 Cross-Correlation Results	17
3.3.1 Cross-Correlation Between Springs	18
3.3.2 Time Lag Analysis Between Springs	20
4 Discussion and Conclusion	21
4.1 Interpretation of Results	21
4.2 Impact of Climate Change	27
4.3 Limitations of the Study	28
4.4 Recommendations for Future Research	29
Bibliography	30
Appendix A	32

Abstract

This thesis investigates the long-term thermal and hydrological dynamics of five alpine springs-Cheserod, Entrebin, Gabiet, Promiod, and Promise-located in the Aosta Valley (Northwestern Italy) over a 14-year period (2010–2024). Using high resolution hourly data, the study analyzes water temperature, water level, flow rate, and conductivity to identify trends, seasonal behaviors, and inter-spring relationships in response to climatic drivers.

The time series were first cleaned using Kalman filtering and interpolation techniques, followed by sinusoidal modeling with linear trends. Results reveal that most springs exhibit a subtle but consistent warming trend in water temperature, with varying rates of increase. In contrast, water level trends are more heterogeneous, with some springs showing slight declines. In particular, Gabiet and Cheserod, both used for municipal supply, exhibit distinct patterns of variability.

Cross-correlation analyzes were performed between all spring pairs to evaluate temporal dependencies and regional hydrological coherence. Strong atmospheric correlations (e.g., air temperature $r > 0.95$) were found across all sites, while hydrological linkage varied significantly. In particular, only some springs showed synchronous flow or conductivity responses, indicating that even under shared climatic conditions, their aquifer systems respond differently.

These findings highlight the influence of climate change on alpine groundwater systems and underscore the importance of site-specific analysis for sustainable water resource management in mountainous regions.

1. Introduction

1.1 Introduction and Research Objectives

Mountain aquifers are among the most important and valuable water resources in northern Italy, crucial for supplying local populations. In recent decades, studies across the Italian Alps and Apennines have documented several hydrological challenges, including the gradual drying of many springs, reduced discharge during dry months, and the transition of formerly perennial springs into seasonal ones (Cambi and Dragoni 2000; Fiorillo et al. 2007; Gattinoni and Francani 2010; Forestieri et al. 2018; Padulano et al. 2019). These trends have been attributed to both the overexploitation of groundwater resources and the impacts of climate change.

The Aosta Valley, located in the northwest Italian Alps, is characterized by a complex hydrogeological environment where mountain springs are vital to regional water supply. Springs such as Cheserod, Entrebin, Gabiet, Promiod, and Promise serve as key sources and are closely monitored for their hydrological behavior. Integrating hydrological data from these springs with meteorological observations is essential to understanding the cause-and-effect dynamics within the hydro-meteorological system, providing valuable insights into how climatic factors influence spring responses over time.

Analyzing hydrological data from mountain springs presents several challenges that complicate reliable interpretation. A primary issue is the frequent presence of gaps and missing values within individual spring datasets, often caused by sensor malfunctions, harsh weather conditions, or logistical difficulties in remote monitoring locations. These interruptions reduce data continuity and complicate time series analyses. Additionally, the spatial separation between spring monitoring stations and meteorological measurement sites introduces uncertainty, as local climatic conditions may differ significantly over short distances and altitudinal gradients, limiting the accuracy of correlating meteorological inputs with spring responses. Furthermore, the complex geological heterogeneity of alpine aquifers and the highly dynamic, localized responses of springs to environmental factors demand advanced modeling approaches capable of capturing non-linear and time-lagged interactions.

Finally, measurement noise and outliers inherent in field data require careful preprocessing techniques such as Kalman filtering and interpolation to ensure data quality and robustness of subsequent analyses. Addressing these challenges is critical to develop a reliable understanding of hydrological behavior and the influence of climatic variables on mountain springs.

The primary aim of this study is to analyze the hydrological variables of mountain springs in the Aosta Valley by applying advanced data processing and mathematical

modeling techniques.

The study begins with detailed data cleaning, using methods such as Kalman filtering to reduce noise and interpolation to fill missing values, ensuring high-quality and continuous time series for each spring. Following this, Fourier series combined with linear trend models are fitted to temperature data to capture seasonal cycles and long-term trends. The reliability of these models is evaluated through statistical metrics like the coefficient of determination (R^2). This modeling approach is then extended to other hydrological parameters including electrical conductivity and flow rate, with the goal of identifying consistent mathematical relationships that describe the behavior of multiple springs.

Finally, the study investigates correlations and time-lagged interactions between spring variables and environmental factors such as precipitation and air temperature, to better understand the dynamic relationships within the hydro-meteorological system governing alpine springs.

1.2 State of the Art

Mountain springs are vital components of alpine hydrological systems, serving as important freshwater sources and indicators of broader environmental changes. The hydrodynamics of springs, particularly those in complex geological settings such as the Aosta Valley, have been extensively studied through hydrograph and time series analyses, which provide insights into aquifer properties and groundwater flow dynamics. Historically, hydrograph recession curve analysis has been a fundamental method to characterize aquifer drainage and storage behavior, distinguishing fast conduit flows from slower matrix flows in karst and porous aquifers (Lo Russo et al., 2014; Mangin, 1975; Atkinson, 1977). These approaches help understand the response of springs to hydrological inputs such as rainfall and snowmelt, which are critical in mountainous environments subject to seasonal variability.

Time series methods, including autocorrelation and cross-correlation analyses, have been increasingly employed to evaluate the interrelationships among hydrological variables such as discharge, temperature, and electrical conductivity, as well as their responses to external meteorological drivers.

These statistical tools facilitate the detection of time-lags and coupling effects between climatic inputs and spring behavior, which are essential for managing and predicting spring flow regimes (Lo Russo et al., 2014; Padilla et al., 1994). However, most existing studies have focused on individual springs or aquifer systems, often limited by the availability and quality of continuous datasets.

In recent years, advances in data cleaning techniques, such as the application of Kalman filtering for noise reduction and interpolation for gap filling, have improved the robustness of time series datasets. These preprocessing steps are crucial for accurate modeling and trend detection in spring temperature and flow data (Gizzi et al., 2023). Moreover, the fitting of mathematical models, including Fourier series with linear trends, has enabled quantification of seasonal cycles and long-term changes, providing a structured way to represent complex hydrological signals (Fiorillo, 2009).

Despite these methodological developments, challenges persist in fully understanding the complex hydro-meteorological interactions in mountain spring systems. The spatial variability among springs, often located in geologically heterogeneous and remote terrains, complicates the generalization of findings. Data gaps and inconsistent monitoring intervals further hinder continuous analysis.

Additionally, the impacts of climate change on snowmelt timing, precipitation patterns, and temperature regimes introduce non-stationarities that challenge traditional modeling approaches (Barbieri et al., 2021; Duratorre et al., 2020).

Within the Aosta Valley, the combination of geological complexity and evolving climatic influences necessitates comprehensive, data-driven approaches to discern the correlations and causal links between spring variables and environmental factors. Automated and semi-automated tools for spring monitoring data analysis have emerged to address the need for scalable and repeatable hydrogeological characterizations (Gizzi et al., 2023).

Nonetheless, there remains a gap in integrating multiple parameters across different springs to develop universal or transferable models that capture both local specificity and regional trends.

In summary, the current state of knowledge reflects significant progress in spring hydrograph analysis, time series modeling, and data processing techniques, but also highlights the need for integrated studies addressing multi-variable interactions and the impacts of climate variability in mountain aquifers. This research contributes to this gap by applying advanced data cleaning, mathematical modeling, and correlation analyses to multiple springs in the Aosta Valley, aiming to identify underlying patterns and interactions among hydrological variables in a changing environment.

2. Data and Methodology

2.1 Case Study

This research centers on five mountain springs in the Aosta Valley—Cheserod, Entrebin, Gabiet, Promiod, and Promise—selected based on their hydrological relevance and the availability of comprehensive monitoring data.

These springs are distributed across different parts of the valley, each situated within distinct hydrogeological contexts characterized by varying lithologies, geological structures, and aquifer types. This spatial diversity provides a valuable cross-section of the regional groundwater system, allowing for a robust examination of spring behavior under different environmental and geological influences.

The hydrological variables measured at each spring include water level, water temperature, electrical conductivity, and flow rate. These four parameters collectively characterize the physical state and chemical properties of spring water, reflecting the response of the underlying aquifer to climatic and hydrological inputs.

To contextualize the springs' hydrological behavior within the broader climate system, meteorological data were also collected from nearby weather stations. Key climatic variables include precipitation, air temperature, and humidity, which influence recharge processes, evapotranspiration rates, and snowmelt dynamics in this alpine environment. The close spatial proximity of meteorological stations to the springs enhances the reliability of correlating hydro-meteorological interactions.

The monitoring period for this dataset spans from 2011 to 2024, encompassing multiple years of continuous observations. This temporal coverage enables the analysis of both short-term fluctuations and long-term trends, capturing seasonal cycles as well as potential impacts from climate variability and anthropogenic pressures.

The data underwent rigorous quality control procedures including outlier detection, gap filling using interpolation and Kalman filtering, and validation against physical and statistical criteria. These preprocessing steps ensure the integrity and usability of the data for advanced time series analysis and modeling.

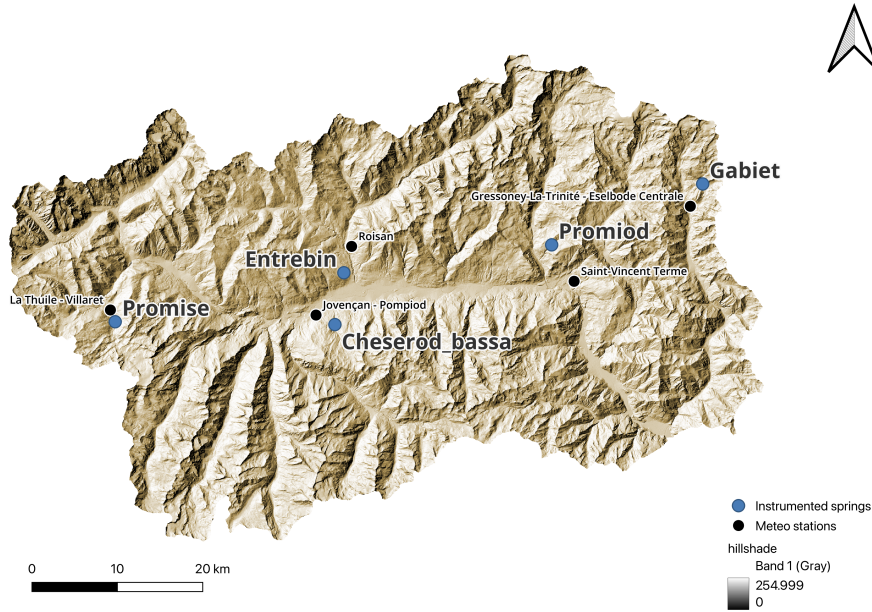


Figure 2.1: figure shows the geographic locations of the five springs and meteo stations the Aosta Valley.

2.1.1 Cheserod spring

Cheserod spring is located in the Gressan municipality at approximately 1095 meters above sea level. Its geology is characterized by soluble Triassic carbonate rocks, including carbonate breccias, gypsum. The spring area is overlain by glacial deposits extensively reworked by eluvial and colluvial processes, which significantly influence groundwater recharge and flow patterns. The spring is monitored continuously with multi-parameter probes measuring water level, temperature, electrical conductivity, and flow rate, providing valuable hydrogeological insights (Mondani et al., 2022).

2.1.2 Entrebin Spring

Entrebin spring situated near Arpuilles-Entrebin at an elevation of about 981 meters above sea level, the Entrebin spring emerges from an aquifer mainly composed of glacial detrital deposits resting on impermeable metamorphic bedrock. The spring's hydrological parameters are continuously recorded alongside meteorological data from the nearby Roisan-Preyl weather station, helping characterize recharge processes influenced by snowmelt and precipitation (Bolognini et al., 2010; Mondani et al., 2022).

2.1.3 Gabiet Spring

Gabiet spring is located in the Gressoney-la-Trinité area. It is positioned in a high alpine basin at an elevation generally near 2300 to 2400 meters, where fractured metamorphic rocks and porous glacial sediments form a complex aquifer system. Monitoring includes water level, temperature, conductivity, and flow rate measurements, supported by meteorological data on precipitation, temperature, and snowpack (Mondani

et al., 2022; Cerino Abdin et al., 2021).

2.1.4 Promiod Spring

The Promiod spring is situated near Chatilln municipality, with an elevation near 1600 to 1700 meters a.s.l. The hydrogeological context consists of fractured bedrock overlain by glacial deposits, influencing groundwater dynamics. It is paired with the Saint Vincent meteorological station, which provides essential climate data such as precipitation and temperature. Continuous hydrological monitoring of spring parameters in conjunction with meteorological inputs enables evaluation of recharge and discharge patterns under varying climatic conditions (Mondani et al., 2022).

2.1.5 Promise Spring

Promise spring is located in the La Thuile municipality. The spring sits at an elevation of approximately 1600 meters a.s.l., within a hydrogeological environment dominated by quaternary glacial sediments atop metamorphic and carbonate bedrock. The spring and meteorological stations collectively provide data essential for understanding groundwater recharge influenced by snowpack dynamics and precipitation seasonality (Mondani et al., 2022).

