



# Politecnico di Bari

Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Data-driven methods for qualitative and quantitative characterisation of coastal aquifers

This is a PhD Thesis

*Original Citation:*

Data-driven methods for qualitative and quantitative characterisation of coastal aquifers / Alfio, Maria Rosaria. -  
ELETTRONICO. - (2024). [10.60576/poliba/iris/alfio-maria-rosaria\_phd2024]

*Availability:*

This version is available at <http://hdl.handle.net/11589/264661> since: 2024-01-13

*Published version*

<http://hdl.handle.net/11589/264661>  
DOI: 10.60576/poliba/iris/alfio-maria-rosaria\_phd2024

*Terms of use:*

Altro tipo di accesso

(Article begins on next page)

28 April 2024



# D.R.S.A.T.E.

## POLITECNICO DI BARI

# 01

Doctor in Risk And Environmental, Territorial And Building Development

2024

Coordinator: Prof. Vito Iacobellis

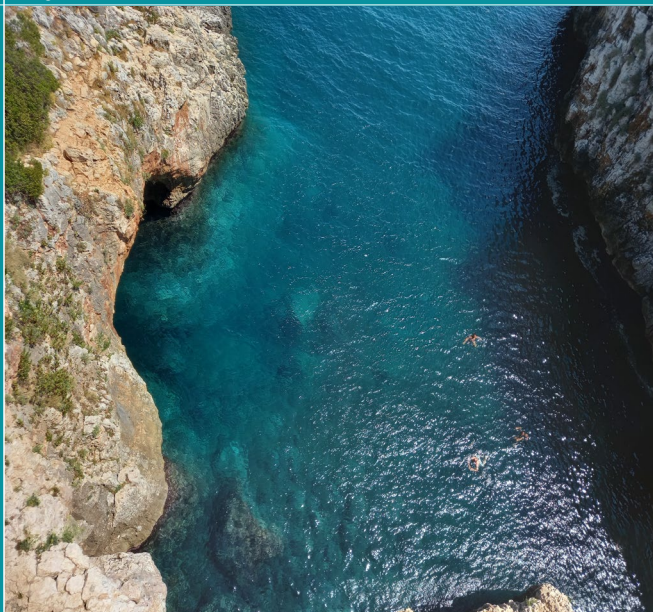
XXXVI CYCLE  
Curriculum: ICAR/02

DICATECh  
Department of Civil, Environmental,  
Building Engineering and Chemistry

Maria Rosaria Alfio

**Data-driven methods for qualitative and quantitative characterisation of coastal aquifers**

Prof. Gabriella Balacco  
Prof. Maria Dolores Fidelibus  
Department of Civil, Environmental, Building Engineering and Chemistry - Polytechnic University of Bari  
Dr. Andreas Panagopoulos  
Soil and Water Resources Institute (SWRI) - Hellenic Agricultural Organisation





La sottoscritta **ALFIO MARIA ROSARIA** nata a **GROTTAGLIE (TA)** il **20/03/1993**

residente a **GROTTAGLIE (TA)** in via **GIOVANNI XXIII, 16** e-mail **mariarosaria.alfio@poliba.it**

iscritto al 3° anno di Corso di Dottorato di Ricerca in **Rischio, Sviluppo Ambientale, Territoriale ed Edilizio** ciclo **XXXVI**

ed essendo stato ammesso a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

**Data-Driven methods for qualitative and quantitative characterisation of coastal aquifers**

#### DICHIARA

- 1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorreranno dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
- 2) di essere iscritto al Corso di Dottorato di ricerca **Rischio, Sviluppo Ambientale, Territoriale ed Edilizio** ciclo **XXXVI**, corso attivato ai sensi del "Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari", emanato con D.R. n.286 del 01.07.2013;
- 3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
- 4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archiverà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito <http://www.creativecommons.it/Licenze>), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
- 5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviata/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
- 6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
- 7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data Bari, 28/12/2023

Firma Maria Rosaria Alfio

La sottoscritta, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

#### CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data Bari, 28/12/2023

Firma Maria Rosaria Alfio



**D.R.S.A.T.E.**

**POLITECNICO DI BARI**

**01**

Doctor in Risk And Environmental, Territorial And Building Development

**2024**

Coordinator: Prof. Vito Iacobellis

XXXVI CYCLE  
Curriculum: ICAR/02

DICATECh  
Department of Civil, Environmental,  
Building Engineering and Chemistry

Maria Rosaria Alfio

**Data-driven methods for qualitative and quantitative characterisation of coastal aquifers**

Prof. Gabriella Balacco  
Department of Civil, Environmental, Building Engineering and Chemistry - Polytechnic University of Bari

Prof. Maria Dolores Fidelibus  
Department of Civil, Environmental, Building Engineering and Chemistry - Polytechnic University of Bari

Dr. Andreas Panagopoulos  
Soil and Water Resources Institute (SWRI) - Hellenic Agricultural Organisation





**D.R.S.A.T.E.**

**POLITECNICO DI BARI**

**01**

Dottorato in Rischio e Sviluppo Ambientale,  
Territoriale ed Edilizio

**2024**

Coordinatore: Prof. Vito Iacobellis

XXXVI CICLO  
Curriculum: ICAR/02

DICATECh  
Dipartimento di Ingegneria Civile,  
Ambientale, del Territorio, Edile e di  
Chimica

Maria Rosaria Alfio

**Metodi Data-driven per la caratterizzazione  
qualitativa e quantitativa degli acquiferi costieri**

Prof.ssa Gabriella Balacco  
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e  
di Chimica – Politecnico di Bari

Prof.ssa Maria Dolores Fidelibus  
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e  
di Chimica – Politecnico di Bari

Dr. Andreas Panagopoulos  
Soil and Water Resources Institute (SWRI) - Hellenic Agricultural  
Organisation



## ***EXTENDED ABSTRACT (eng)***

This research investigates a set of data-driven methods for characterising coastal aquifers under data scarcity using available short hydrogeological time series and chemical surveys. Correlation between groundwater levels and climate indexes and time series analyses, with particular attention to the autocorrelation and cross-correlation functions, wavelet analyses and seasonal and trend decomposition, are extensively analysed using precipitation and groundwater level recordings to define the hydrodynamic mechanism of an aquifer system in response to climate factors. These methods rely on the hypothesis that the aquifer system is considered as a filter that modifies, retains, or attenuates the input signal, i.e., precipitation, into an output signal, such as spring discharge, groundwater level, river flow rate or other physical or chemical parameters. They can provide valuable insights into various aspects, encompassing the nature of the aquifer, the influence of climatic conditions, and significant abstractions.

Multivariate statistical analysis is instead a valuable approach handling multiple geochemical and physical parameters to reveal spatial and temporal variations in groundwater quality, identify key hydrochemical processes, and assess how they change over time. Together with the Hydrogeochemical Facies Evolution-Diagram, these techniques allow to explore the salinisation process at the case study. Through Geostatistic, instead, the salinisation process and nitrate pollution are investigated in space and compared in time, allowing the identification of the areas more vulnerable.

A part of the Thesis is dedicated to a comprehensive assessment of climate change projections and bias-correction techniques, employing historical and regional climate data, to discuss the potential impacts of weather projections on a coastal aquifer.



The Thesis focuses on the complex coastal karst aquifer of Salento in Southern Italy, which presents numerous challenges, including geomorphological complexity, regional size, limited surface water resources, and significant water withdrawals for various human activities. The primary objective of this research is to investigate the hydrodynamic mechanism of such aquifer and discuss the potential problematics to which groundwater resources are exposed due to climate change and human pressure. The scope is to raise awareness among water utilities and political stakeholders in actuating mitigative actions and restrictions on the use of groundwater and exploring alternative measures to supply water demand. The study encourages for the establishment and consistent implementation of a comprehensive and strategic monitoring plan encompassing groundwater levels, water quality parameters, and other relevant variables aimed at ensuring the long-term sustainability and availability of groundwater resources for current and future generations.

### ***keywords***

Groundwater, Coastal aquifer, Data-scarcity, Time series analyses, Salinisation, Nitrate pollution, Climate change

## ***EXTENDED ABSTRACT (ita)***

Nella presente tesi è stata analizzata una serie di metodologie per caratterizzare gli acquiferi costieri in condizioni di scarsità di dati utilizzando brevi serie temporali idrogeologiche disponibili e rilevamenti chimici. Al fine di esaminare la dinamica delle acque sotterranee e la loro risposta alle precipitazioni, vengono ampiamente analizzate le correlazioni tra i livelli delle acque sotterranee e gli indici climatici, nonché le analisi delle serie temporali, con particolare attenzione alle funzioni di autocorrelazione e cross-correlazione, alle analisi *wavelet* e alla decomposizione stagionale e di tendenza, utilizzando registrazioni di precipitazioni e livelli. Questi metodi si basano sull'ipotesi che il sistema acquifero venga considerato come un filtro che modifica, trattiene o attenua il segnale in ingresso, cioè le precipitazioni, in un segnale in uscita, come la portata sorgentizia, il livello delle acque sotterranee, la portata dei fiumi o altri parametri fisici o chimici. Essi possono fornire preziose intuizioni su vari aspetti, compresi la natura dell'acquifero, l'influenza delle condizioni climatiche e i significativi prelievi.

L'analisi statistica multivariata rappresenta invece un approccio prezioso per gestire molteplici parametri geochimici e fisici al fine di rivelare variazioni spaziali e temporali nella qualità delle acque sotterranee, identificare processi idrochimici chiave e valutare come essi cambiano nel tempo. Insieme al Diagramma di Evoluzione delle *Facies* Idrogeochimiche, queste tecniche consentono di esplorare il processo di salinizzazione negli acquiferi costieri. Tramite l'impiego di tecniche geostatistiche, il processo di salinizzazione o di inquinamento (da nitrati per esempio) può essere analizzato spazialmente e comparato temporalmente con l'obiettivo di identificare le aree maggiormente vulnerabili.

Una parte significativa della tesi è dedicata all'esame delle proiezioni riguardanti i cambiamenti climatici e alle strategie di correzione dei potenziali bias nei modelli climatici, facendo ricorso alle serie storiche dei dati climatici e ai modelli climatici regionali al fine di discutere in che misura le previsioni meteorologiche potrebbero influenzare un acquifero costiero.

L'acquifero carsico costiero del Salento, ubicato nel Sud Italia, rappresenta il caso studio di tale ricerca in quanto presenta molteplici sfide, tra cui la complessità geomorfologica, le considerevoli dimensioni regionali, la limitata disponibilità di risorse idriche superficiali e il notevole utilizzo delle risorse idriche per svariate attività umane. Il principale obiettivo della presente ricerca è di comprendere il meccanismo idrodinamico di tale acquifero e di discutere le possibili problematiche a cui le risorse idriche sotterranee potrebbero essere esposte a causa dei cambiamenti climatici e delle attività antropiche. Si pone inoltre l'obiettivo di sensibilizzare le entità gestionali delle risorse idriche e gli attori politici a intraprendere azioni mitigative e restrizioni sull'uso delle acque sotterranee, esplorando altresì misure alternative per soddisfare la domanda idrica. Lo studio promuove l'istituzione e l'attuazione coerente di un piano di monitoraggio completo e strategico che includa la misura in continuo dei livelli dell'acqua sotterranea, dei parametri relativi alla qualità e di altre variabili rilevanti, al fine di garantire la sostenibilità e la disponibilità a lungo termine delle risorse idriche per le generazioni attuali e future.

### ***keywords***

Acque sotterranee, Acquiferi costieri, Data-scarcity, Analisi delle serie temporali, Salinizzazione, Inquinamento da nitrati, Cambiamenti climatici

# INDEX

INTRODUCTION.....	1
<b>1. THE SALENTO AQUIFER .....</b>	<b>7</b>
<b>1.1 General characteristics of the study area .....</b>	<b>7</b>
1.1.1 Geological setting of the study area .....	8
1.1.2 Geomorphology and surface hydrogeology of the study area.....	13
1.1.3 Hydrogeology of the study area .....	16
<b>1.2 Meteorological data .....</b>	<b>18</b>
1.2.1 Precipitation and temperature time series.....	18
<b>1.3 Hydrogeological data .....</b>	<b>19</b>
1.3.1 Groundwater levels.....	19
1.3.2 Chemical analyses for the salinisation process assessment .....	21
1.3.3 Chemical analyses for the nitrate pollution assessment .....	23
<b>2. MISSING VALUES FILLING-IN METHODS.....</b>	<b>25</b>
<b>2.1 Missing values in weather data.....</b>	<b>25</b>
<b>2.2 Filling-in methods and adopted metrics.....</b>	<b>29</b>
2.2.1 Mean Substitution .....	29
2.2.2 Thiessen polygons .....	30
2.2.3 Inverse Distance Weighting .....	30
2.2.4 Multiple Imputation by Chained Equations .....	30
<b>2.3 Metric coefficients.....</b>	<b>31</b>
2.3.1 Pearson's correlation coefficient.....	31
2.3.2 Nash-Sutcliffe Efficiency.....	32
2.3.3 Similarity Index.....	32
<b>3. DATA-DRIVEN METHODS FOR QUANTITATIVE CHARACTERISATION .....</b>	<b>33</b>
<b>3.1 Correlation indices .....</b>	<b>33</b>
3.1.1 Kendall's tau correlation coefficient .....	35
3.1.2 Spearman's rank-order correlation coefficient .....	36
3.1.3 Meteorological indices.....	36
<b>3.2 Seasonal and trend decomposition.....</b>	<b>39</b>

<b>3.3</b>	<b><i>Time series analyses</i></b> .....	<b>41</b>
3.3.1	Time Domain Analysis .....	41
3.3.2	Frequency Domain Analysis .....	43
<b>3.4</b>	<b><i>Wavelet Analysis</i></b> .....	<b>46</b>
<b>4.</b>	<b>DATA-DRIVEN METHODS FOR QUALITATIVE CHARACTERISATION</b> .....	<b>49</b>
<b>4.1</b>	<b><i>Multivariate statistical analysis</i></b> .....	<b>49</b>
4.1.1	Preliminary investigation .....	49
4.1.2	Q-mode Hierarchical Cluster Analysis.....	51
4.1.3	R-mode Factor Analysis.....	51
<b>4.2</b>	<b><i>Hydrogeochemical Facies Evolution-Diagram</i></b> .....	<b>53</b>
<b>4.3</b>	<b><i>Geostatistic</i></b> .....	<b>55</b>
4.3.1	Ordinary Kriging.....	57
4.3.2	Indicator Kriging .....	59
<b>5.</b>	<b>CLIMATE CHANGE MODELS</b> .....	<b>61</b>
<b>5.1</b>	<b><i>Climate Models and uncertainty related to BC methods</i></b> .....	<b>61</b>
5.1.1	Climate projections .....	63
5.1.2	Historical dataset .....	65
<b>5.2</b>	<b><i>Bias-correction methods</i></b> .....	<b>66</b>
5.2.1	Linear Scaling (LS) .....	67
5.2.2	Local Intensity scaling of precipitation (LOCI) .....	67
5.2.3	Power Transformation of Precipitation (PTR).....	68
5.2.4	Variance scaling of temperature (VAR) .....	68
5.2.5	Empirical Quantile Mapping (EQM) .....	68
5.2.6	Parametric Quantile Mapping (PQM).....	69
5.2.7	Generalised Quantile Mapping (GPQM).....	69
5.2.8	Detrended Quantile Matching (DQM) .....	70
5.2.9	Quantile Delta Mapping (QDM) .....	70
<b>5.3</b>	<b><i>BC Performance evaluation</i></b> .....	<b>71</b>
<b>6.</b>	<b>RESULTS AND DISCUSSION</b> .....	<b>73</b>
<b>6.1</b>	<b><i>Missing values production</i></b> .....	<b>73</b>
<b>6.2</b>	<b><i>Correlation between GWL and meteorological indicators</i></b> .....	<b>77</b>

<b>6.3</b>	<b><i>Analyses of GWL and precipitation in time and frequency domain</i></b> .....	<b>89</b>
<b>6.4</b>	<b><i>GWL and precipitation decomposition</i></b> .....	<b>98</b>
<b>6.5</b>	<b><i>Wavelet results</i></b> .....	<b>101</b>
<b>6.6</b>	<b><i>MVSA and HFE comparison</i></b> .....	<b>107</b>
6.6.1	Tiziano Project database.....	107
6.6.2	Maggiore Project database .....	119
<b>6.7</b>	<b><i>Indicator Kriging for the nitrate pollution assessment</i></b> .....	<b>126</b>
<b>6.8</b>	<b><i>Potential climate change in the Salento aquifer</i></b> .....	<b>137</b>
6.8.1	Results of bias correction methods.....	137
6.8.2	Future trends and potential climate change over the Salento aquifer...	147
<b>CONCLUSIONS</b> .....		<b>163</b>
<b>BIBLIOGRAPHY</b> .....		<b>168</b>
<b>List of figures</b> .....		<b>187</b>
<b>List of tables</b> .....		<b>191</b>
<b>Annex</b> .....		<b>193</b>
<b>CURRICULUM VITAE</b> .....		<b>200</b>



## INTRODUCTION

Freshwater is a vital resource for sustaining life and economic development. However, it is scarce on Earth, being less than three per cent, while the remaining 97 per cent is saline (National Geographic, 2023). Furthermore, only a limited portion of this freshwater is easily accessible, as a significant part is either trapped in glaciers or located beneath the Earth's surface as groundwater. In this context, groundwater is of utmost importance for human welfare and development, constituting approximately 97 per cent of the world's freshwater resources and serving as the primary source of water for drinking, irrigation, and industrial purposes (Ravenscroft & Lytton, 2022), especially in arid and semi-arid regions. It also plays an essential role in the hydrological cycle and is indispensable for the sustainability of various dependent aquatic and terrestrial ecosystems.

In many parts of the world, where the combination of geology, topography, and climate creates conditions that render the use of surface water resources unfeasible (technically and/or financially), groundwater is often the only option to cover the water demand. However, groundwater extraction may create numerous adverse effects, such as water table decline, land subsidence following the dewatering of phreatic aquifer, and water pollution. Climate change and the growing global population contribute to groundwater quality and quantity degradation, leading to potentially irreversible impacts (UNESCO, 2022). Groundwater overexploitation has become more severe in many areas since the beginning of the 21st century, resulting in a substantial reduction of groundwater reserves (Caretta et al., 2022). Extensive research conducted by Bagheri-Gavkosh et al. (2021) involving 290 case studies across 41 countries demonstrated that groundwater extraction contributes to land subsidence by approximately 59.75%. Furthermore, groundwater worldwide contains chemical species exceeding the World Health Organization (2012) quality standards for drinking and irrigation water limits (Algieri et al., 2022), with nitrate pollution being the most prevalent form of groundwater contamination globally (Arumi et al., 2009).



Hence, globally, groundwater resources present significant issues regarding their quality and quantity status that may result from geogenic factors, but most of the time are the impacts of anthropogenic activities often exacerbated by climate change effects. These impacts are even more severe and challenging in coastal aquifers where freshwater floats on saltwater due to their different fluid density. Under a decrease in water levels, the transition zone between fresh and saltwater expands, thus reducing the thickness of freshwater with a concurrent increase in salt content. Depending on the aquifer scale, groundwater recovery can take a very long time, and its quality may remain compromised considering the permanence of human pressures (Alfio et al., 2020). A slight amount of saltwater is sufficient to contaminate a fresh groundwater supply. Only a 3%–4% addition of salinity can make a freshwater supply unsuitable for drinking water and even irrigation (Morris et al. 2003). Besides seawater intrusion, groundwater salinisation may also locally or regionally occur because of geogenic sources or anthropogenic factors.

The delicate balance and susceptibility to human activities and environmental changes in coastal aquifers are crucial to control and mitigate the risk of saltwater contamination of water resources (Baena-Ruiz et al., 2020). The combination of global warming and the population increase will exacerbate groundwater issues in many regions across the Mediterranean. These areas are expected to experience higher temperatures and reduced precipitation patterns, leading to increased evapotranspiration rates, more frequent and severe droughts, and increased risk of flooding events (IPCC, 2014; Giorgi & Lionello, 2008). Sordo-Ward et al. (2019) have indicated that water scarcity issues are poised to intensify in several southern European basins, including countries such as Portugal, Spain, France, Switzerland, Italy, Greece, Northern Macedonia, Bulgaria, and Turkey. Pardo-Iguzquiza et al. (2019) have investigated the influence of climate change on the recharge of a karst aquifer in the Sierra de las Nieves (southern Spain), estimating a potential reduction on average recharge by approximately 53% under the extreme RCP8.5 scenario in 2071-2100. Mediterranean climatic conditions and landscape create favourable agricultural production and tourism conditions, which thrive. However, the region also faces numerous environmental and socio-economic

challenges relating to surface water scarcity, groundwater pollution due to human activities, including salinisation, whose effects on groundwater status should worsen under climate change (Aureli et al., 2008; Nikolaidis et al., 2014).