2.1.6 Meteorological Stations

The hydrogeological monitoring of the five springs in the Aosta Valley is supported by data from several nearby meteorological stations strategically positioned to capture local climate variables. Each spring is paired with one or more meteorological stations that provide continuous measurements of key parameters such as precipitation, air temperature, relative humidity conditions. These stations are equipped with a suite of sensors to continuously measure key atmospheric parameters. Precipitation is recorded using tipping bucket rain gauges (non-heated), which accurately capture rainfall amounts but do not melt snow, making snow measurement reliant on complementary sensors. Air temperature is monitored with digital sensors designed for precise atmospheric readings. Additionally, stations measure relative humidity, wind speed, and wind direction, providing a comprehensive climatic dataset.

Table 2.1: *Selected meteorological stations.*

Meteorological Stations	Springs	Longitude (WGS84)	Latitude (WGS84)	Elevation (m)
Saint Vincent-Terme	Promiod	7.6526	45.7495	626
La Thuile-Villaret	Promise	6.95609	45.7095	1488
Gressoney-la-Trinité-Eselbode	Gabiet	7.82587	45.8306	1642
Roisán-Preyl	Entrebin	7.31667	45.7819	935
Jovençan - Pompiod	Cheserod	7.2653	45.7087	670

2.2 Data Description

The data used in this study were collected and managed by the Department of Environmental, Land and Infrastructure Engineering (DIATI) at Politecnico di Torino. The primary datasets consist of time series measurements of hydrological variables including water level, water temperature, and electrical conductivity for each spring.

For each spring, measurements include:

- Water level (m)
- Temperature (degree Celsius)
- Electrical conductivity (micros/cm)

Meteorological data collected from nearby weather stations include:

- Precipitation (mm)
- Air temperature (degree Celsius)
- Humidity (%)

2.2.1 Data file format

The raw data files are provided as plain text files, with each file corresponding to a specific spring and variable. The files are structured as semicolon-separated values (CSV) with three main columns:

- Date: The date of the measurement in DD/MM/YYYY format.
- Time: The time of the measurement in HH:MM:SS format.
- Value: The recorded measurement value.

The datasets span the period from 2010 to 2024, providing a comprehensive record for the analysis. However, the Entrebin spring exhibits a notable data gap, with approximately two years of missing measurements between July 2014 and February 2016. This interruption is attributed to sensor downtime or maintenance.

```
12/07/2011;01:00:00;---  
12/07/2011;02:00:00;---  
12/07/2011;03:00:00;---  
12/07/2011;04:00:00;---  
12/07/2011;05:00:00;---  
12/07/2011;06:00:00;---  
12/07/2011;07:00:00;---  
12/07/2011;08:00:00;---  
12/07/2011;09:00:00;---  
12/07/2011;10:00:00;---  
12/07/2011;11:00:00;0.029  
12/07/2011;12:00:00;0.030  
12/07/2011;13:00:00;0.030  
12/07/2011;14:00:00;0.030  
12/07/2011;15:00:00;0.030
```

Figure 2.2: *Example of the dataset*

2.2.2 Instrumentation and Measurement Accuracy

Hydrological measurements were obtained using OTT CTD sensors, which simultaneously record water level (pressure), temperature, and electrical conductivity.

According to the manufacturer's specifications, the water level sensors cover ranges from 0 to 4 m up to 0 to 100 m, with a resolution as fine as 0.001 m and an overall accuracy of $\pm 0.05\%$ of full scale, accounting for linearity and hysteresis effects.

The temperature sensors operate reliably within a compensation range from $-5\text{ }^{\circ}\text{C}$ to $+45\text{ }^{\circ}\text{C}$ (ice-free conditions) and can measure temperatures from $-25\text{ }^{\circ}\text{C}$ to $+70\text{ }^{\circ}\text{C}$ with a resolution of $0.01\text{ }^{\circ}\text{C}$ and an accuracy of $\pm 0.1\text{ }^{\circ}\text{C}$.

Electrical conductivity is measured within a range of 0.001 to 2.000 mS/cm or 0.10 to 100.00 mS/cm depending on the model, with a resolution of 0.001 mS/cm (or 0.01 mS/cm in some ranges) and an accuracy of $\pm 1.5\%$ or $\pm 0.5\%$ of the measured value, with a minimum uncertainty of $\pm 0.001\text{ mS/cm}$.

2.3 Data Filtering and Cleaning

Raw field data often contain errors, noise, and missing values due to sensor limitations, environmental interference, or equipment malfunctions. Therefore, data cleaning is a crucial step to ensure the reliability and accuracy of subsequent analyses.

The time series data for each spring and variable were preprocessed following a structured approach:

- **Handling missing and invalid data:** Initial cleaning involved converting the raw measurements to numeric values and removing entries with missing or invalid data points.
- **Outlier detection and smoothing:** To reduce noise and correct for outliers, a Kalman filter was applied to the cleaned time series. The Kalman filter assumes the observed variable follows a nearly constant state model with small process and observation noise, allowing it to smooth fluctuations while preserving underlying trends.
- **Interpolation:** After filtering, polynomial interpolation of order two was applied to fill remaining gaps caused by missing data or removed outliers. This step ensures continuous time series suitable for further analysis.

The data were loaded from semicolon-separated text files with separate Date and Time columns, which were combined and converted into a datetime index to facilitate time series operations. These preprocessing steps were applied independently to each variable, including water level, temperature, electrical conductivity, across all springs.

To address noise and outliers, a univariate Kalman filter was employed. The Kalman filter is a recursive algorithm designed to estimate the true state of a dynamic system from noisy measurements by modeling uncertainties in both the process and the observations. It operates in two steps at each time increment: prediction and update.

The Kalman filter was configured with a simple univariate model, where the system state is assumed to remain nearly constant over time (random walk). This is reflected by setting the state transition matrix to 1, indicating no change in the state between consecutive time steps. The observation matrix was also set to 1, meaning the observed

measurements directly correspond to the true state without any transformation.

The context of the Kalman filter, two key parameters govern the behavior of the algorithm: the transition covariance, also known as process noise, and the observation covariance, often called measurement noise. The transition covariance represents the uncertainty in how the system evolves over time. It models the variability or randomness in the true state's progression, capturing unpredictable changes that the underlying process may experience between time steps. A larger process noise implies the system state can change more abruptly, while a smaller value assumes smoother, more predictable dynamics.

On the other hand, the observation covariance quantifies the uncertainty in the measurements themselves. It reflects the noise and errors inherent in the data collection process, such as sensor inaccuracies or external disturbances. Higher measurement noise indicates less reliable observations, prompting the filter to rely more heavily on the predicted state rather than the raw measurements. Properly balancing these two covariances is crucial for the Kalman filter's effectiveness, as it determines the trade-off between trusting the model's prediction and the incoming observations.

2.4 Trend Analysis

To characterize the temporal behavior of the filtered environmental variables, a combined sinusoidal and linear trend model was employed. This approach allows the simultaneous capture of both periodic seasonal fluctuations and long-term systematic changes within the time series. After data filtering and interpolation, the time series were modeled using a *Fourier series with an added linear trend component*. The Fourier series decomposes the signal into a sum of sinusoidal functions at multiple frequencies, which represent cyclical seasonal patterns of different periods. The linear trend term accounts for any consistent increase or decrease over time.

Mathematically, the model can be expressed as:

$$y(t) = \sum_{i=1}^n A_i \sin(B_i t + C_i) + mt + c \quad (2.1)$$

where A_i , B_i , and C_i are amplitude, frequency, and phase shift parameters of the i -th sinusoidal term, respectively; m is the slope of the linear trend; c is the intercept; and t is time expressed in days since the start of the measurement period.

The model parameters were estimated by nonlinear least squares curve fitting, using initial guesses based on expected seasonal frequencies (e.g., annual, semiannual, multi-year cycles) and a linear trend near zero slope. The curve fitting procedure minimized the residual sum of squares between observed data and the model output.

The goodness of fit was evaluated through the coefficient of determination (R^2), confirming the model's ability to represent the main patterns in the data accurately. The linear trend slope parameter (m) was converted into an annual change rate by multiplying by 365 days, providing a quantitative measure of long-term trends.

After fitting the model, residuals were computed as the difference between the observed values and the modeled outputs. These residuals were then analyzed to evaluate model adequacy and to assess whether the model sufficiently captured the seasonal and long-term structure of the data. A horizontal reference line was added to indicate

the zero level, allowing easy identification of potential over- or under-fitting. This step helped identify any systematic patterns or abrupt deviations not accounted for by the model.

2.5 Correlation Analysis

Understanding the relationships among hydrological variables across different springs is essential for uncovering potential shared recharge mechanisms and climatic influences. Accordingly, both non-parametric correlation and time-lagged cross-correlation analyses were conducted on daily-resolved data from all five springs: Cheserod, Promiod, Gabiet, Entrebin, and Promise.

Data Preprocessing and Resampling

Raw time series data for each variable and spring were initially processed to ensure temporal alignment and consistency. The steps included:

- Combining separate date and time columns into a unified datetime format.
- Converting variable values to numeric, removing non-numeric and missing entries.
- Resampling the cleaned time series to daily mean values using a fixed calendar day.
- Merging the resulting daily series for all variables into a common dataframe, aligned on the datetime index.

This standardized format facilitated direct pairwise comparison of variables across different springs.

Spearman Correlation Analysis

To evaluate the general monotonic relationships among all variables, the Spearman rank correlation matrix was computed. This non-parametric method is robust to outliers and non-linear associations, making it particularly appropriate for environmental time series data.

The resulting matrix highlights variable pairs that tend to increase or decrease concurrently over time, regardless of the magnitude or precise functional relationship. Both intra-spring (within a single spring) and inter-spring (across different springs) correlations were considered in the analysis.

Correlation coefficients were visualized using a heatmap, where warm colors represent strong positive associations, cool colors indicate negative associations, and neutral tones correspond to weak or no correlation. This visualization facilitated the identification of springs exhibiting similar temporal patterns for specific variables.

Cross-Correlation Analysis

Although Spearman correlation is useful for identifying relationships that occur at the same time, it doesn't reveal whether one variable might influence another with a delay. To explore these possible time-lagged connections, a cross-correlation analysis was carried out for all variable pairs across the springs. This helped examine whether a change in one variable could lead or lag behind a corresponding change in another.

For each pair, the cross-correlation function (CCF) was calculated using a lag window of ± 30 days. Before performing the analysis, all time series were standardized through z-score normalization to make them directly comparable. From the CCF, the highest absolute correlation and its associated time lag were identified and documented.

The resulting matrix presents a broad picture of the most significant delayed interactions between the variables. These patterns help highlight possible causal links or delayed reactions, which could reflect differences in subsurface flow paths, recharge timing, or storage behavior among the springs.

To deepen the analysis, some of the cross-correlation curves were plotted, allowing for a closer look at how these lagged relationships play out over time and making it easier to spot meaningful peaks and trends.

3. Results

3.1 Data Quality and Preprocessing

The raw datasets collected from the five monitored springs exhibited, several common issues typical of environmental time series: missing values, abrupt outliers, and sensor-related noise. These problems were caused by harsh alpine conditions, instrument malfunctions, and interruptions in power or data logging.

To ensure the integrity and usability of the temperature and water level time series, a structured preprocessing pipeline was applied. First, invalid and missing entries were removed. Then, a univariate Kalman filter was implemented to smooth the data and reduce measurement noise, while preserving underlying seasonal trends. After filtering, second-order polynomial interpolation was used to reconstruct missing segments, resulting in continuous time series for each spring.

To assess the statistical distribution of the temperature and water level data and the impact of the cleaning process, Q–Q plots were generated for all springs (Figure 3.1). As illustrated for the Cheserod spring, deviations from the reference line indicate a non-normal distribution, particularly at the tails—a pattern observed across all springs (see Appendix A). These characteristics are consistent with bounded, seasonally driven hydrological variables.

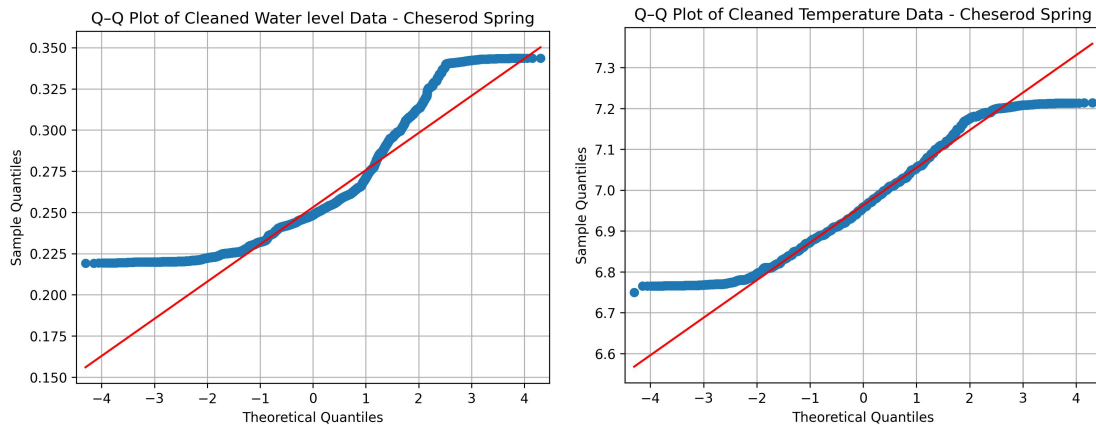


Figure 3.1: Q–Q plot of cleaned temperature and water level data – Cheserod spring.

Following this, a comparison between the original and cleaned temperature and water level series was performed to visually validate the effectiveness of the filtering and interpolation steps. As shown in Figure 3.2, the cleaned data follows the original signal closely while eliminating noise and filling gaps. This preprocessing ensured that the time series were suitable for trend analysis, modeling, and correlation studies.