The enormous societal and ecological importance of coastal aquifers, along with the impactful effects of climate change, should direct political efforts towards taking serious measures for monitoring and managing groundwater resources. Their associated risk is high, as the three factors generally considered for risk assessment are simultaneously significant: (i) Hazard is high because it refers to human pressures and climate change, which are both continuously increasing; (ii) Vulnerability is also severe, considering that the quantitative and qualitative depletion of groundwater is by now in constant progress in various regions worldwide; (iii) Exposed Value, associated with the economic, social, and environmental damage, is also significant in the highly urbanised coastal regions.

Identifying the effective groundwater monitoring and management strategy requires a deep understanding of the system and its cause-effect relationships with environmental and human pressures (Jiao & Post, 2019). Management practices in coastal aquifers strive to reduce the overexploitation of freshwater to prevent seawater intrusion and mitigate the impacts on groundwater-dependent ecosystems (Werner et al., 2011). Knowledge that leads to setting up appropriate management practices, requires systematic measurements of parameters related to the relevant hydrological processes, which often lack or are spatially (and temporally) inadequate. The Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) argued that the lack of long observational time series prevents evaluating climate change impacts on groundwater quantity and quality (Jiménez Cisneros et al., 2014). Despite the significant advancement since AR5, nowadays, AR6 still claims that the spatial and temporal coverage of groundwater monitoring networks, pumping rates, and recharge data constrain a practical impact assessment on groundwater (Douville et al., 2021). While hydrological time series (i.e., precipitation and temperature data) are available for extended periods, hydrogeological datasets like groundwater levels (GWs) or chemical analyses often lack or present fragmented recordings. Under data scarcity, setting up

a comprehensive numerical model of a groundwater system (i.e., groundwater flow, pollution transport, salinisation transition zone propagation) can be challenging, especially if the system dynamic is poorly known. Nonetheless, data scarcity often affects regions where groundwater is crucial in sustaining ecosystems and human activities and is particularly susceptible to hydro-climatic hazards and human-induced pressures. Consequently, exploring alternative approaches based on available information becomes imperative and vital.

To ensure sustainable groundwater resources management and develop appropriate adaptation and mitigation strategies, it is firstly essential to properly understand the response of groundwater flow systems to meteorological and anthropogenic pressures. This Thesis explores a series of methodologies for the quantitative and qualitative characterisation of coastal aquifers, aiming at providing a suitable toolbox for assisting policymakers and water utilities in efficiently monitoring and managing groundwater resources under data scarcity. This integrated approach has been developed and tested on the coastal karst aquifer of Salento (Apulia, Southern Italy), which can be considered a benchmark for all discussed topics related to coastal aquifers, human pressures, and climate change impacts. Salento currently relies on groundwater for drinking and irrigation purposes due to a karst geological structure that prevents the existence of surface water bodies. Its complex geological and hydrogeological characteristics and the limited availability of hydrogeological data hinder a deep understanding of the recharge process, more so the development of a comprehensive numerical model. Moreover, its geographic location and geomorphological features make it sensitive to climatic hazards exacerbated by increasingly intense and frequent drought events and rising temperatures.

The Thesis is organised as in the following:

### *1. Hydrogeological characterisation*

Firstly, with reference to the Salento study area, the direct correlations between meteorological indices based on precipitation and temperature time series and available GWL recordings were established to investigate the aquifer response (Balacco et al., 2022a). To account for the propagation time from climate stimuli to the response of the

aquifer system, D’Oria et al. (2023) considered different cumulated periods to calculate meteorological indices and temporal shifts (lags) between time series. Despite the complexity of the study area, a potential linear relationship and locally variable lags between climate indices and GWLs were found for the Salento aquifer, as also demonstrated worldwide by several studies (Bloomfield & Marchant, 2013; Secci et al., 2021; Babre et al., 2022; Ndehedehe et al., 2023; Aydin-Kandemir & Erilat, 2023).

Considering these findings, Time Series (Balacco et al., 2022b) and Wavelet Analyses confirmed these preliminary results. These methods rely on the principle of representing an aquifer system as a filter where precipitation (input signal) infiltrates through the soil and percolates to the saturated zone of the aquifer, generating a change in the groundwater level (output signal) with a site-specific lag that can be interpreted as a transfer function (Delbart et al., 2014). However, groundwater fluctuations depend on multiple drivers (Shapoori et al., 2015). Time series and wavelet analysis can lead to inaccurate interpretations because additional seasonal or inter-annual, long, and random components influence the same fluctuations. Lafare et al. (2016) suggested combining time series decomposition techniques to avoid erroneous assessments of the behaviour of the aquifer and its characteristics. In the scientific literature, Seasonal and Trend decomposition using Loess (STL) technique was widely applied for different goals, including to account for the meteorological and geological heterogeneity impacts (Colyer et al., 2022), to detect GWL and salinity spatio-temporal dynamic patterns (Fan et al., 2023), to correlate GWL with rainfall and abstractions (Niranjannaik et al., 2022), and to investigate memory effect and response times between superficial and deep aquifers (Duy et al., 2021). Results of these alternative approaches suggest their effectiveness and general applicability to other regions where at least groundwater level, precipitation, and temperature time series are available. They are effective tools for hydrogeological characterisation and references for groundwater management.

## *2. Qualitative assessment*

Regarding the qualitative characterisation of the study area, Parisi et al. (2023) selected a few statistical and hydrogeochemical methodologies like Multivariate Statistical Analyses (i.e., Hierarchical Cluster Analysis and Factor Analysis) and Hydrochemical Facies

Evolution-Diagram (HFE-D) requiring only discrete measurements in time and space. Referring to the study area, MVSA and HFE-D led to outline the study of the spatio-temporal evolution of the salinisation process through chemical analyses belonging to two Regional Monitoring Projects, *Tiziano* (2007-2011) and *Maggiore* (2015-2018). While HFE exclusively focuses on groundwater salinisation processes, MVSA can generally outline the significant processes occurring in an aquifer and identify the most affected areas. Balacco et al. (2023) implemented the Indicator Kriging (IK) method to evaluate the spatio-temporal evolution of nitrate pollution and detect the more vulnerable areas.

### 3. *Future climatic projections*

After completing the quantitative and qualitative characterisation of the study area, projected precipitation and temperature have been used to assess the potential climate change for the near to medium-term (2031-2060) and distant (2071-2100) future (Alfio et al., 2023). To this aim, an ensemble of twelve Regional Climate Models (RCMs) driven by several General Circulation Models (GCMs) was initially adapted to the local climatic variability using eight bias-correction (BC) methods at daily time steps. A performance analysis was then carried out, comparing the bias-corrected simulated data with observations during the reference period (1971-2005) for each BC method. Subsequently, future patterns were estimated based on the most effective BC precipitation and temperature data methods. Results showed a general decrease in rainfall of about 6% and an increase in temperature of 2°C at the end of this century, compared to the historical period (1971-2005).

This research aims to develop an understanding of the evolution of the coastal aquifer to increase institutional and societal awareness of the risk to which water resources are being exposed, and design a feasible groundwater management tool of general interest. It is worth noting that the implemented approaches are of wide-ranging applicability and suitable under data scarcity as they successfully worked in a high geomorphological and structural complexity, regional size, surface water scarcity, and significant but unknown water withdrawals for human activities.

# 1. THE SALENTO AQUIFER

## 1.1 *General characteristics of the study area*

Salento aquifer locates in the homonymous Peninsula covering the southern part of the Apulia Region in Southern Italy. The Adriatic Sea limits the peninsula to the northeast, the Ionian Sea to the southwest, and the Murgia territory to the northwest. Its hydrogeological limits roughly coincide with the administrative borders of Lecce province, covering 2,799 km<sup>2</sup> with a coastline of 281 km (Fig. 1). This territory is mainly flat, with the highest altitude of around 200 meters AMSL. The climate is Mediterranean, characterised by mild and moderately rainy winters and hot and dry summers. Due to its flat geography and proximity to the sea, it is often quite windy. Temperatures in Salento are significantly influenced by the mitigating presence of the Ionian and Adriatic Seas, whose winds blowing on both sides establish distinct weather conditions. As already mentioned, winters are generally mild, although occasional cold periods can occur when winds blow in from the nearby Balkan Peninsula. Conversely, summers are typically pleasant because of a refreshing breeze from the north. However, there are times when the Sirocco wind blows in from Africa, resulting in scorching days with temperatures surpassing 40 °C.

Starting from the 1960s, the population began growing with a concurrent economic development mainly based on tourism, agriculture, and small family-size manufacturing activities. Data from Corine Land Cover 2018 (<https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download>) show that agricultural activities occupy 82.81% of the total area of the province. However, in recent years, the area dedicated to agriculture, as well as that occupied by permanent meadows, pastures, and woods, has experienced a gradual decrease due to the economic transformation and urban development of the entire province.

Significant and enduring characteristics of Salento territory are the scarcity of surface water and hydraulic disorders, and persistent drought periods, which have occurred over time due to human practices and implemented policies. These environmental

challenges have influenced the type of settlement, agricultural practices, the overall landscape, urban development, and the living conditions of the local population for centuries.

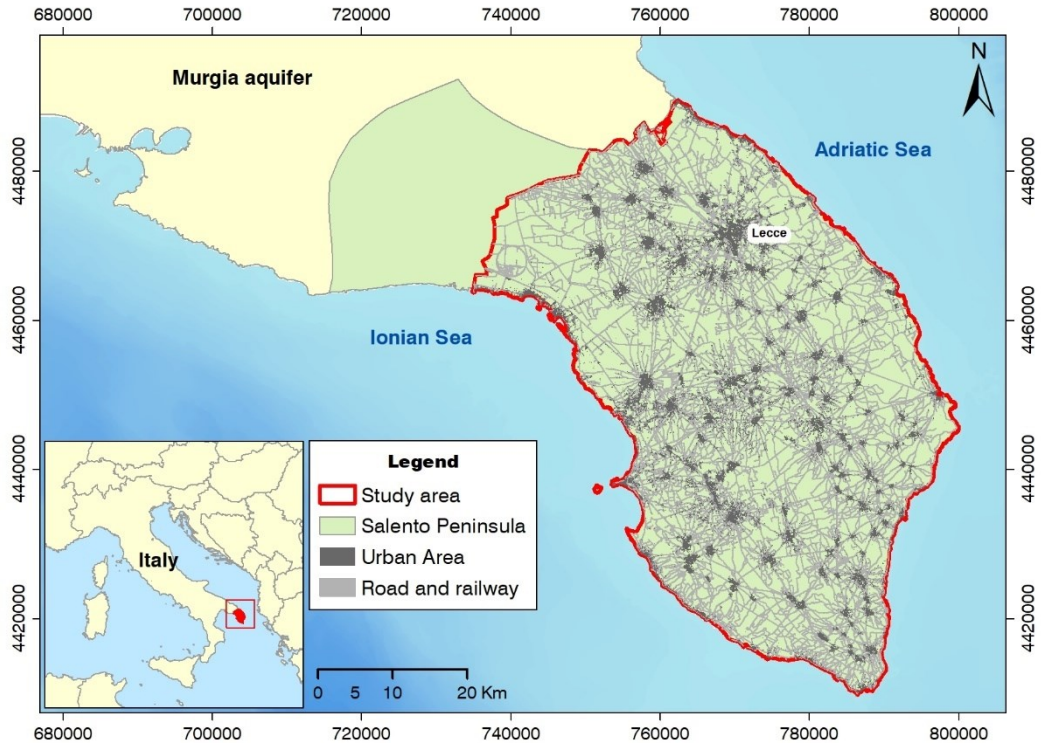


Fig. 1 - Location of the Salento Peninsula and delimitation of the study area.