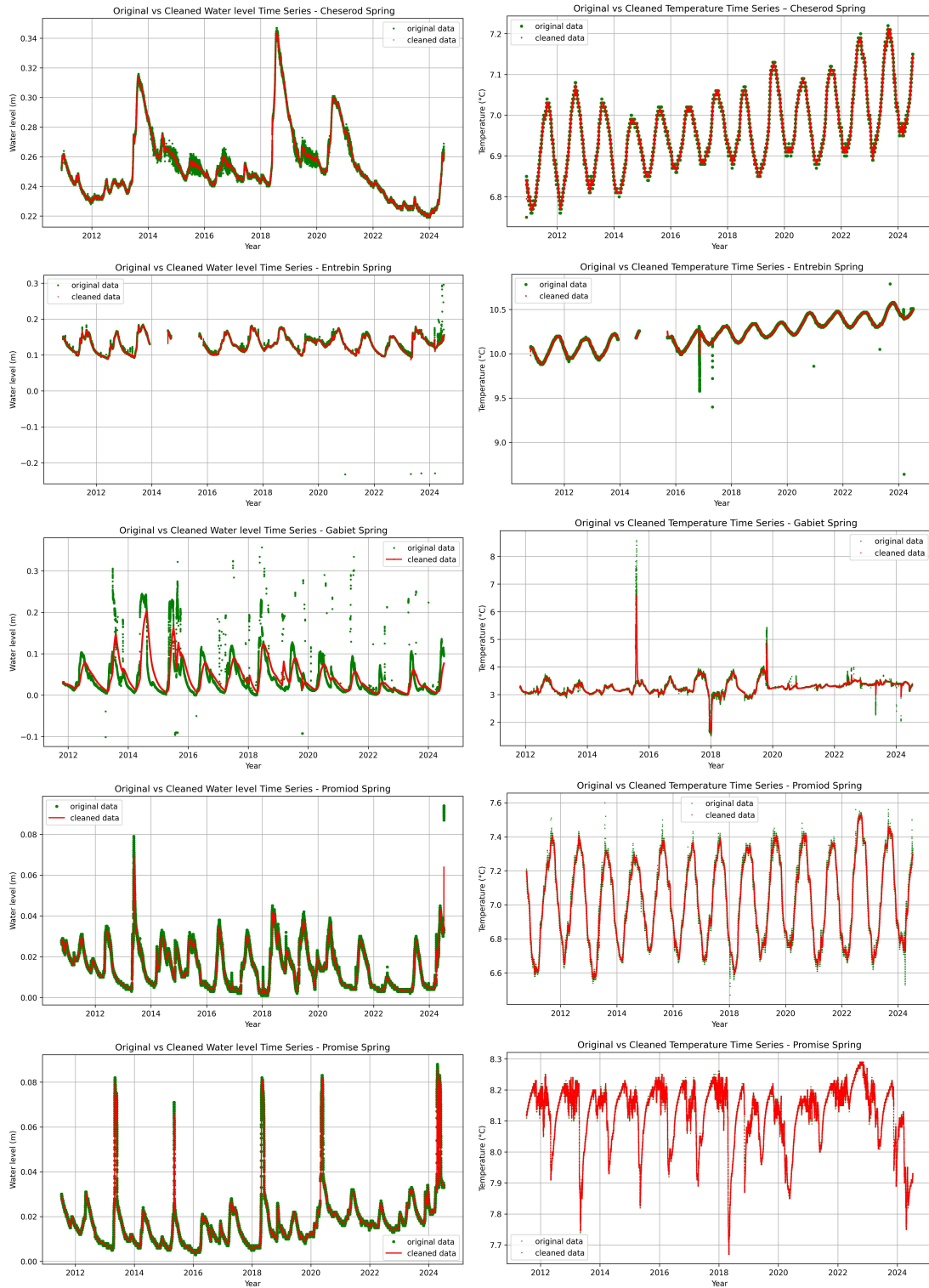


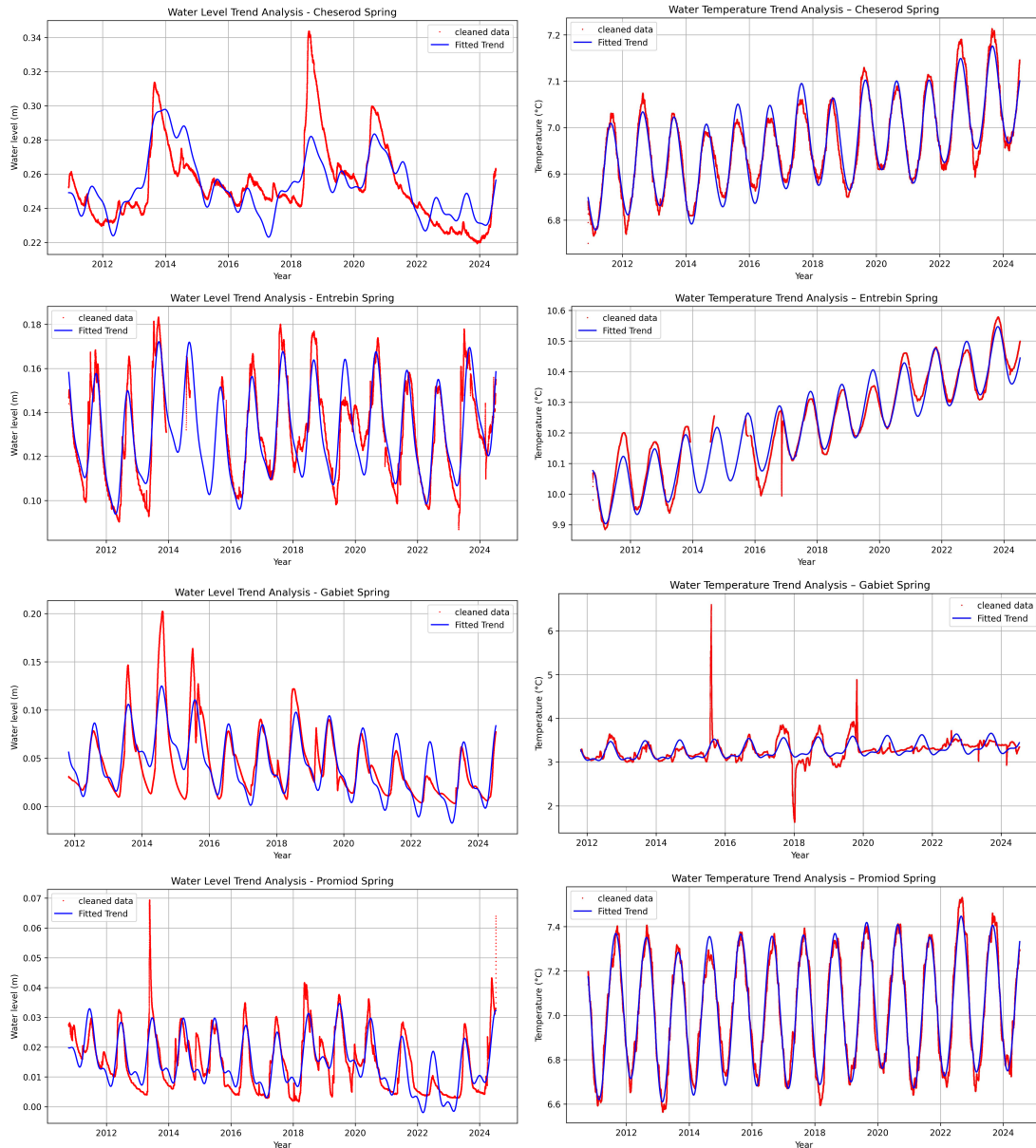
Figure 3.2: Comparison of original and cleaned temperature and water level time series.

3.2 Hydrological Trend Analysis

Figure 3.3 presents the time series of water level and temperature measurements for the five springs studied in the Aosta region. These plots illustrate the seasonal patterns and long-term variations observed between 2010 and 2024.

Each temperature graph is based on cleaned and filtered data, enabling clear visualization of trends without the interference of noise or outliers. The plots exhibit the expected annual cyclic behavior, with higher temperatures during the summer months and lower values in winter. In addition to this seasonal pattern, subtle interannual fluctuations are visible, potentially reflecting the influence of broader climatic variability over the study period.

Each water level plot is similarly derived from pre-processed data and captures both seasonal variations and long-term trends in spring discharge from 2010 to 2024.



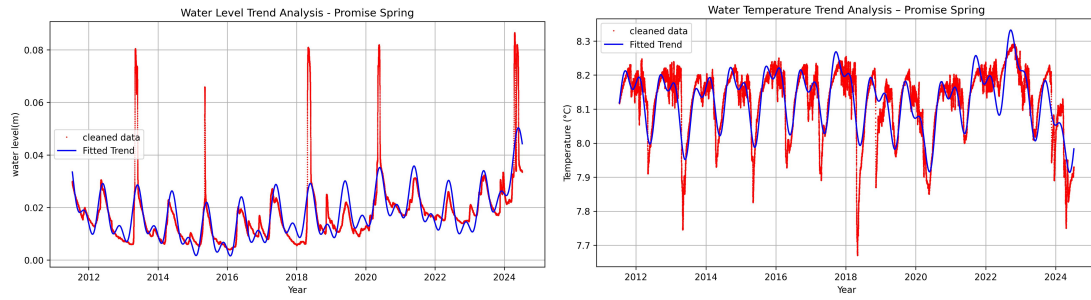


Figure 3.3: Observed and fitted water level and temperature time series for springs, showing seasonal oscillations and the trend.

3.2.1 Residual Analysis

To further evaluate the performance of the sinusoidal model with linear trend, residual analysis was carried out. Residuals were computed as the difference between the observed water level and the model prediction for each time step. Figure A.1 in the Appendix shows the residual time series for the Cheserod spring. The residuals fluctuated closely around zero throughout the observation period, with no clear long-term bias, indicating a satisfactory model fit. Occasional deviations, visible as spikes in the residual series, were observed during abrupt changes in water level. These may correspond to short-term recharge events (e.g., snowmelt or rainfall) that are not fully captured by the relatively smooth Fourier model. A red dashed line was used as a visual reference for the zero-residual level, facilitating the identification of over- or under-estimation periods.

Overall, the residuals support the model's ability to represent the main seasonal and trend dynamics in the water level of the Cheserod spring, while also highlighting its limitations in capturing high-frequency or irregular variations. Residual plots for other springs and variables are provided in Appendix A to support broader evaluation of model performance across the dataset.

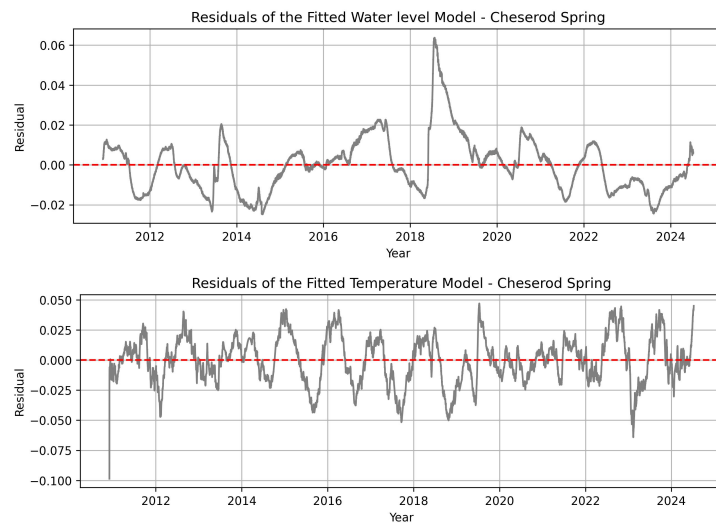


Figure 3.4: Residuals between observed and modeled water level and temperature for the Cheserod spring. The red dashed line indicates the zero-residual reference level.

3.3 Cross-Correlation Results

Cross-correlation analysis was performed between all available hydro-meteorological variables within each spring. The time series were first resampled to daily means, and pairwise cross-correlations were computed over a lag range of ± 30 days.

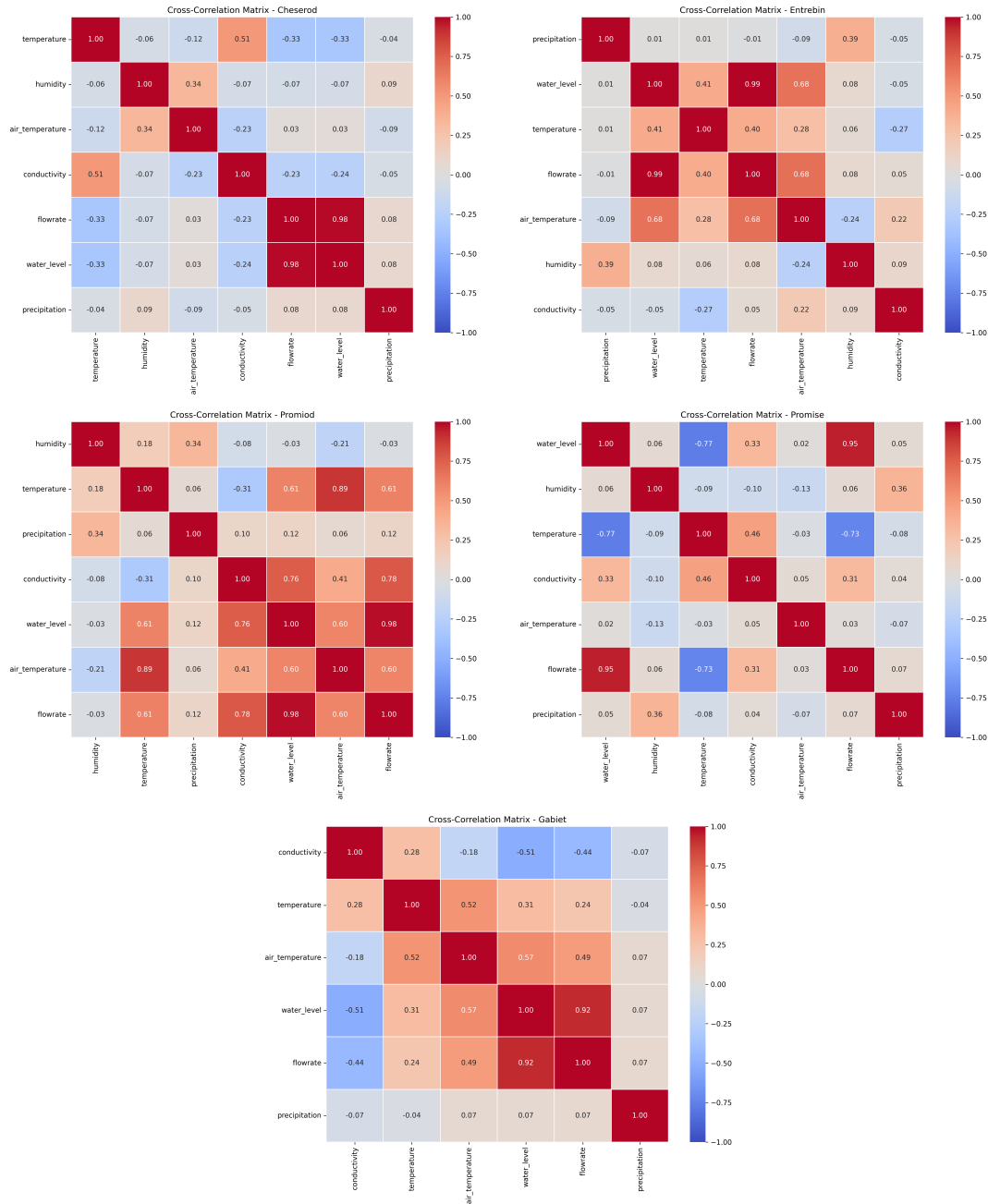


Figure 3.5: *Observed cross correlation for each spring*

To better understand the delayed responses of the springs to climatic and hydrological influences, a cross-correlation analysis was conducted to identify the time lags between key variable pairs. For each combination, the correlation was computed across a range of lags, and the lag corresponding to the highest absolute correlation was identified. The resulting plots illustrate how correlation values change with varying lags.

A positive lag indicates that the first variable precedes the second in time. The findings demonstrate that the springs do not respond instantaneously to external factors; instead, their reactions are often delayed. For example, in the Entrebin spring, the water level exhibited the strongest correlation with flowrate at a lag of 0 days, with a coefficient of 0.99.

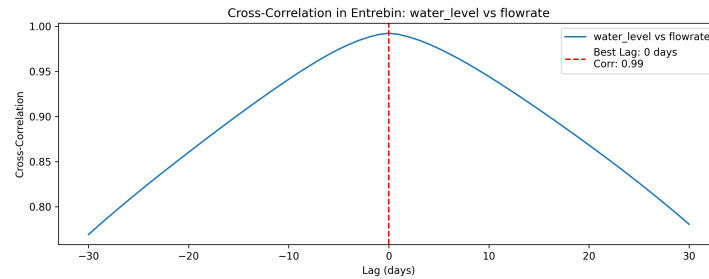


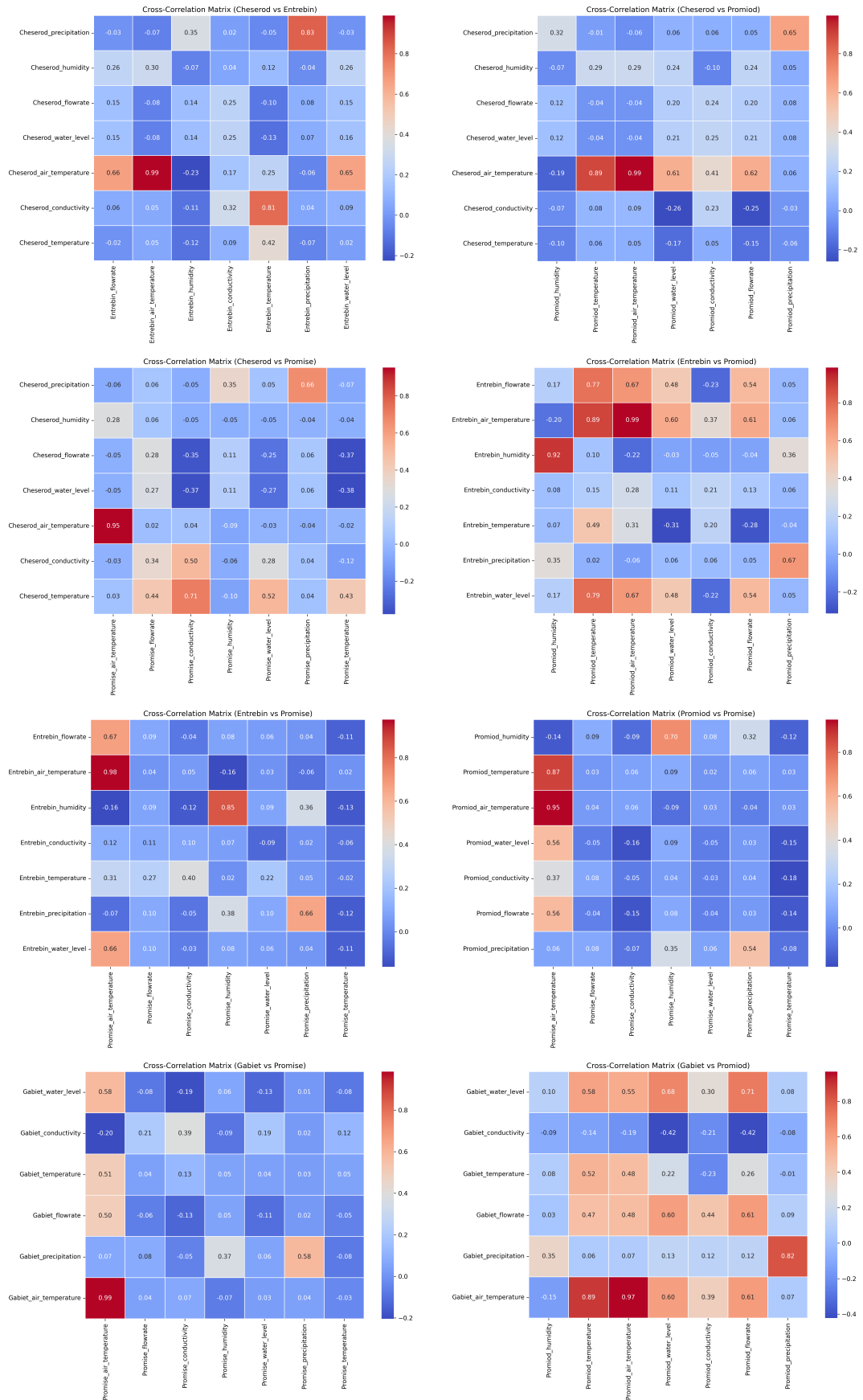
Figure 3.6: *Time-lagged cross-correlation plots between spring variables. The red dashed line marks the lag with the strongest correlation.*

3.3.1 Cross-Correlation Between Springs

Cross-correlation analysis was performed between identical variables measured at different springs. All time series were resampled to daily means to ensure temporal alignment. For each pair of springs and variables, the cross-correlation was computed over a lag range of ± 30 days, and the maximum absolute correlation coefficient and its corresponding lag were recorded.

The results are summarized in a heatmap (Figure 3.6), where each cell indicates the highest correlation value observed between a pair of springs for a given variable. This format provides a visual overview of the degree of similarity in hydrological behavior across the different springs.

Spring pairs with strong correlations appear with warmer colors, while low or negative correlations are represented with cooler tones.



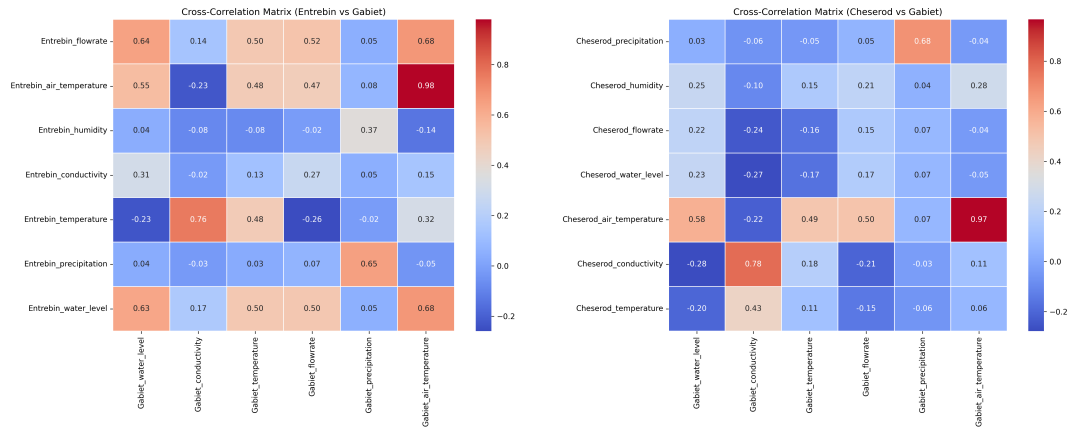


Figure 3.7: Heatmap showing the maximum absolute cross-correlation values between identical variables across different springs. Colors represent correlation strength; values are computed over a ± 30 day lag window.

3.3.2 Time Lag Analysis Between Springs

To evaluate the delayed influence between different springs, cross-correlation analysis was performed on their key variables. For each pair, the correlation was calculated across a range of time lags and the lag with the highest absolute correlation was identified. A positive lag indicates that variations in the first spring precede those in the second. Results show that springs do not respond simultaneously; instead, changes in one may appear in another after several days.

4. Discussion and Conclusion

4.1 Interpretation of Results

This study analyzed 14 years (2010–2024) of hydrological data from five alpine springs in the Aosta Valley—Gabiet, Promiod, Promise, Entrebin, and Cheserod—using advanced time series modeling and correlation analysis. The primary goal was to understand the temporal trends and inter-variable relationships affecting spring behavior under the influence of environmental drivers.

Temporal Trends and Long-Term Changes

sinusoidal models with linear trend components were applied to all spring's variables. For temperature in the Cheserod spring, the model demonstrated a strong fit, with a coefficient of determination of $R^2 = 0.953$, indicating that the seasonal and linear trend components explain a substantial portion of the temperature variability. The estimated rate of temperature change was approximately 0.0120°C per year, resulting in a total increase of about 0.1637°C over the 4967 days observation period (approximately 13.6 years). This reflects a subtle but consistent warming trend in the thermal behavior of the spring.

For water level, the model achieved a moderate fit, with a coefficient of determination of $R^2 = 0.7060$, suggesting that the seasonal and trend components account for a significant portion of the variability in the data. The estimated rate of change in water level was -0.0003 m per year, leading to a total decrease of approximately -0.0041 m over the same period. This indicates a very slight but consistent decline in water level during the monitoring period.

For temperature in the Entrebin spring, the model demonstrated a strong fit, with a coefficient of determination of $R^2 = 0.9502$, indicating that the seasonal and trend components explain most of the temperature variability. The estimated rate of temperature change was 0.0349°C per year, corresponding to a total increase of approximately 0.4792°C over the full observation period of 5012 days (about 13.7 years).

For water level, the model also showed a good fit, with a coefficient of determination of $R^2 = 0.8132$, suggesting that seasonal fluctuations and the linear trend account for a substantial portion of the observed variability. The estimated rate of change in water level was 0.0004 m per year, resulting in a total increase of approximately 0.0050 m over the same observation period. This suggests a slight but steady rise in water level.

For temperature in the Gabiet spring, the model fit was considerably weaker compared to other springs, with a coefficient of determination of $R^2 = 0.3021$. This suggests that

the seasonal and linear trend components only partially explain the observed variability in temperature, indicating the presence of additional factors or a higher degree of short-term variability in the thermal signal. Despite the lower model performance, the estimated rate of temperature change was 0.0177°C per year, resulting in a total temperature increase of approximately 0.2246°C over the observation period of 4641 days (around 12.7 years).

For water level, the model similarly exhibited a weak fit, with a coefficient of determination of $R^2 = 0.3099$. This low value indicates that the model captures only a limited portion of the observed variability, pointing to the influence of irregular or non-seasonal factors, such as glacial melt pulses or anthropogenic impacts. The estimated rate of change in water level was -0.0038 m per year, corresponding to a total decrease of approximately -0.0485 m over the same period. This reflects a slight but persistent decline in water level over time.

For temperature in the Promiod spring, the model exhibited a strong fit, with a coefficient of determination of $R^2 = 0.9576$, indicating that the seasonal and long-term trend patterns account for most of the variability in the observed data. The estimated rate of temperature change was 0.0063°C per year, corresponding to a total increase of approximately 0.0860°C over the full observation period of 5017 days (about 13.7 years). This reflects a minor but steady warming trend in the thermal behavior of the spring. For water level, the model showed a moderate fit, with a coefficient of determination of $R^2 = 0.6991$, indicating that the seasonal and trend components explain a considerable portion of the variability in the water level data. The estimated rate of change was -0.0004 m per year, resulting in a total decrease of approximately -0.0058 m throughout the observation period.