### 1.1.1 Geological setting of the study area

The Salento peninsula is part of the Apulian carbonate platform, consisting of several litho-stratigraphic units of carbonate rocks from the Jurassic and Cretaceous periods, spatially distributed in variable order and thickness, ranging from approximately 3 to 5 kilometres (De Filippis et al., 2019). The geological basement comprises Cretaceous limestone and dolomitic limestone, which prominently outcrop in extensive areas. The covers include Miocene to Pleistocene silty or sandy clay, sand, and calcarenite (Ciaranfi et al., 1988). Fig. 2 shows the Reference Geological Map (RCM) for the Salento area provided by recent studies conducted during the MEDSAL project

([www.medsal.net](http://www.medsal.net)). Table 1 synthesizes the features of the Local Geological Formations (LGF), which description is detailed as follows:

- SAL\_GF1: a well-stratified succession of irregular and discontinuous alternation of greyish-white limestones, dolomite limestones, and grey-black dolomites, with subordinate intercalations of large layers full of Rudist fossils constitutes the unit. The upper part consists of layers and banks of fine-grained limestone (pack-stone/grainstone) white or light havana coloured, usually characterised by the constant presence of numerous colonial types coral clusters. Usually, this layer has a cavernous appearance, with cavities filled with lithified glauconitic mud with a laminar structure.
- SAL\_GF2: this unit is constituted by a well-stratified succession of micritic limestones and calcirudite stones, and locally by macrofossiliferous oolitic biosparites with variable diagenesis, white or light grey, subordinately havana or purple. The macrofossiliferous content consists of marine shellfish of medium and small size belonging mainly to lamellibranchs (*Cardium*, *Tapes*) and gastropods (*Cerithium*), with shells of brachiopods, echinoid skeletons, bryozoans, and annelids of internal and shallow neritic environment.
- SAL\_GF3: this unit is composed by yellow limestones, sometimes greenish due to the presence of Glauconite, which is either massive or with signs of bank stratification, rich in pelagic foraminifera (*Globigerina* and *Orbulina*). Biomicrites with medium and fine grain, semi-diagenized or incoherent, with thin, irregular, and discontinuous stratiform calcarenite intercalations represent the dominant lithofacies.
- SAL\_GF4: massive compact grey-blue marly-silty clay, somewhere rich in macro- and microfossils constitute this unit. The sedimentation environment corresponds to a deep neritic seabed.
- SAL\_GF5: this unit is composed of a set of clastic covers, referring to different and short sedimentary phases consequent to the repeated variations in sea level since the middle Pleistocene, owing to discontinuous and differentiated regional uplift associated with the contemporary glacial eustasy. The lithofacies (mainly calcarenitic) are constituted by terrigenous components of the coastal environment (from



subtidal-intertidal to emerged beach) of a limited thickness (max 30 meters) with sub-horizontal orientation. They lay on abrasion platforms dislocated towards the coast at progressively decreasing elevations between 100 m and 1 m AMSL, often with the interposition of continental deposits.

- SAL\_GF6: the dominant lithotype of the unit is composed of yellowish-grey calcarenites with coarse to fine grain size, scarcely macrofossiliferous, arranged in discontinuous layers with a centimetric and planar laminar structure, generally parallel to the stratification. On the top part, there are dune deposits constituted by incoherent yellowish-grey sands. This unit includes terrigenous deposits resulting from both chemical (karstic processes) and physical (erosive processes due to water flow processes) meteoric degradation, progressively accumulated in endoreic and exoreic depressions of various amplitude and shape. Deposits of a more detrital composition from medium to coarse grain size cover the bottom of episodically active endorheic basins and temporary water stream valleys engraved on the planes constituted by Plio-Pleistocene deposits.

Table 1 - Geology of the Salento aquifer (source: MEDSAL Project, report D.2.3)

<b>MEDSAL Code</b>	<b>LGF</b>	<b>Age</b>	<b>Description</b>
SAL_GF1	Limestone	Upper Cretaceous - Palaeocene	Layers and banks of variously fractured and karstified limestones
SAL_GF2	Calcarenites	Miocene - Lower Pliocene	Greyish yellow and organogeneous marly calcarenites
SAL_GF3	Calcarenites	Upper Pliocene - Lower Pleistocene	Calcareous sands from weakly to tightly cemented
SAL_GF4	Clays	Upper Pliocene - Lower Pleistocene	Silty or sandy clays
SAL_GF5	Calcarenites	Upper Pleistocene - Holocene	Soft and weakly cemented calcarenites
SAL_GF6	Sands and Clays	Holocene	Sands, sandy clays and red silts

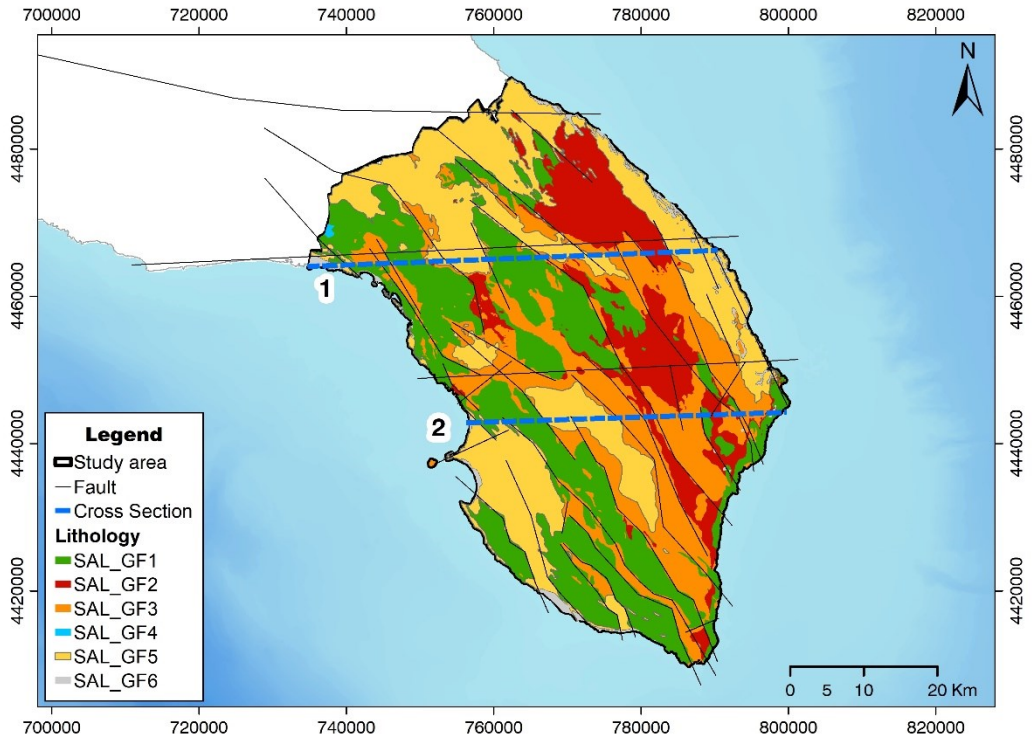


Fig. 2 - Reference Geological Map of Salento aquifer with traces of geological sections. The legend details are in Table 1.

The individual geological formations found in the stratigraphic columns of a few wells located in the Salento area were grouped into the following formational units considering the paleogeographical-geodynamic aspects (Fig. 3):

- Apulian Platform and Units of the Apulia Foreland (SAL\_GF1 and SAL\_GF2);
- Apennine-Dinaric Foredeep (SAL\_GF3 and SAL\_GF4);
- Marine Terrace Deposits (SAL\_GF5 and SAL\_GF6).

The first group is constituted by the sedimentary units that document the evolution of the Apulian Platform and the formations connected to the main phases of the Apennine-Dinaric tectogenesis. The second group includes the terrigenous units referable to the late tectonic phases of Upper Pliocene and Lower Pleistocene, responsible for the fore-deep evolution and filling. The last groups refer to the sedimentary units generated by the repeated changes in sea level, which were caused by the polyphasic uplift of the entire system constituted by the Apennine belt, the Foredeep, and the Foreland, as well

as by the concomitant glacial eustasy, started in the middle Pleistocene and still in progress.

Two geological-cross sections were considered. The geological-cross section no. 1 shows to the West the most recent lithotype units (SAL\_GF5 and SAL\_GF6), which are directly in transgression on the cretaceous carbonate basement (SAL\_GF1). In the central part, Pleistocene clays of high thickness fill a tectonic depression overlaying the Pliocene facies (SAL\_GF3). A horst and graben structure lowers the carbonate basement top towards the Adriatic Sea eastwards. All terms of the Salento stratigraphy appear in this part of the geological-cross section. The geological-cross section no. 2, located further south of cross section no. 1, shows in the central part a vast tectonic depression. Within this tectonic depression, the geological layers that constitute the complete Salento stratigraphy exhibit their maximum thicknesses.

The basement exhibits structural variations, with Horst and Graben separated by EW dextral and sinistral strike-slip faults and NW-SE sub-vertical normal faults (Gambini & Tozzi, 1996). The fault system, represented in Fig. 2, causes compartmentalisation of the aquifer at the regional scale, leading groundwater to flow by exhibiting conduit, conduit-barrier, or barrier behaviour from place to place (Fidelibus & Pulido-Bosch, 2019).

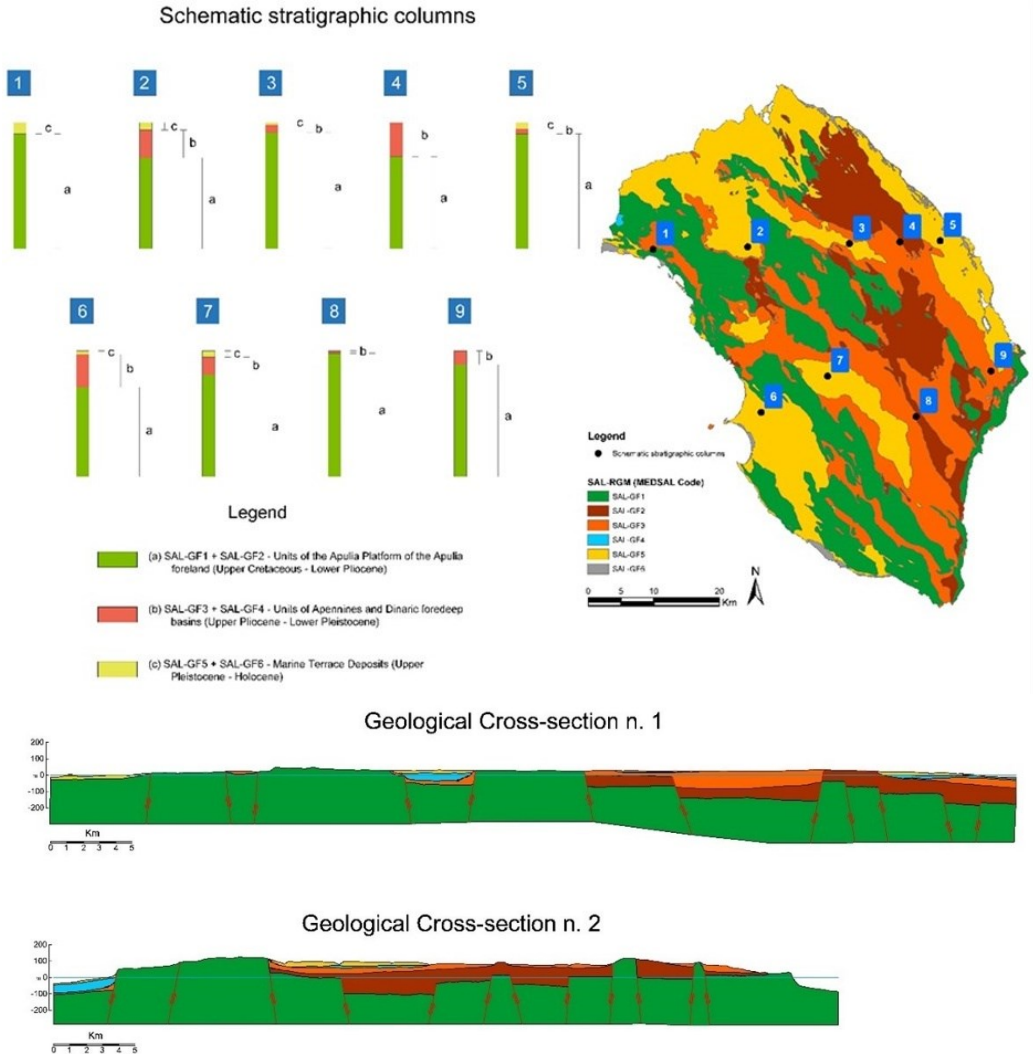


Fig. 3 - Schematic stratigraphic column and representative geological cross-sections (traces in Fig. 2).

### 1.1.2 Geomorphology and surface hydrogeology of the study area

The geomorphology of the Salento area is characterised by several features that reflect, in the landscape forms, the lithological matrix of the system and its tectonic-structural structure, as well as the cyclicity and spatial overlapping of the action of morphodynamic agents (Fig. 4). Even though the climatic conditions of the area have allowed the modelling action of both physical and chemical agents, the areal distribution

and number of the morphotypes currently recognisable in the territory suggest the prevalence and/or concomitance on certain areas and in particular geological periods of one or the other agent. The scarce relief energy characterising nearly the entire study area (except its south-west portion where the tectonic reliefs (Horst) reach maximum altitudes around 200 m AMSL) and the absence of extensive hydrographic networks, mean that the morphotypes of the physical gravity or erosive processes attributable to water action concentrate almost exclusively along with the coastal areas. Along the Adriatic coastal strip, mainly made up of high and rocky coasts, morphologies are frequent due to gravitational collapse phenomena (Bruno & Zezza, 1992). The Ionian coast is characterised by an alternating succession of low rocky cliffs with small islands, and rocks lined up a short distance from the shore, interspersed with extensive sandy shores bordered by dune belts with reclaimed marshes behind. The most characteristic morphotypes of rocky coastlines are depressions generated by the gravity collapse (sinkhole) of coastal cavities of tectonic-karst origin. Marine terraces and/or the paleo coastlines are other morphologies well represented in the Salento territory, both on the Adriatic and Ionian sides. These are determined by the erosive action of wave motion, give evidence not only of the cyclical glacio-eustatic oscillations especially during the Quaternary period but also of the uplifts and downturns undergone by the earth's crust as a result of tectonic deformation.

About the carbonated nature of the rocks existent in the territory, both outcropping and covered, the morphotypes due to the morphodynamic action of chemical agents (commonly known as karstification) are widespread. The Salento area displays a variety of distinctive karst features, including karst plains, fracture zones, dolines, sinkholes, and sub-horizontal karst levels, forming a system of discontinuities that significantly influence the hydraulic conductivity of the area. Karstic caves and submarine springs represent the subsurface karst morphologies and are almost exclusively located along the coast, close to or below sea level. The presence of this interconnected karst system results in a high degree of anisotropy and regional-scale permeability.

Ephemeral stream networks, usually dried, drain the karst surface during heavy and intense precipitation, causing flash floods. They drain exoreic basins outflowing along

the Adriatic and Ionian Sea coasts. On the contrary, hundreds of endorheic basins, which cover more than 40% of the study area, contribute to the aquifer recharge and cause-focused pollution transport processes. The behaviour of the endorheic basins in the Salento area depends on the magnitude and type of rainfall events. If the volume of internal runoff from precipitation exceeds the capacity of the endorheic basins, excess water flows out, feeding the network of exoreic basins. The intricate recharge and mass transport processes of the region reflect this complex water movement and redistribution behaviour. However, understanding of these processes is still limited and not fully understood, leaving some aspects of the hydrologic dynamics in the Salento area unclear.

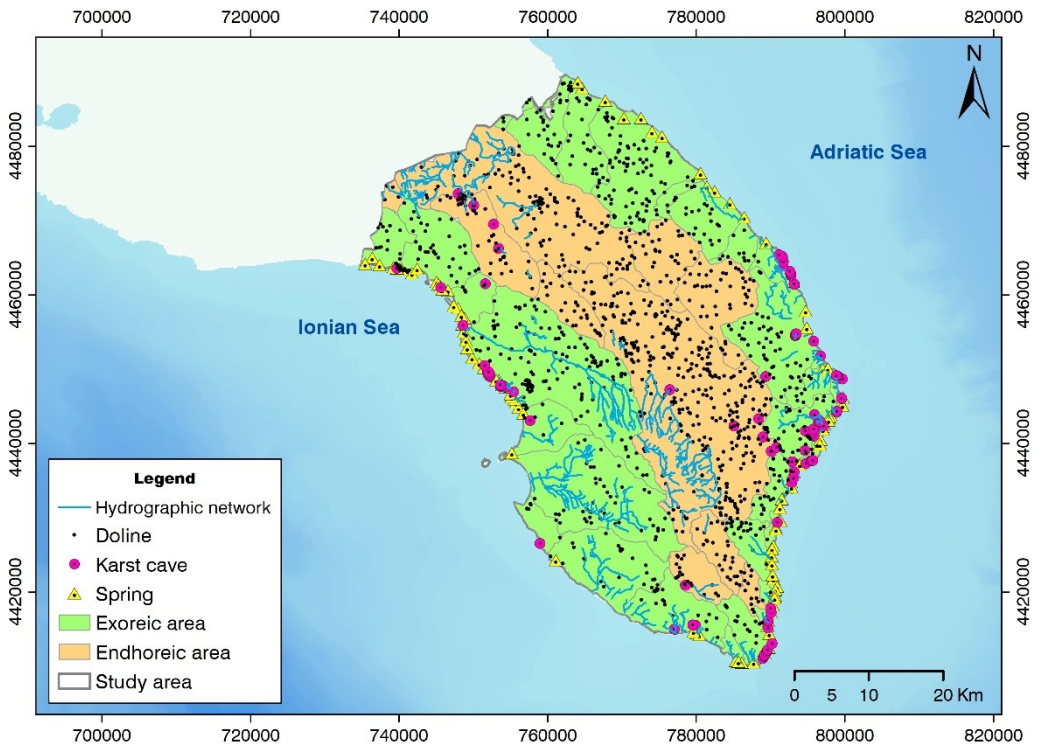


Fig. 4 - Geomorphological features of the Salento study area.

### 1.1.3 Hydrogeology of the study area

The primary groundwater body of the Salento study area is represented by the karst aquifer (also named “deep” aquifer) hosted in the Mesozoic carbonate rocks and bordered by the sea. It represents the main water resource for the Salento territory due to the karst nature of this environment that prevents the development of consistent surface water bodies. Groundwater also circulates in the Miocene and Plio-Pleistocene deposits, constituting modest shallow aquifers (Cotecchia, 1977).

The deep aquifer is a coastal aquifer, and the groundwater flow system is regional. Groundwater flows in predominantly phreatic conditions under hydraulic heads of the order of 3 m AMSL and hydraulic gradients of the order of a few tenths per thousand. The highest groundwater levels are located at the NW and SE sectors of the peninsula. Groundwater locally may be found in confined conditions since the presence of micritic carbonate levels of low permeability, or when the carbonate basement is below the mean sea level and covered by barely permeable Miocene and, somewhere, Plio-Pleistocene sediments. Fig. 5 shows the water level map of the Salento aquifer, defined at a regional scale by the Apulian government and published in *Piano di Tutela delle acque* (Water Protection Plan, 2009).

Freshwater floats on saltwater of marine origin as a lens of around 120 m maximum thickness at the highest hydraulic heads on the peninsula; the salt content varies between 0.2 and 0.5 g/L (Fidelibus & Pulido-Bosch, 2019). Groundwater discharges occur through coastal springs (sub-aerial and submarine, concentrated and diffuse), with Total Dissolved Solids (TDS) varying between 3.5 and 20 g/L.