For temperature in the Promise spring, the model achieved a moderate fit, with a coefficient of determination of $R^2 = 0.7213$, suggesting that while seasonal and trend components explain a substantial part of the variability, other environmental or hydrogeological factors may also influence the temperature dynamics. The estimated rate of temperature change was -0.0023°C per year, indicating a slight cooling trend over the observation period. The total change amounted to approximately -0.0298°C across 4745 days (around 13 years).

For water level, the model also achieved a moderate fit, with a coefficient of determination of $R^2 = 0.5474$, indicating that seasonal and trend components explain a fair portion of the variability, though other factors may contribute to water level fluctuations. The estimated rate of change was -0.0001 m per year, resulting in a total decrease of approximately -0.0017 m over the same observation period. This suggests a very slight but consistent declining trend in water level at the Promise spring.

Intra-Spring Correlation Analysis

To better understand the internal hydroclimatic dynamics of each spring, correlation matrices were computed among the key measured variables: water level, flow rate, water temperature, conductivity, air temperature, and precipitation. This analysis provided insight into the dominant interactions between environmental drivers and spring responses within each system.

For the Cheserod spring, strong positive correlations were observed between water level and flow rate, consistent with their shared dependence on aquifer recharge conditions. A significant correlation was also found between water temperature and air temperature, indicating a thermal sensitivity to atmospheric conditions. Conductivity, on the other hand, showed a weak negative correlation with water level and precipitation, suggesting a dilution effect during high-recharge periods. These patterns reflect a system where surface inputs and atmospheric conditions directly influence both quantity and quality indicators.

For the Entrebin spring, the correlation between water temperature and air temperature was particularly strong, supporting the findings from the trend analysis that this spring exhibits high thermal responsiveness. Flow rate and water level were also highly correlated, suggesting efficient and immediate hydrological connectivity. Conductivity exhibited a moderate negative correlation with both water level and flow rate, further supporting the hypothesis of dilution effects during high-discharge events.

In the Gabiet spring, correlation patterns were more complex and generally weaker. While a positive relationship between water level and flow rate was still evident, correlations involving temperature and conductivity were more variable and less pronounced. This may reflect the influence of non-linear processes such as glacial melt inputs or variable recharge mechanisms, which introduce irregularities in the hydrological and thermal signals.

The Promiod spring displayed relatively strong and consistent correlations. Flow rate and water level were again tightly linked, while temperature and air temperature were moderately correlated. Notably, conductivity showed a positive correlation with flow rate, which may indicate mineral mobilization during higher flows, suggesting different geochemical dynamics compared to other springs.

Finally, the Promise spring presented an unusual pattern: a negative correlation between water temperature and water level, indicating that higher discharge periods may be associated with colder water inputs — possibly from snowmelt or deeper groundwater. While flow rate and water level remained positively correlated, other relationships were weaker or more variable, pointing to more complex or isolated hydrological behavior.

Overall, the intra-spring correlation analysis highlights the diversity of spring behavior even within a relatively small geographic region, reflecting differences in aquifer properties, recharge sources, and exposure to climatic forcing.

Cross-Correlation Analysis Between Springs

The cross-correlation analyses performed between the various springs across the Aosta Valley provide a detailed understanding of the spatial coherence and hydrometeorological interactions among the monitored sites. By examining the relationships between key environmental variables—such as temperature, flowrate, water level, precipitation, conductivity, and humidity—the study revealed both atmospheric synchrony

and hydrogeological divergence among the springs.

The cross-correlation analysis between Cheserod and Entrebin springs reveals strong atmospheric coherence and moderate hydrological connectivity. The most prominent correlation is observed between the air temperature at both sites, with a near-perfect correlation coefficient of $r = 0.99$, indicating that both springs are exposed to virtually identical climatic conditions. Additionally, Cheserod's air temperature is strongly correlated with Entrebin's flow rate $r = 0.66$ and water temperature $r = 0.65$, suggesting that climatic variability at Cheserod may have a direct and almost synchronous influence on the hydrological and thermal response at Entrebin.

Conductivity values at both springs also show a relatively strong positive correlation $r = 0.81$, pointing to similar geochemical influences or shared recharge sources with comparable mineral content. However, correlations between flow rate and water level variables are generally weaker (typically below $r = 0.25$), indicating only partial synchronization in aquifer response or differing recharge dynamics.

The cross-correlation matrix between Cheserod and Gabiet springs highlights a high degree of atmospheric coherence and modest hydrological connectivity. The strongest correlation is observed between air temperature at the two sites, with a coefficient of $r = 0.97$, indicating almost identical thermal conditions. This is consistent with both springs being located in similar climatic zones and exposed to the same synoptic-scale weather patterns.

A notable secondary correlation is seen between Cheserod's conductivity and Gabiet's conductivity $r = 0.78$, suggesting potentially similar geological substrates or shared mineral sources within the recharge area. However, water-related parameters—such as flow rate and water level—exhibit generally low to negative correlations (typically between $r = -0.27$ and $r = 0.23$), indicating that despite similar atmospheric input, the two springs operate within hydraulically distinct systems. This is further supported by the weak correlation between temperature series at the two sites $r = 0.43$, implying differences in thermal buffering or groundwater residence time.

The cross-correlation matrix between Cheserod and Promiod springs shows a very high level of climatic similarity and weak-to-moderate hydrological coupling. The strongest observed relationship is between air temperatures at the two sites, with a correlation coefficient of $r = 0.99$, confirming an almost identical thermal regime. Additionally, Cheserod's air temperature is highly correlated with Promiod's temperature $r = 0.89$, flowrate $r = 0.62$, and water level $r = 0.61$, suggesting that atmospheric forcing at Cheserod influences the hydrological dynamics at Promiod with relatively little delay or attenuation.

Other parameters show limited connectivity. For instance, conductivity and humidity variables present generally weak correlations (mostly $r < 0.30$), while precipitation at Cheserod is only modestly correlated with Promiod's precipitation $r = 0.65$, hinting at spatial variability in rainfall distribution or localized recharge characteristics. Correlations between flow and water level are also low $r \approx 0.20$, indicating that while both springs may respond to common meteorological inputs, their aquifer systems behave independently.

The cross-correlation matrix between Cheserod and Promise reveals a weaker overall

relationship compared to other spring pairings, particularly in hydrological parameters. The highest correlation is between air temperature at Cheserod and air temperature at Promise $r = 0.95$, confirming a shared climatic regime in terms of atmospheric conditions. Additionally, Cheserod's temperature shows moderate correlations with Promise's conductivity $r = 0.50$ and temperature $r = 0.43$, indicating a partial alignment in thermal and geochemical behavior.

On the other hand, water-related parameters show very weak or even negative correlations. For instance, Cheserod's water level and flowrate exhibit negative correlations with Promise's corresponding variables (e.g., $r = -0.37$ for water level and $r = -0.35$ for flowrate), suggesting distinct aquifer dynamics and response characteristics between the two sites. Correlations for precipitation and humidity are generally close to zero, reflecting either a lack of shared meteorological forcing or localized variations in precipitation distribution and infiltration behavior.

The cross-correlation matrix between Entrebin and Gabiet reveals strong synchronization in several hydrometeorological and physicochemical variables, suggesting a potential hydraulic or climatic linkage between the two springs. The highest correlation is observed between air temperature at Entrebin and Gabiet $r = 0.98$, indicating nearly identical atmospheric conditions, possibly due to geographic proximity and similar elevation.

Flow-related variables show notable alignment as well: Entrebin's water level and flowrate correlate strongly with Gabiet's water level $r = 0.63$ and $r = 0.64$ and moderately with Gabiet's flowrate $r = 0.50$ and $r = 0.52$. These results suggest shared recharge timing or connected aquifer systems responding similarly to climatic inputs.

Entrebin's temperature also shows a relatively high correlation with Gabiet's conductivity $r = 0.76$, possibly reflecting similar geochemical evolution processes influenced by temperature-driven mineral dissolution. Additionally, Entrebin's temperature correlates moderately with Gabiet's temperature $r = 0.48$ and flowrate $r = 0.48$, reinforcing the notion of thermal and hydrological synchronization.

In contrast, humidity and precipitation correlations are weak to moderate across the matrix, suggesting these parameters may be more localized or affected by microclimatic differences.

The cross-correlation analysis between Entrebin and Promiod springs indicates strong climatic consistency and moderate hydrological synchronization. The most dominant correlation is found between the air temperatures at both sites $r = 0.99$, demonstrating a nearly identical atmospheric temperature profile. Furthermore, Entrebin air temperature is also strongly correlated with Promiod's temperature $r = 0.89$, flowrate $r = 0.61$, and water level $r = 0.60$, suggesting a shared seasonal response to climate variability. Water-related variables, such as Entrebin's flowrate and water level, also correlate strongly with Promiod's temperature $r = 0.77$ and $r = 0.79$, respectively, indicating that temperature fluctuations may drive comparable hydrological responses in both springs. These patterns suggest a possible similarity in recharge timing or aquifer behavior.

Interestingly, Entrebin's humidity is highly correlated with Promiod's humidity $r = 0.92$, reinforcing the notion of a shared microclimatic environment. In contrast, the correlation for other variables such as conductivity and precipitation remains low to moderate ($r < 0.30$ in most cases), which could reflect localized differences in subsur-

face properties or rainfall distribution.

The cross-correlation matrix between Entrebin and Promise reveals mixed patterns of interaction, with strong climatic agreement but weaker hydrological and geochemical correlations. The most significant relationship is seen in air temperature, where Entrebin and Promise exhibit a very strong correlation $r = 0.98$, indicating nearly identical atmospheric temperature variations and confirming regional climatic coherence.

Water level and flowrate correlations are moderately strong: Entrebin's flowrate and water level show good agreement with Promise's air temperature $r = 0.67$ and $r = 0.66$, respectively, implying that air temperature plays a key role in driving hydrological responses across both springs. However, the direct correlations between corresponding hydrological variables (e.g., flowrate-to-flowrate or water level-to-water level) remain weak or near zero, suggesting distinct aquifer characteristics or local recharge behaviors.

One particularly notable correlation is between Entrebin's humidity and Promise's humidity $r = 0.85$, reinforcing the consistency of microclimatic conditions across the two sites. In contrast, correlations involving conductivity, temperature, and precipitation are generally weak $r < 0.40$, reflecting either localized geochemical evolution or differences in catchment-specific meteorological inputs.

The cross-correlation matrix between Gabiet and Promiod springs highlights strong atmospheric consistency, with moderate to high hydrological alignment and weaker geochemical connectivity. The air temperature at Gabiet is highly correlated with that of Promiod $r = 0.97$, confirming synchronized seasonal temperature patterns across the two locations. Additionally, strong correlations are observed between Gabiet air temperature and Promiod temperature $r = 0.89$, flowrate $r = 0.61$, and water level $r = 0.60$, implying that rising air temperatures likely drive similar hydrological responses at both springs.

Hydrological indicators such as water level and flowrate also show significant cross-correlations. Gabiet's water level strongly aligns with Promiod's flowrate $r = 0.71$, water level $r = 0.68$, and temperature $r = 0.58$, suggesting a parallel seasonal response or possibly similar aquifer recharge timing. Gabiet flowrate also correlates notably with Promiod's flowrate $r = 0.61$ and water level $r = 0.60$, further emphasizing this connection.

Promiod precipitation demonstrates a strong correlation with Gabiet precipitation $r = 0.82$, reinforcing the notion of shared climatic input. On the other hand, Gabiet conductivity exhibits weak or negative correlations with all Promiod variables (as low as $r = -0.42$), suggesting distinct geochemical characteristics or different lithological influences between the two catchments.

The cross-correlation matrix between Gabiet and Promise springs shows a clear alignment in atmospheric conditions, but weaker hydrological and geochemical links. The air temperature at Gabiet is almost perfectly correlated with that of Promise $r = 0.99$, highlighting synchronized regional temperature variations. A moderate correlation also appears between precipitation values from both springs $r = 0.58$, reinforcing the presence of shared meteorological patterns.

Gabiet's water level demonstrates a moderate correlation with Promise air temperature $r = 0.58$, which may reflect temperature-dependent effects on evapotranspiration or

snowmelt recharge. Similarly, Gabiet's flowrate $r = 0.50$ and temperature $r = 0.51$ show moderate correlations with Promise air temperature, suggesting parallel seasonal responses driven by regional atmospheric forcing.

However, other variable pairs show notably weaker or inconsistent relationships. For instance, conductivity values are generally uncorrelated or weakly associated, and Gabiet's conductivity even shows a slight negative correlation with Promise air temperature $r = -0.20$. Flowrate, water level, and temperature at Gabiet do not show strong direct associations with the corresponding hydrochemical variables in Promise, with most correlations falling below $r = 0.10$.

The cross-correlation matrix between Promiod and Promise springs reveals strong atmospheric synchronization but weak hydrogeochemical connectivity. The air temperatures of the two springs are highly correlated $r = 0.95$, consistent with regional climate forcing. Promiod temperature also shows a strong correlation with Promise air temperature $r = 0.87$, reinforcing the dominant influence of shared seasonal patterns across the region.

A moderate relationship is observed between Promiod humidity and Promise humidity $r = 0.70$, suggesting similar atmospheric moisture conditions. Promiod water level and flowrate both show moderate correlation with Promise air temperature $r = 0.56$, possibly reflecting indirect climatic influence on local hydrology. However, their relationships with Promise's own water level and flowrate are near zero or even slightly negative, highlighting localized hydrogeological processes at each spring.

Precipitation correlations are generally modest. Promiod precipitation shows a moderate correlation with Promise precipitation $r = 0.54$, aligning with shared meteorological input. However, most hydrochemical variables—including conductivity and water temperature—exhibit weak or negative correlations, indicating divergent geochemical behaviors between the two sites.

In summary, while the region exhibits high atmospheric synchrony, hydrological and geochemical responses are far more heterogeneous and reflect the influence of local hydrogeological settings. These results underscore the importance of considering both climatic drivers and subsurface properties when evaluating spring behavior and managing groundwater resources in mountainous regions.

4.2 Impact of Climate Change

The results of this study reveal subtle but persistent trends consistent with the expected impacts of climate change on mountainous hydrological systems. Across all springs, long-term temperature trends indicate a gradual warming, with estimated rates of change ranging from approximately 0.0063°C to 0.0177°C per year. This regional warming is further supported by the nearly perfect cross-correlation of air temperature between all springs ($r > 0.95$), highlighting a shared climatic signal likely driven by broader atmospheric trends such as increasing air temperatures and altered seasonal patterns.

These thermal changes have potential cascading effects on spring hydrology. Slight but consistent declines in water levels were observed in several springs—most notably Gabiet and Promiod—suggesting a possible reduction in recharge or a shift in snowmelt

timing. While individual flowrate and water level correlations between springs are often modest, the overall pattern points to more irregular and localized hydrological responses, potentially exacerbated by climate-driven changes in precipitation distribution, evapotranspiration, and snowpack dynamics.

Moreover, the limited synchronization in geochemical parameters such as conductivity suggests that warming may be influencing subsurface water–rock interactions in spring-specific ways, driven by variations in groundwater residence times, flow paths, or recharge sources. As climate change continues to affect temperature and precipitation regimes in alpine environments, such diverging spring behaviors may become more pronounced, necessitating locally adapted monitoring and management strategies.

In conclusion, the data support the hypothesis that climate change is already exerting measurable influence on both the thermal and hydrological dynamics of alpine springs, emphasizing the importance of long-term, high-resolution monitoring for future water resource planning and ecological conservation in mountainous regions.

4.3 Limitations of the Study

While this study provides valuable insights into the thermal and hydrological dynamics of alpine springs in the Aosta region, several limitations must be acknowledged. First, the analysis is based on available historical data from 2010 to 2024, which, despite its multi-year span, may not fully capture the long-term variability or rare extreme events that can strongly influence hydrological systems. Additionally, some variables (e.g., humidity) were missing for Gabiet, limiting the completeness and consistency of cross-correlation analyses.

Second, while sinusoidal models with linear trends and Kalman filtering were effective for identifying seasonal and long-term patterns, they assume relatively smooth and regular behavior. These methods may not fully capture sudden changes, nonlinear trends, or multi-scale variability that could arise from shifts in climate, land use, or hydrogeological disturbances. Furthermore, while cross-correlation analyses offer a valuable view of synchronous and lagged relationships, they do not establish causality, and are sensitive to data gaps, outliers, and autocorrelation in the time series.

Third, the spatial representativeness of the selected springs may be limited. Although the five springs studied are distributed across different parts of the Aosta Valley, they may not reflect the full diversity of hydrogeological conditions in the region. Local geological heterogeneities, anthropogenic influences, or site-specific recharge mechanisms may introduce behaviors not generalizable to other locations.

Finally, this study does not include direct modeling of future climate scenarios or water demand projections. As a result, while current trends are highlighted, their future implications under different climate trajectories remain uncertain. Future research could benefit from the integration of physically-based hydrological models, coupled climate-hydrology simulations, and the inclusion of socioeconomic factors affecting water use and management.

4.4 Recommendations for Future Research

This study has demonstrated the value of high-resolution (hourly) data in capturing seasonal, interannual, and cross-spring hydrological dynamics. Building on these results, future research could further enhance temporal analysis by integrating additional environmental variables such as snow depth, evapotranspiration, or groundwater level from deeper observation wells. Such additions would provide a more comprehensive view of the processes governing spring dynamics.

Future studies should also consider coupling the statistical models used here with physically-based hydrological models (e.g., MODFLOW, MIKE SHE, or SWAT) to simulate flow paths, recharge dynamics, and responses to extreme weather events. This would help move from descriptive analysis toward process-based understanding and prediction.

Expanding spatial coverage to include more springs with varied geological and topographic settings would strengthen the regional interpretation of results. Incorporating hydrochemical or isotopic tracers could also provide insight into groundwater residence times and mixing processes—key factors in understanding system resilience to climate change.

Finally, integrating climate projection data and land-use change scenarios would allow for forecasting how alpine spring systems may evolve under future environmental pressures. This would support adaptive water resource planning, especially in regions where spring-fed water is critical for ecosystems and human consumption.

Bibliography

- [1] Atkinson TC (1977) Hydrogeological properties of karstic aquifers. In: Ford AP, Williams CH (eds) *Karst Hydrology and Physical Speleology*. Springer, Dordrecht, pp 53–72
- [2] Barbieri M, Lo Russo S, Guida D (2021) Impact of climate change on hydrological regimes of mountain springs: a case study in the Italian Alps. *Environmental Earth Sciences* 80(7):1–15
- [3] Duratorre J, Padilla J, Pulido-Bosch A (2020) Non-stationarity in groundwater systems under climate variability. *Hydrogeology Journal* 28(4):1079–1093
- [4] Fiorillo F (2009) Hydrograph analysis and karst aquifer characterization in southern Italy. *Hydrogeology Journal* 17(5):1187–1202
- [5] Gizzi FT, Lo Russo S, Sappa G (2023) Semi-automatic tools for spring-monitoring data analysis: enhancing hydrogeological characterizations. *Hydrological Sciences Journal* 68(1):1–15
- [6] Lo Russo S, Maréchal J-C (2014) Recession hydrographs and time series analysis of springs monitoring data. *Hydrogeology Journal* 22(3):789–802
- [7] Mangin A (1975) Contribution à l'étude hydrodynamique des aquifères karstiques. Université de Dijon, Dijon, France
- [8] Padilla J, Pulido-Bosch A (1994) Application of autocorrelation and cross-correlation analysis to the study of karstic spring discharge. *Journal of Hydrology* 156(1-4):285–303
- [9] Gizzi M, Vagnon F(2023) A review of groundwater heat pump systems in the Italian framework: technological potential and environmental limits. *Energies* 16:4813.
- [10] Gizzi M, Mondani M (2022) Aosta Valley mountain springs: a preliminary analysis for understanding variations in water resource availability under climate change. *Water* 14:1004
- [11] Grewal MS, Andrews AP (2010) Applications of Kalman filtering in aerospace 1960 to the present. *IEEE Control Systems Magazine* 30(3):69–78
- [12] Barbieri M, Barberio MD (2023) Climate change and its effect on groundwater quality. *Environmental Geochemistry and Health* 45:1133–1144

- [13] Gizzi M, Narcisi R (2023) Comprehending mountain springs' hydrogeological perspectives under climate change in Aosta Valley (northwestern Italy): new automated tools and simplified approaches. *Italian Journal of Engineering Geology and Environment, Special Issue* 1:73–80.
- [14] Duratorre T, Bombelli GM (2020) Hydropower potential in the Alps under climate change scenarios: the Chavonne plant, Val d'Aosta. *Water* 12:2011.
- [15] Cerino Abdin E, Taddia G (2021) Reliability of spring recession curve analysis as a function of the temporal resolution of the monitoring dataset. *Environmental Earth Sciences* 80:249
- [16] Mondani M, Gizzi M, Taddia G (2022) Role of snowpack-hydrometeorological sensors for hydrogeological system comprehension inside an alpine closed-basin. *Sensors* 22:7130.
- [17] Luino F (2005) Sequence of instability processes triggered by heavy rainfall in northern Italy. *Geomorphology* 66(1-4):13–39.
- [18] Viglietti D, Letey S (2010) Snow avalanche release in forest ecosystems: a case study in the Aosta Valley region (NW-Italy). *Cold Regions Science and Technology* 64:167–173
- [19] Marti S (2004) The subducted Tethys in the Aosta Valley (Italian Western Alps). In: Pre-Congress, Florence, Italy, August 20-28, 2004, p. B02

Appendix A

This appendix contains the plots referenced in the main chapters.