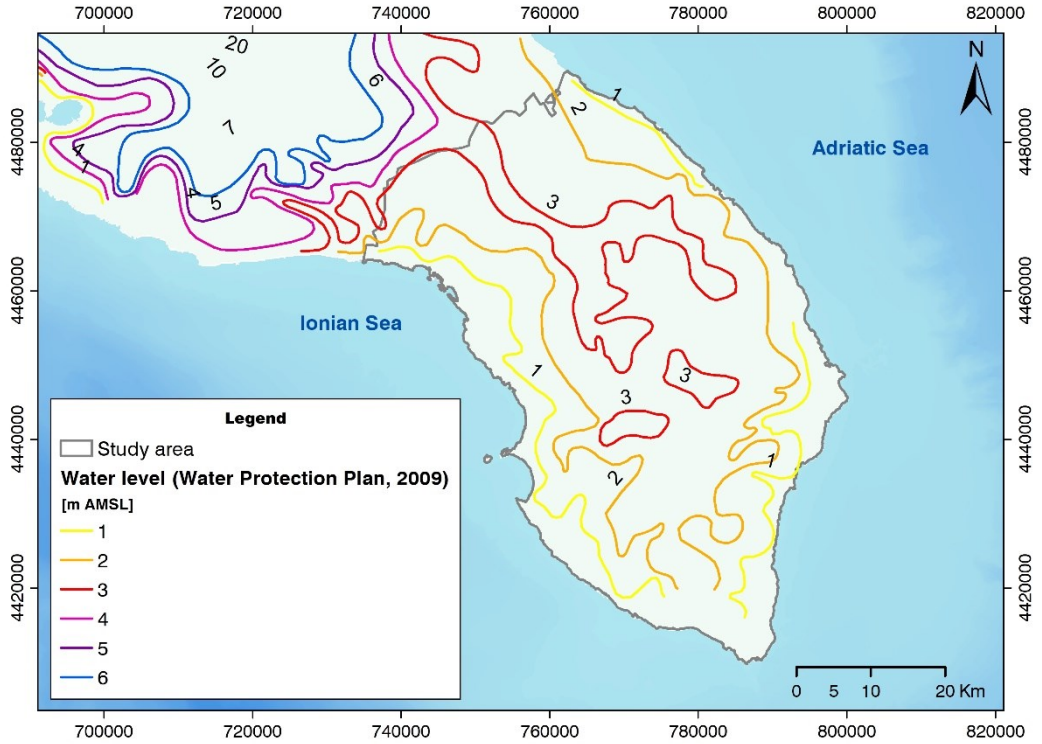


Fig. 5 - Water level map for the Salento aquifer (Water Protection Plan, 2009).

As a whole, the Salento aquifer represents a complex system that combines the complexity of a tectonically controlled karst with that of a coastal aquifer. The aquifer is, therefore, highly vulnerable to salinisation because of the structure of this system, where discontinuities and karst forms are routes of fast and deep intrusion of seawater and saltwater (Fidelibus & Pulido-Bosch, 2019). In Salento, droughts may easily propagate their effects to the coastal aquifer, deteriorating the groundwater qualitative and quantitative status and causing cascade crises (Parisi et al., 2018). Climate change currently threatens these issues, leading to increased groundwater exploitation as in other areas typified by a high level of urbanisation and low natural availability of water resources (Alsumaiei, 2020). Therefore, the knowledge and interpretation of the key mechanisms that control the hydrodynamic evolution of the aquifer systems are vital for the sustainable management of groundwater resources and urban development (Kong et al., 2021).



## 1.2 Meteorological data

### 1.2.1 Precipitation and temperature time series

Precipitation ( $Pr$ ), minimum ( $T_{min}$ ), and maximum ( $T_{max}$ ) air temperature data for the Salento study area were provided by the Civil Protection Service of the Apulian Government. Meteorological stations automatically record data at an hourly interval. After validation, they are published in the Bulletins of Civil Protection daily (source: <https://protezionecivile.puglia.it/centro-funzionale-decentrato/>). They report the amount of rain (mm) and the average minimum and maximum air temperature ( $^{\circ}C$ ) measured in 24 hours. Registrations are available with some gaps from the fifties of the past century; a few meteorological stations were recently installed. Fig. 6 shows the locations of the twenty-one rain gauge stations, of which nineteen also measure the minimum and maximum air temperature. Their main characteristics are listed in Table A1.

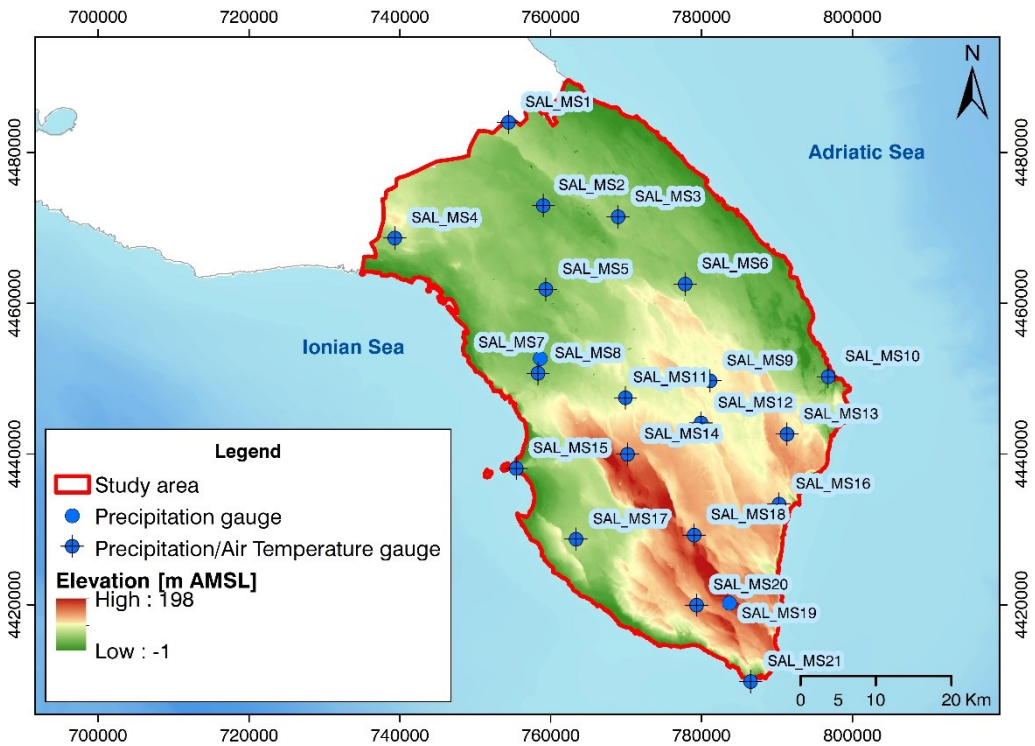


Fig. 6 - Location of meteorological gauge stations.

### **1.3 Hydrogeological data**

#### 1.3.1 Groundwater levels

The first efforts of groundwater monitoring of the Salento aquifer date back to the sixties when a net of 13 deep observation wells was drilled and purposely equipped for the control of seawater intrusion (Cotecchia, 2017; Tadolini & Tulipano, 1979). From 1970 up to 1986, researchers at Bari University regularly performed temperature and electrical conductivity logs. Monitoring went along with a few time gaps until 1994 when the first official groundwater network was established under the responsibility of the Apulian Regional Government. The *Ente Irrigazione Project* was carried out from the end of 1994 to early 1997 on a total of 127 wells (in static condition) distributed in the whole Apulian territory; multi-parameter logs and groundwater level measurements were sampled quarterly by bailers at different depths. Afterwards, the monitoring was interrupted for 11 years, starting again in 2007 with the establishment of a more complex network (*Tiziano Project*). Apart from several monitoring wells aimed at controlling superficial aquifers, the deep aquifers were monitored through the previous 127 wells in static condition with the addition of 262 private wells equipped with pumps. Sampling surveys (and water level/multi-parameter logs in the dedicated monitoring wells) were conducted with a bi-annual frequency from July 2007 to December 2011. Furthermore, sensors were installed in most wells for water level and/or multi-parameter continuous monitoring with data registered every hour. In the framework of the *Tiziano Project*, 28 wells were equipped with groundwater level monitoring sensors; administrative inconveniences and technical equipment and personnel changes caused some time gaps in the water level registrations. Therefore, only 11 out of 28 groundwater level recordings were used in this research for their continuity. The monitoring stopped again for five years. From 2015 to 2018, a new groundwater monitoring phase started on an updated net (*Maggiore Project*), including groundwater sampling, multi-parameter logging, and water level measurements carried out with bi-yearly frequency.

In conclusion, from 1995 to the present, the monitoring surveys were discontinuous, with significant time gaps and many changes at each new phase in the number of monitoring points, type of wells used, and monitoring frequency. *Tiziano Project*

constitutes the unique groundwater monitoring of the Apulian Region with a consistent hydrogeological dataset, i.e., groundwater levels, electrical conductivity, and multi-parameter time series.

Fig. 7 illustrates the monitoring well locations in the *Tiziano Project* over the Salento area, along with the wells selected for the current research on the basis of the consistency of available hydrogeological data. Their main characteristics are summarised in the Annex section (Table A2).