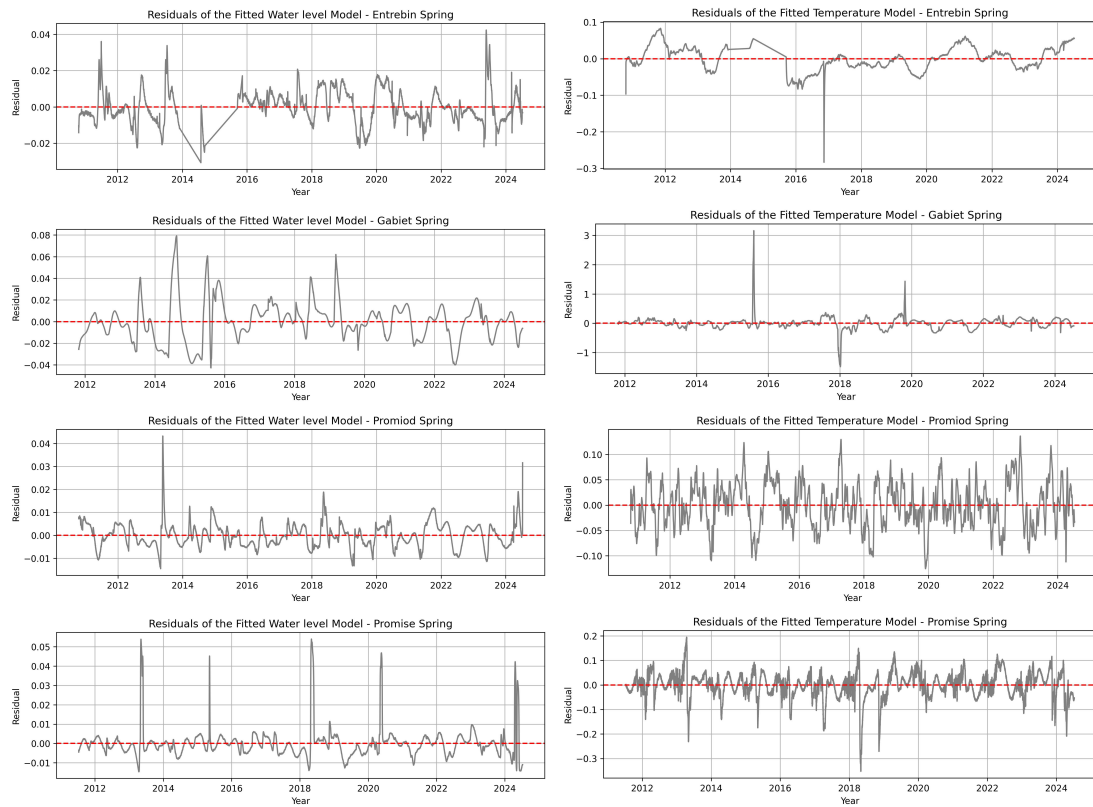


Figure 4.2: *Residuals of the fitted model springs*

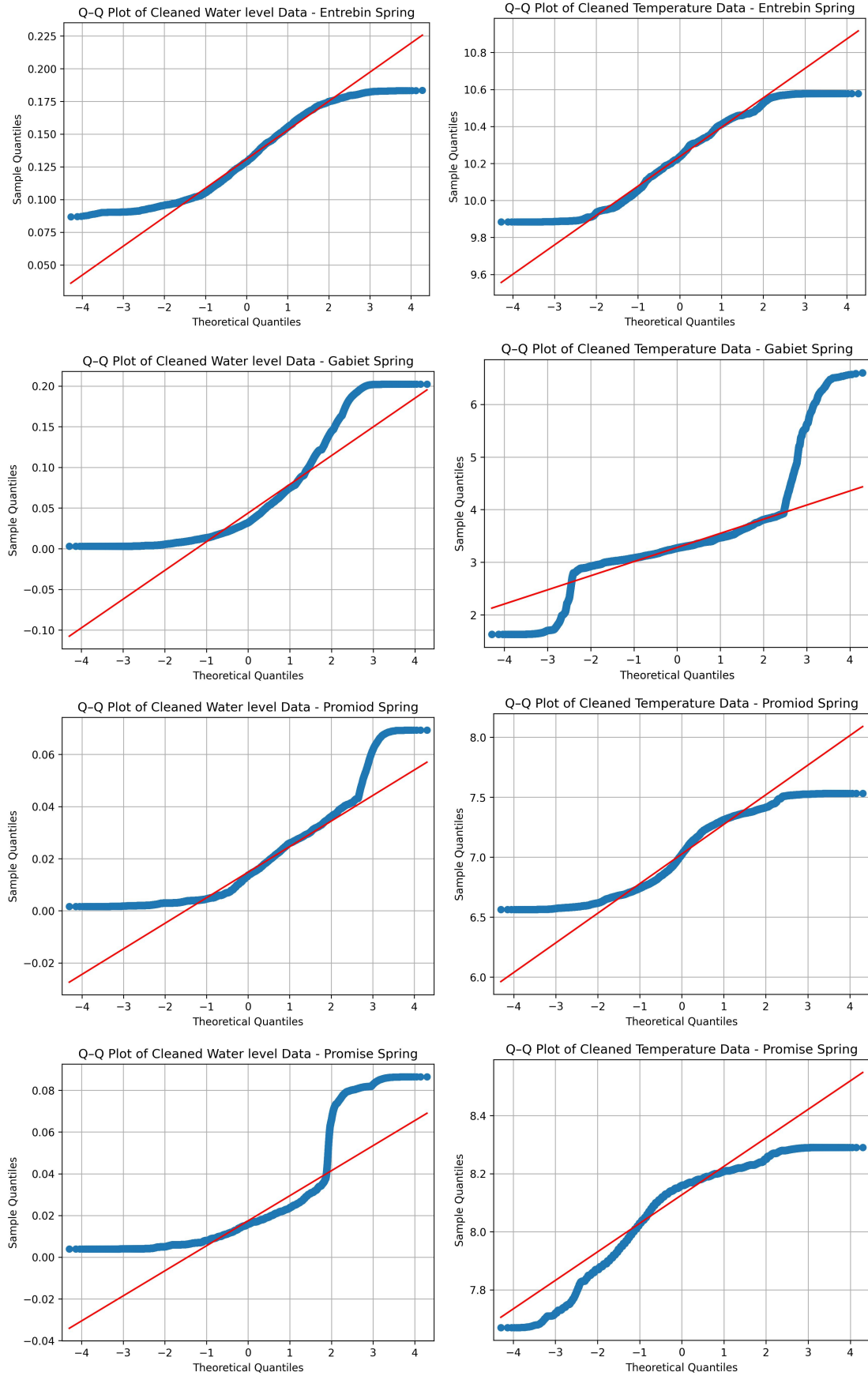


Figure 4.1: Q-Q plot of cleaned data springs

Appendix B

B.1: Importing Packages and Extracting Data

The following script imports the required Python libraries, extracts ZIP data containing the spring measurements, and defines a list of spring names used throughout the analysis.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import os
5 import glob
6 import zipfile
7 from numpy.polynomial.polynomial import Polynomial
8 from scipy import stats
9
10 zip_file_path = '/content/springs-data.zip'
11
12 with zipfile.ZipFile(zip_file_path, 'r') as zip_ref:
13     zip_ref.extractall()
```

B.2: Cleaning and Fitting Temperature Data

The following Python script performs data cleaning and modeling for the Cheserod spring temperature and water level time series. It applies a Kalman filter, interpolates missing values, fits a Fourier series with a linear trend, generates visualizations (Q-Q plots, residuals, trend), and computes performance metrics such as the coefficient of determination (R^2).

Note: The data cleaning, interpolation, Fourier modeling, and residual analysis were performed using the same method for all springs. For clarity and brevity, only the full Python script for the Cheserod spring temperature and water level analysis is included in this appendix.

```
1 !pip install pykalman
2 from scipy import stats
3 from scipy.optimize import curve_fit
4 from pykalman import KalmanFilter
5 from sklearn.metrics import r2_score
6 from google.colab import files
7 from scipy.stats import norm
8 from statsmodels.graphics.gofplots import qqplot
9 import matplotlib.pyplot as plt
10 import numpy as np
```

```

11
12 cheserod_temperature = ['/content/springs-data/cheserod/
    Cheserod_0002.txt']
13
14 def fourier_series_with_trend(t, *params):
15     n_terms = (len(params) - 2) // 3
16     result = np.zeros_like(t)
17
18     for i in range(n_terms):
19         A = params[i * 3]
20         B = params[i * 3 + 1]
21         C = params[i * 3 + 2]
22         result += A * np.sin(B * t + C)
23
24     # Add the linear trend (slope * t + intercept)
25     trend_slope = params[-2]
26     trend_intercept = params[-1]
27     result += trend_slope * t + trend_intercept # Trend: slope *
        t + intercept
28     return result
29
30 # Loop through each file and plot the data on separate subplots
31 for i, path in enumerate(cheserod_temperature):
32     # Read the data from the file
33     original_data = pd.read_csv(path, sep=';', names=['Date', '
        Time', 'Temperature'])
34
35     # Convert 'Date' and 'Time' columns to datetime and set as
        index
36     original_data['Datetime'] = pd.to_datetime(original_data['
        Date'] + '_' + original_data['Time'], format='%d/%m/%Y_%H:%
        M:%S')
37     original_data.set_index('Datetime', inplace=True)
38
39     # Convert 'Temperature' column to numeric, handling errors
40     original_data['Temperature'] = pd.to_numeric(original_data['
        Temperature'], errors='coerce')
41
42     # Drop rows with missing temperature values
43     df = original_data.dropna(subset=['Temperature'])
44
45     kf = KalmanFilter(
46         transition_matrices=[1], # No change in state
47         observation_matrices=[1], # Direct observation
48         transition_covariance=[[0.00000001]], # Small transition noise
49         observation_covariance=[[0.00001]] # Small observation noise
50     )
51     smoothed_data, _ = kf.filter(df['Temperature'].values)
52
53     # Store smoothed data back into the dataframe
54     df.loc[:, 'Temperature'] = smoothed_data

```

```

55
56 # Interpolation
57 df.loc[:, 'Temperature'] = df['Temperature'].interpolate(
    method='polynomial', order=2)
58
59 # Initial guesses for Fourier terms + trend (slope * t +
    intercept)
60 p0 = [
61     0.6,
62     2 * np.pi / (365*5),
63     0,
64     0.5,
65     2 * np.pi / (365*4),
66     0,
67     0.4,
68     2 * np.pi / (365*3),
69     0,
70     0.3,
71     2 * np.pi / (365*2),
72     0,
73     0.2, # A (first sine term)
74     2 * np.pi / 365, # B (first sine term frequency)
75     0, # C (first sine term phase shift)
76     0.1, # A (second sine term)
77     2 * np.pi / (365/2), # B (second sine term frequency)
78     0, # C (second sine term phase shift)
79     0, # Initial guess for slope (trend slope)
80     np.mean(df['Temperature']) # Initial guess for intercept (
        trend intercept)
81 ]
82
83 # Convert datetime index to numerical values (days since
    start) before curve_fit
84 df.loc[:, 'Time_Num'] = (df.index - df.index[0]).
    total_seconds() / (3600 * 24)
85 # Fit the model
86 params, _ = curve_fit(fourier_series_with_trend, df['Time_Num
    '], df['Temperature'], p0=p0, maxfev=10000)
87
88 # Calculate all_time based on original_data's index
89 all_time = (original_data.index - original_data.index[0]).
    total_seconds() / (3600 * 24) # Recalculate all_time
90
91 # Predict temperature using the fitted model (Fourier
    components + trend)
92 original_data['Predicted_Temperature'] =
    fourier_series_with_trend(all_time, *params)
93
94 # Reconstruct full signal
95 original_data['Final_Temperature'] = original_data['
    Temperature'].fillna(original_data['Predicted_Temperature']

```

```

    ))
96
97 # Only compare on valid (non-NaN) entries
98 mask = ~np.isnan(original_data['Temperature']) & ~np.isnan(
    original_data['Predicted_Temperature'])
99
100 # Compute R
101 r2 = r2_score(original_data.loc[mask, 'Temperature'],
    original_data.loc[mask, 'Predicted_Temperature'])
102 print(f"R    for {cheserod_temperature[0].split('/')[0]}: {r2:.4f
    }")
103
104 trend_slope = params[-2] # The second-to-last parameter is the
    slope (trend_slope)
105 change_per_year = trend_slope * 365 # Temperature change per
    year
106 print(f"Change in temperature per year for {path.split('/')[0]}:
    {change_per_year:.4f} C /year")
107
108 total_days = (original_data.index[-1] - original_data.index[0]).
    days
109 total_change = trend_slope * total_days # Total temperature
    change
110 print(f"Total change in temperature for {path.split('/')[0]}: {
    total_change:.4f} C over {total_days} days")
111
112 # Plot results
113 plt.figure(figsize=(10, 5))
114 plt.plot(df.index, df['Temperature'], 'rx', markersize= 1, label=
    'cleaned_data')
115 plt.plot(original_data.index, original_data['
    Predicted_Temperature'], color= 'blue', label='Fitted_Trend')
116 plt.xlabel('Year')
117 plt.ylabel('Temperature ( C )')
118 plt.legend()
119 plt.grid(True)
120 plt.title(f'Water Temperature Trend Analysis    Cheserod Spring'
    )
121 plt.savefig(f'/content/springs-data/Cheserod_temperature_plot.jpg
    ', dpi=300, bbox_inches='tight')
122 plt.show()
123 plt.close()
124
125 #Q-Q Plot
126 qqplot(df['Temperature'], line='s')
127 plt.title(' Q Q Plot of Cleaned Temperature Data - Cheserod
    Spring')
128 plt.grid(True)
129 save_path = '/content/springs-data/Cheserod_Q-Q_plot.jpg'
130 plt.savefig(save_path, dpi=300, bbox_inches='tight')
131 plt.show()

```

```

132
133 #Original vs Cleaned data
134 plt.figure(figsize=(10, 5))
135 plt.plot(original_data.index, original_data['Temperature'], color
        ='green', marker='o', linestyle='None', markersize=3, label='
        original_data')
136 plt.plot(df.index, df['Temperature'], color='red', marker='o',
        linestyle='None', markersize=1, label='cleaned_data')
137 plt.title('Original vs Cleaned Temperature Time Series
        Cheserod Spring')
138 plt.xlabel('Year')
139 plt.ylabel('Temperature ( C )')
140 plt.legend(loc='upper_left')
141 plt.grid()
142 save_path = '/content/springs-data/Cheserod_temperature_plot.jpg'
143 plt.savefig(save_path, dpi=300, bbox_inches='tight')
144 plt.show()
145
146 #Residuals
147 residuals = df['Temperature'] - fourier_series_with_trend(df['
        Time_Num'], *params)
148 plt.figure(figsize=(10, 3))
149 plt.plot(df.index, residuals, color='gray')
150 plt.axhline(0, color='red', linestyle='--')
151 plt.title('Residuals of the Fitted Temperature Model -
        Cheserod Spring')
152 plt.xlabel('Year')
153 plt.ylabel('Residual')
154 plt.grid(True)
155 save_path = '/content/springs-data/Cheserod_residuals_plot.jpg'
156 plt.savefig(save_path, dpi=300, bbox_inches='tight')
157 plt.show()
158
159 # Plot to visually confirm
160 df['Temperature'].isna().sum() # Count of missing values
161
162 plt.figure(figsize=(10, 5))
163 plt.plot(original_data.index, original_data['Temperature'], color
        ='green', markersize=1,)
164 plt.title('Visual Inspection of Missing Values in Temperature
        Time Series - Cheserod Spring')
165 plt.xlabel('Year')
166 plt.ylabel('Temperature ( C )')
167 plt.grid()
168 save_path = '/content/springs-data/
        Cheserod_missing_values_temperature.jpg'
169 plt.savefig(save_path, dpi=300, bbox_inches='tight')
170 plt.show()

```

B.3: Extracting Cleaned Data for Correlation Analysis

The following script extracts cleaned spring datasets from a compressed archive and prepares them for further analysis, such as correlation evaluation. It loads the necessary Python libraries and unzips the data files into the working directory.

```
1 import zipfile
2 import os
3 import glob
4 import pandas as pd
5 import numpy as np
6 import matplotlib.pyplot as plt
7 import seaborn as sns
8 from datetime import datetime
9
10 # Unzip the folder
11 zip_path = '/content/springs_cleaned_data.zip'
12 extract_path = '/content'
13
14 with zipfile.ZipFile(zip_path, 'r') as zip_ref:
15     zip_ref.extractall(extract_path)
```

B.4: Loading and Organizing Spring Data for Analysis

This script collects all cleaned variable data files (temperature, water level, conductivity, etc.) from the five springs and organizes them into a structured Python dictionary. The script handles different delimiters, parses datetime fields, and maps variable codes to readable names.