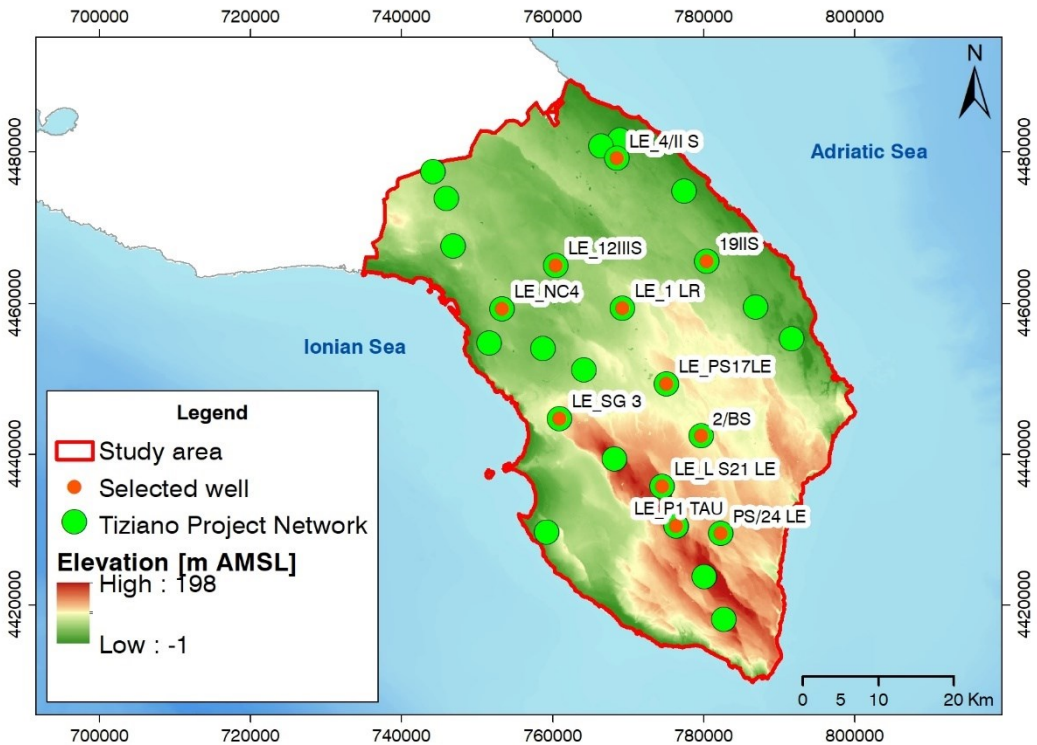


Fig. 7 - Tiziano Project's monitoring net over the Salento territory with the indication of the selected wells. Fig. 8 shows the daily hydrographs of the 11 selected wells over the available hydrological years. The evolution of groundwater levels indicates a smooth seasonal change between recharge and recession periods (Cai & Offerdinger 2016). During the summer dry season, water levels occasionally decline, which may be attributed to excessive abstractions to satisfy drinking and irrigation demand.

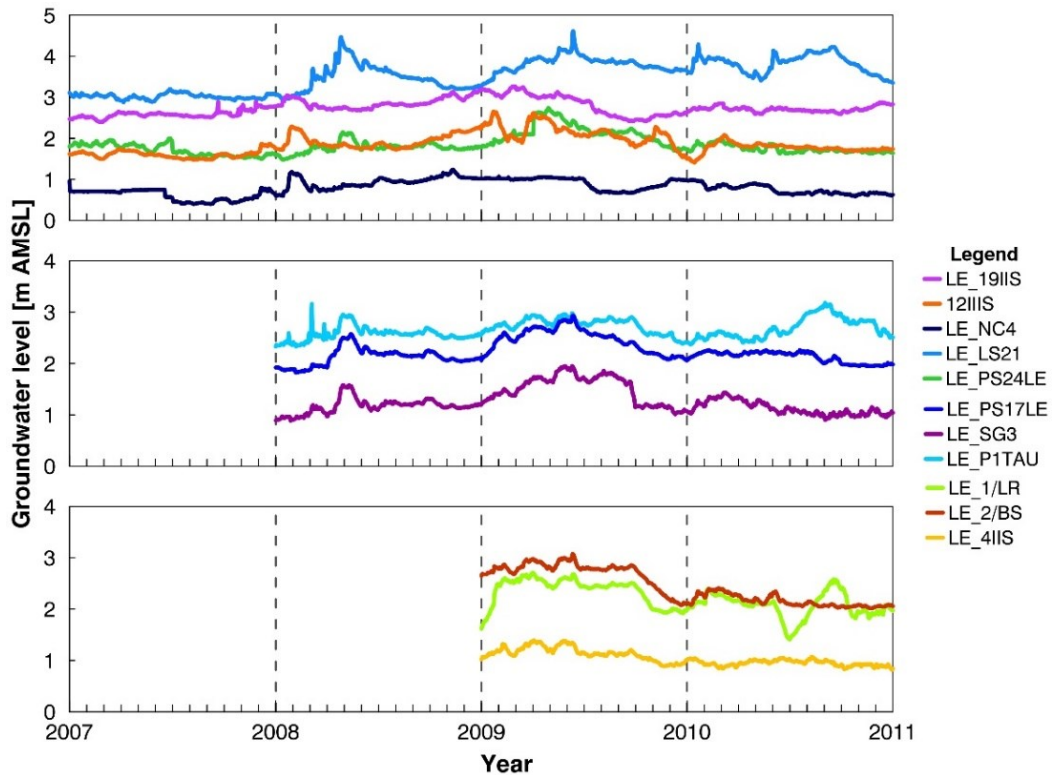


Fig. 8 - Daily hydrograph of the selected eleven monitoring wells.

### 1.3.2 Chemical analyses for the salinisation process assessment

The groundwater chemical analyses considered for this research refer to seven monitoring survey campaigns (*Tiziano Project*) from 2007 to 2011 and six surveys (*Maggiore Project*) from 2016 to 2018. These chemical analyses were processed through various statistical methodologies to investigate the salinisation process. The monitoring program of the *Tiziano Project* was carried out with a bi-annual frequency, with sampling at the end of each wet season (April-June) and at the end of each dry season (September-October). 449 samples were collected from 204 static and 245 pumping wells and analysed for numerous physical and chemical parameters. The final dataset includes 268 groundwater samples collected from not repeated wells (106 static and 162 pumping) over the seven sampling periods (three wet and four dry seasons). Fig. 9 illustrates the location of the considered wells.

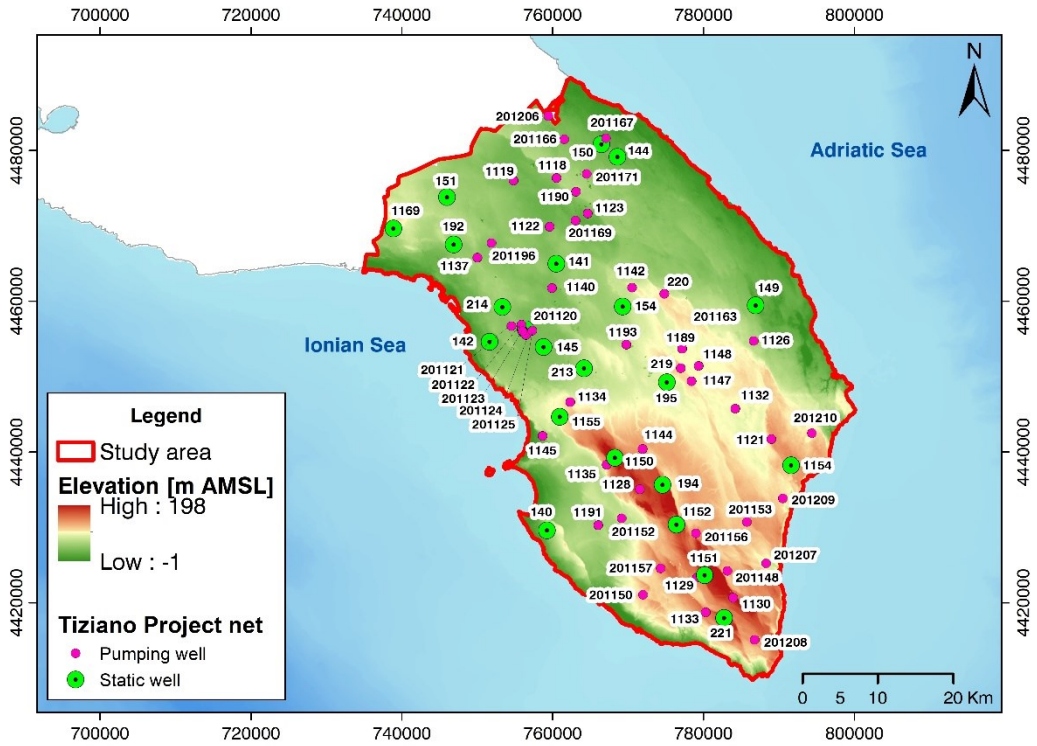


Fig. 9 - Groundwater quality monitoring network of the Tiziano Project. The numerical tag refers to the code of each sampling point.

Selection of the dataset to be used was based on the following criteria that were to be cumulatively met: (i) the shallowest measurement is considered for wells with samplings at various depths; (ii) monitoring wells should be located within the extent of the Salento aquifer; (iii) no or minimal missing/non-numerical values should occur; (iv) charge balance error should be less than 10%.

Likewise, the *Maggiore Project* had a bi-annual frequency with sampling referring to the end of each wet and dry season. The six monitoring surveys concerning the Salento aquifer comprise 290 groundwater samples (collected from 22 static and 35 pumping wells). The final dataset, applying the same set of multi-criteria pre-screening, includes groundwater samples collected from 31 monitoring wells (12 static and 19 pumping wells) with repeated sampling in the six sampling periods (three wet and three dry seasons) (Fig. 10).

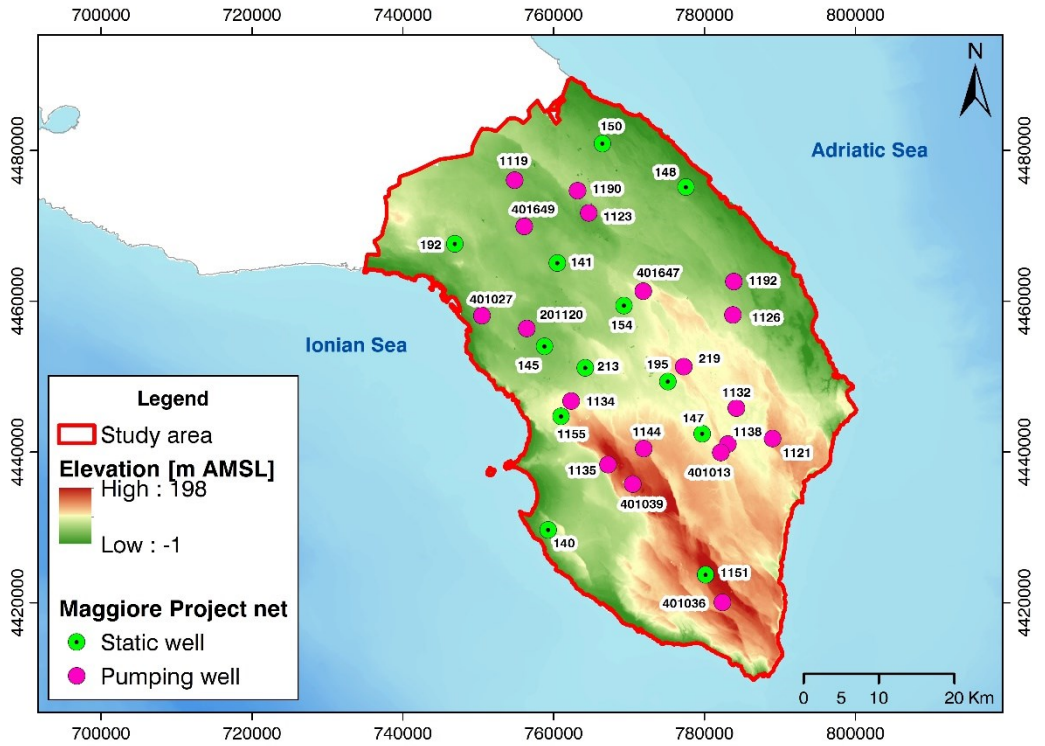


Fig. 10 - Chemical groundwater monitoring network of Maggiore Project.