```
1 # Define paths to each spring
2 Cheserod = "/content/springs_cleaned_data/Cheserod"
3 Gabiet = "/content/springs_cleaned_data/Gabiet"
4 Entrebin = "/content/springs_cleaned_data/Entrebin"
5 Promiod = "/content/springs_cleaned_data/Promiod"
6 Promise = "/content/springs_cleaned_data/Promise"
7
8 # Collect all .txt files from each spring directory
9 variable_files = (
10     glob.glob(f"{Cheserod}/*.txt") +
11     glob.glob(f"{Gabiet}/*.txt") +
12     glob.glob(f"{Entrebin}/*.txt") +
13     glob.glob(f"{Promiod}/*.txt") +
14     glob.glob(f"{Promise}/*.txt")
15 )
16
17 # Mapping of variable codes to names
18 variable_map = {
19     '0001': 'water_level',
20     '0002': 'temperature',
21     '0004': 'conductivity',
22     '0005': 'flowrate',
23     '0006': 'precipitation',
```

```

24     '0007': 'air_temperature',
25     '0008': 'humidity'
26 }
27
28 def load_custom_txt(file_path, variable_name):
29     # Detect delimiter using the first line
30     with open(file_path, 'r', encoding='ISO-8859-1') as f:
31         first_line = f.readline()
32         sniffer = csv.Sniffer()
33         try:
34             dialect = sniffer.sniff(first_line)
35             delimiter = dialect.delimiter
36         except csv.Error:
37             delimiter = ';' # fallback if detection fails
38
39     # Read the file, skipping the first row (header or metadata)
40     df = pd.read_csv(file_path, sep=delimiter, skiprows=1,
41                     encoding='ISO-8859-1', header=None)
42
43     # If not enough columns, raise an error
44     if df.shape[1] < 4:
45         raise ValueError(f"File {file_path} has only {df.shape[1]} column(s), expected at least 4.")
46
47     # Assign column names
48     df.columns = ['datetime', 'date', 'time', 'variable']
49
50     # Convert 'datetime' to datetime object
51     df['datetime'] = pd.to_datetime(df['datetime'], errors='coerce', dayfirst=True)
52
53     return df
54
55 # Initialize dictionary with all springs
56 dataframes_dict = {
57     'Cheserod': {},
58     'Gabiet': {},
59     'Entrebin': {},
60     'Promiod': {},
61     'Promise': {}
62 }
63
64 # Process each file
65 for file in variable_files:
66     filename = os.path.basename(file)
67     variable_code = filename.split('_')[-1].split('.')[0]
68
69     if variable_code in variable_map:
70         variable_name = variable_map[variable_code]
71

```

```

72     # Identify the spring name from the file path
73     for spring in dataframes_dict.keys():
74         if spring in file:
75             source = spring
76             break
77     else:
78         print(f"Could not determine source for {file}")
79         continue
80
81     print(f"Reading {file} as {variable_name} from {source}"
82         )
83     df = load_custom_txt(file, variable_name)
84     dataframes_dict[source][variable_name] = df
85 else:
86     print(f"Skipping {file} as it does not match any known
87         variable codes.")

```

B.5: Cross-Correlation Analysis for Spring Variables

```

1  import pandas as pd
2  import matplotlib.pyplot as plt
3  import seaborn as sns
4  from google.colab import files
5
6  for spring_name, variable_dict in dataframes_dict.items():
7      print(f"\nProcessing {spring_name}")
8      merged_data = pd.DataFrame()
9
10     # Merge variables daily
11     for var_name, df in variable_dict.items():
12         df = df.dropna(subset=['datetime'])
13         df = df.set_index('datetime')
14         df_daily = df['variable'].resample('D').mean().rename(
15             var_name)
16
17         if merged_data.empty:
18             merged_data = df_daily.to_frame()
19         else:
20             merged_data = merged_data.join(df_daily, how='outer')
21
22     clean_data = merged_data.dropna()
23
24     if clean_data.shape[0] < 100:
25         print(f"Skipping {spring_name} Not enough data ({
26             clean_data.shape[0]} rows)")
27         continue
28
29     variable_names = clean_data.columns.tolist()
30     for i, var1 in enumerate(variable_names):
31         for j, var2 in enumerate(variable_names):
32             if i >= j:

```

```

31         continue # skip duplicates & self-correlation
32
33     series1 = clean_data[var1]
34     series2 = clean_data[var2]
35
36     lags, corr = cross_correlation_with_lags(series1,
37                                             series2, max_lag=MAX_LAG)
38
39     if np.all(np.isnan(corr)):
40         print(f"Skipping {var1} vs {var2} - insufficient data overlap")
41         continue
42
43     best_idx = np.nanargmax(np.abs(corr))
44     best_lag = lags[best_idx]
45     best_corr = corr[best_idx]
46
47     # Plotting
48     plt.figure(figsize=(10, 4))
49     plt.plot(lags, corr, label=f'{var1} vs {var2}')
50     plt.axvline(best_lag, color='red', linestyle='--',
51                label=f'Best Lag: {best_lag} days\nCorr: {best_corr:.2f}')
52     plt.title(f'Cross-Correlation in {spring_name}: {var1} vs {var2}')
53     plt.xlabel('Lag (days)')
54     plt.ylabel('Cross-Correlation')
55     plt.legend()
56     plt.tight_layout()
57     plt.show()

```

```

1 MAX_LAG = 30 # max lag in days
2
3 def cross_correlation_with_lags(series1, series2, max_lag=30):
4     lags = np.arange(-max_lag, max_lag + 1)
5     correlations = []
6
7     for lag in lags:
8         if lag < 0:
9             shifted_series1 = series1.shift(-lag)
10            shifted_series2 = series2
11        else:
12            shifted_series1 = series1
13            shifted_series2 = series2.shift(lag)
14
15        valid = shifted_series1.notna() & shifted_series2.notna()
16        if valid.sum() < 30: # require minimum overlap
17            correlations.append(np.nan)
18        else:
19            corr = shifted_series1[valid].corr(shifted_series2[valid])

```

```

20         correlations.append(corr)
21
22     return lags, correlations
23
24     OUTPUT_DIR = "cross_correlation_heatmaps"
25     os.makedirs(OUTPUT_DIR, exist_ok=True)
26
27     for spring_name, variable_dict in dataframes_dict.items():
28         print(f"\nProcessing_{spring_name}")
29
30         # Merge data for all variables
31         merged_data = pd.DataFrame()
32         for var_name, df in variable_dict.items():
33             df = df.dropna(subset=['datetime'])
34             df = df.set_index('datetime')
35             df_daily = df['variable'].resample('D').mean().rename(
36                 var_name)
37
38             if merged_data.empty:
39                 merged_data = df_daily.to_frame()
40             else:
41                 merged_data = merged_data.join(df_daily, how='outer')
42
43         clean_data = merged_data.dropna()
44
45         if clean_data.shape[0] < 100:
46             print(f"Skipping_{spring_name}_    _Not_enough_data_{(
47                 clean_data.shape[0]}_rows)")
48             continue
49
50         variable_names = clean_data.columns.tolist()
51         n_vars = len(variable_names)
52
53         # Prepare matrices to store best correlation and lag
54         best_corr_matrix = pd.DataFrame(np.nan, index=variable_names,
55             columns=variable_names)
56         best_lag_matrix = pd.DataFrame(np.nan, index=variable_names,
57             columns=variable_names)
58
59         for i, var1 in enumerate(variable_names):
60             for j, var2 in enumerate(variable_names):
61                 if i == j:
62                     best_corr_matrix.loc[var1, var2] = 1.0
63                     best_lag_matrix.loc[var1, var2] = 0
64                     continue
65
66                 series1 = clean_data[var1]
67                 series2 = clean_data[var2]
68
69                 lags, corr_values = cross_correlation_with_lags(
70                     series1, series2, max_lag=MAX_LAG)

```

```

66
67         if all(np.isnan(corr_values)):
68             continue
69
70         # Get index of max absolute correlation
71         best_idx = np.nanargmax(np.abs(corr_values))
72         best_corr = corr_values[best_idx]
73         best_lag = lags[best_idx]
74
75         best_corr_matrix.loc[var1, var2] = best_corr
76         best_lag_matrix.loc[var1, var2] = best_lag
77
78     # Plot correlation heatmap
79     plt.figure(figsize=(10, 8))
80     sns.heatmap(best_corr_matrix, annot=True, cmap='coolwarm',
81                 center=0, vmin=-1, vmax=1,
82                 linewidths=0.5, fmt=".2f")
83     plt.xticks(rotation=90)
84     plt.yticks(rotation=0)
85     plt.title(f'Cross-Correlation Matrix_{spring_name}')
86     plt.tight_layout()
87     plt.savefig(os.path.join(OUTPUT_DIR, f"{spring_name}
88         _best_correlation_heatmap.png"), dpi=300)
89     plt.show()
90     plt.close()

```

B.6: Cross-Correlation Analysis between Spring Variables

Note: The following procedure is identical for all springs. For illustration purposes, we show the full code for Cheserod-Gabiet.

```

1 import zipfile
2 import os
3 import glob
4 import pandas as pd
5 import numpy as np
6 import matplotlib.pyplot as plt
7 import seaborn as sns
8 from datetime import datetime
9
10 # Unzip the folder
11 zip_path = '/content/springs_cleaned_data.zip'
12 extract_path = '/content'
13
14 with zipfile.ZipFile(zip_path, 'r') as zip_ref:
15     zip_ref.extractall(extract_path)

```

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 from google.colab import files

```

```

5
6 for spring_name, variable_dict in dataframes_dict.items():
7     print(f"\nProcessing_{spring_name}")
8     merged_data = pd.DataFrame()
9
10    # Merge variables daily
11    for var_name, df in variable_dict.items():
12        df = df.dropna(subset=['datetime'])
13        df = df.set_index('datetime')
14        df_daily = df['variable'].resample('D').mean().rename(
            var_name)
15
16        if merged_data.empty:
17            merged_data = df_daily.to_frame()
18        else:
19            merged_data = merged_data.join(df_daily, how='outer')
20
21    clean_data = merged_data.dropna()
22
23    if clean_data.shape[0] < 100:
24        print(f"Skipping_{spring_name}_    _Not_enough_data_{(
            clean_data.shape[0]}_rows)")
25        continue
26
27    variable_names = clean_data.columns.tolist()
28    for i, var1 in enumerate(variable_names):
29        for j, var2 in enumerate(variable_names):
30            if i >= j:
31                continue # skip duplicates & self-correlation
32
33            series1 = clean_data[var1]
34            series2 = clean_data[var2]
35
36            lags, corr = cross_correlation_with_lags(series1,
                series2, max_lag=MAX_LAG)
37
38            if np.all(np.isnan(corr)):
39                print(f"Skipping_{var1}_vs_{var2}_    _
                    insufficient_data_overlap")
40                continue
41
42            best_idx = np.nanargmax(np.abs(corr))
43            best_lag = lags[best_idx]
44            best_corr = corr[best_idx]
45
46            # Plotting
47            plt.figure(figsize=(10, 4))
48            plt.plot(lags, corr, label=f'_{var1}_vs_{var2}_')
49            plt.axvline(best_lag, color='red', linestyle='--',
50                        label=f'Best_Lag:_{best_lag}_days\nCorr:_{
                            best_corr:.2f}_')

```

```

51         plt.title(f'Cross-Correlation in {spring_name}: {var1}
52                vs {var2}')
53         plt.xlabel('Lag (days)')
54         plt.ylabel('Cross-Correlation')
55         plt.legend()
56         plt.tight_layout()
57         plt.show()

```

```

1  import numpy as np
2  import pandas as pd
3  import matplotlib.pyplot as plt
4  import seaborn as sns
5  import os
6
7  MAX_LAG = 30 # max lag in days
8
9  def cross_correlation_with_lags(series1, series2, max_lag=30):
10     lags = np.arange(-max_lag, max_lag + 1)
11     correlations = []
12
13     for lag in lags:
14         if lag < 0:
15             shifted_series1 = series1.shift(-lag)
16             shifted_series2 = series2
17         else:
18             shifted_series1 = series1
19             shifted_series2 = series2.shift(lag)
20
21     valid = shifted_series1.notna() & shifted_series2.notna()
22     if valid.sum() < 30: # require minimum overlap
23         correlations.append(np.nan)
24     else:
25         corr = shifted_series1[valid].corr(shifted_series2[
26             valid])
27         correlations.append(corr)
28
29     return lags, correlations
30
31 OUTPUT_DIR = "cross_correlation_heatmaps"
32 os.makedirs(OUTPUT_DIR, exist_ok=True)
33
34 for spring_name, variable_dict in dataframes_dict.items():
35     print(f"\nProcessing {spring_name}")
36
37     # Merge data for all variables
38     merged_data = pd.DataFrame()
39     for var_name, df in variable_dict.items():
40         df = df.dropna(subset=['datetime'])
41         df = df.set_index('datetime')
42         df_daily = df['variable'].resample('D').mean().rename(
43             var_name)

```

```

42
43     if merged_data.empty:
44         merged_data = df_daily.to_frame()
45     else:
46         merged_data = merged_data.join(df_daily, how='outer')
47
48 clean_data = merged_data.dropna()
49
50 if clean_data.shape[0] < 100:
51     print(f"Skipping {spring_name} Not enough data ({
52         clean_data.shape[0]} rows)")
53     continue
54
55 variable_names = clean_data.columns.tolist()
56 n_vars = len(variable_names)
57
58 # Prepare matrices to store best correlation and lag
59 best_corr_matrix = pd.DataFrame(np.nan, index=variable_names,
60     columns=variable_names)
61 best_lag_matrix = pd.DataFrame(np.nan, index=variable_names,
62     columns=variable_names)
63
64 for i, var1 in enumerate(variable_names):
65     for j, var2 in enumerate(variable_names):
66         if i == j:
67             best_corr_matrix.loc[var1, var2] = 1.0
68             best_lag_matrix.loc[var1, var2] = 0
69             continue
70
71     series1 = clean_data[var1]
72     series2 = clean_data[var2]
73
74     lags, corr_values = cross_correlation_with_lags(
75         series1, series2, max_lag=MAX_LAG)
76
77     if all(np.isnan(corr_values)):
78         continue
79
80     # Get index of max absolute correlation
81     best_idx = np.nanargmax(np.abs(corr_values))
82     best_corr = corr_values[best_idx]
83     best_lag = lags[best_idx]
84
85     best_corr_matrix.loc[var1, var2] = best_corr
86     best_lag_matrix.loc[var1, var2] = best_lag
87
88 # Plot correlation heatmap
89 plt.figure(figsize=(10, 8))
90 sns.heatmap(best_corr_matrix, annot=True, cmap='coolwarm',
91     center=0, vmin=-1, vmax=1,
92     linewidths=0.5, fmt=".2f")

```

```
88 plt.xticks(rotation=90)
89 plt.yticks(rotation=0)
90 plt.title(f'Cross-Correlation_Matrix_{spring_name}')
91 plt.tight_layout()
92 plt.savefig(os.path.join(OUTPUT_DIR, f"{spring_name}
    _best_correlation_heatmap.png"), dpi=300)
93 plt.show()
94 plt.close()
```