### 1.3.3 Chemical analyses for the nitrate pollution assessment

The dataset used for the nitrate pollution assessment originates from a regional database, which includes chemical analyses conducted over the last 25 years on static and pumping wells belonging to regional monitoring networks and provincial consortia. Unfortunately, the monitoring protocols differ between the networks in sampling frequency, techniques and instruments, water sample storage, transport, and laboratory procedures. Aware of the non-homogeneity of the data and because of the three-dimensional vulnerability of the study area, this dataset was used in an alternative approach aiming to assess the qualitative evolution of nitrate pollution.

The dataset comprises 13,575 groundwater samples analysed for nitrate concentrations ( $\text{NO}_3^-$ ) from pumping and static wells measured from 1995 to 2021. It is organized into two periods (1995-2006 and 2007-2021) to account for the potential variability of nitrate pollution over time (Fig. 11).

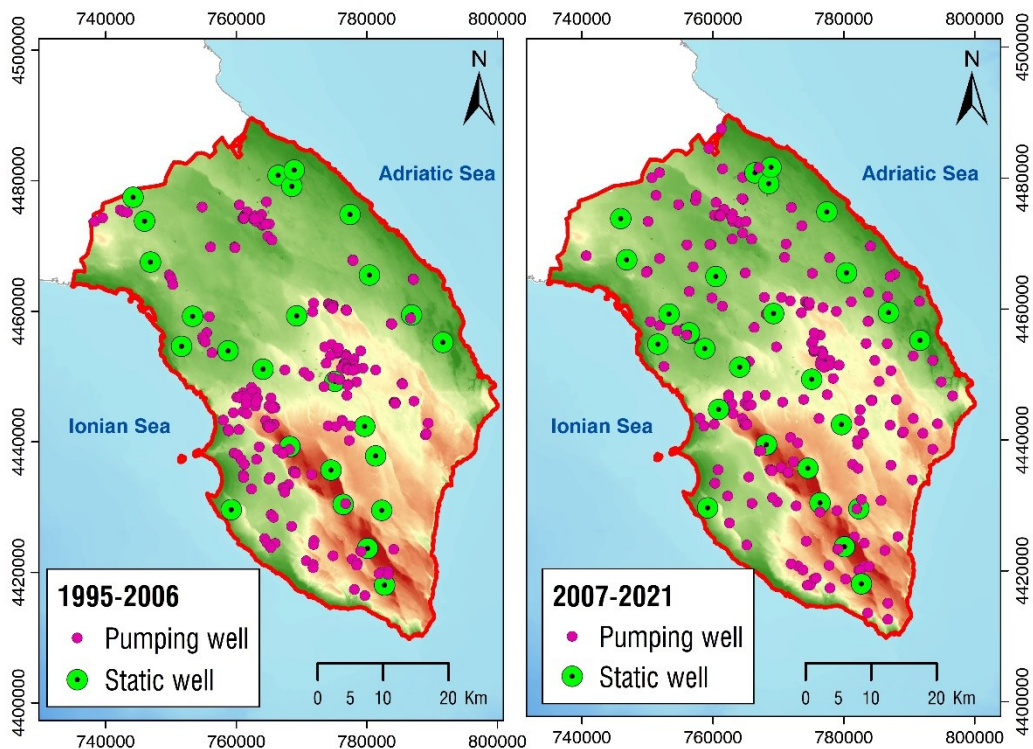


Fig. 11 - Location in the study area of sampling points with nitrate analyses related to 1995-2006 (left-hand side) and 2007-2011 (right-hand side) periods.

Table 2 summarises the main statistics of nitrate concentrations related to the two selected periods. Data show a significant increase in  $\text{NO}_3^-$  concentrations over time for both pumping and static wells. Since the first period, the maximum value has been significantly higher than the Maximum Admissible Concentration (MAC) of 50 mg/L, set by the 2000/60/EC European Water Framework Directive for water intended for human consumption. The minimum values are close to zero in both cases.

Table 2 - Main statistics of  $\text{NO}_3^-$  concentrations. Data are reported in mg/L.

<b>Pumping well</b>	<b>1995-2006</b>	<b>2007-2021</b>	<b>Static well</b>	<b>1995-2006</b>	<b>2007-2021</b>
N° of wells	251	231	N° of wells	26	26
Mean	26.0	30.9	Mean	9.4	17.3
Min	2.6	1.6	Min	0.5	0.06
Max	109	192	Max	56.5	130

## 2. MISSING VALUES FILLING-IN METHODS

### 2.1 *Missing values in weather data*

The availability of complete and accurate precipitation and temperature data is crucial for conducting reliable analyses (Sattari et al., 2017; Armanuos et al., 2020). However, it is not unusual to encounter issues with incomplete data, which can impact the quality and accuracy of successive analyses.

Gaps in time series may depend on several factors (Lotsi et al., 2017), i.e., malfunctioning instruments, maintenance of meteorological stations or human errors in manual measurement (Beaulieu et al., 2007). In the scientific literature, missing data in time series are often categorised into very limited and extensive or consecutive gaps (Aieb et al., 2019). When the gaps are minimal, they can be excluded from the dataset to simplify the analysis (Song et al., 2008). In the second case, various methods for imputing missing values were developed (Li et al., 2007). The effectiveness of these imputation methods can vary depending on several aspects, including the percentage of missing values, the mechanism of data loss, the variables under consideration, and their respective correlations. Specifically, Little & Rubin (1987) identified three categories of missing data: Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR).

In the first case, the missing values occur randomly and do not depend on the values of the variable with missing data or any other variables in the dataset. In mathematical term, it is expressed:

$$P(r|X_{obs}, X_{missing}) = P(r) \quad (1)$$

where  $X_{obs}$  is the observed data,  $X_{missing}$  the missing one and  $(r)$  is the distribution condition of missing values.

In the second case, the missingness is related to other observed variables in the dataset, but not directly to the missing values; thus, the observed data can predict or explain the probability of missingness.



$$P(r|X_{obs}, X_{missing}) = P(r|X_{obs}) \quad (2)$$

The last case is the most challenging because the probability of having missing values depends on the variable in question. Since the missingness is related to the unobserved data, simply modelling the observed data may not be sufficient to impute or estimate the missing values accurately.

$$P(r|X_{obs}, X_{missing}) = P(r|X_{missing}) \quad (3)$$

Researchers have explored different imputation techniques to address this issue, such as interpolation (Simolo et al., 2009), regression-based approaches (Duarte et al., 2022), and machine learning algorithms (Sattari et al., 2020).

Furthermore, two imputation approaches exist: (i) single imputation, when a single plausible value replaces the missing value, and (ii) multiple imputations, when a set of reasonable values is generated and combined.

To address missing data in the time series of precipitation and air temperature at Salento's meteorological stations, we selected and assessed five imputation methods from 1960 to 2005. The evaluation of each algorithm performance relied on metrics such as Pearson's coefficient (R), Nash-Sutcliffe Efficiency (NSE), and Similarity Index (SI), calculated between observed and estimated time series. To facilitate this evaluation, we identified a data recording period without missing values that applied to all stations. For precipitation, the test period ranges from 01/01/1971 to 31/12/1976, while for temperature, it covers the period from 01/01/2000 to 31/12/2022. Successively, a rate of 10% of data was randomly assumed to be missing since the highest missing value percentage found in the dataset was less than 7% (Table 3).

Hereafter, target and reference stations refer to meteorological stations with missing values and those selected to estimate them, respectively.

Table 3 - Missing values percentages over the historical period 1960-2005 (1971-2005 for SAL-MS6 and SAL-MS14).

<b>Code</b>	<b>Name</b>	<b>Precipitation (%)</b>	<b>Max Air Temperature (%)</b>	<b>Min Air Temperature (%)</b>
SAL_MS1	S. Pietro Vernotico	0.00	2.87	3.06
SAL_MS2	Novoli	0.80	-	-
SAL_MS3	Lecce	0.18	2.52	2.92
SAL_MS4	Masseria Monteruga	4.15	-	-
SAL_MS5	Copertino	2.15	-	-
SAL_MS6	Melendugno	1.85	-	-
SAL_MS8	Nardò	0.91	2.20	2.58
SAL_MS10	Otranto	0.00	3.93	4.08
SAL_MS11	Galatina	1.49	-	-
SAL_MS12	Maglie	0.00	3.06	3.07
SAL_MS13	Minervino di Lecce	0.36	7.48	7.91
SAL_MS14	Collepasso	6.95	-	-
SAL_MS15	Gallipoli	0.00	1.80	2.01
SAL_MS16	Vignacas-trisi	0.84	4.36	4.73
SAL_MS17	Taviano	0.91	0.55	0.93
SAL_MS18	Ruffano	0.36	-	-
SAL_MS20	Presicce	0.18	3.91	4.27
SAL_MS21	S. Maria di Leuca	0.98	0.32	0.44

Gaps in each target station were then filled by selecting a set of reference stations close to those affected by missing values according to climatic and physical behaviour. In fact, being a peninsula facing the Adriatic Sea on one side and the Ionian Sea on the other, the weather conditions are not influenced solely by altitude, as is generally the

case inland, but also by other factors such as proximity to the sea. The temperatures in Salento are significantly influenced by the mitigating presence of the Ionian and Adriatic Seas, whose winds blowing on both sides establish distinct weather conditions. Consequently, the selection of reference stations for each target station relied on four criteria to be cumulatively met:

- (i) Pearson's correlation coefficient  $R$  greater than 0.70,
- (ii) distance between them less than 20 Km,
- (iii) altitude difference of less than 80 m
- (iv) exposure to the same sea (Adriatic or Ionian).

A convergence threshold was also set. In this case, the criterion for determining convergence is based on the difference between two consecutive steps. Specifically, the algorithm continues iterating until the difference between the results of two successive steps is less than 0.01. This difference is measured in millimetres for precipitation data and Celsius degrees for temperature data. By setting this convergence criterion, the algorithm ensures that the imputation process continues until a certain level of accuracy is achieved. Once the difference between two consecutive steps falls below the specified threshold, the metrics were calculated between observed ( $O$ ) and estimated ( $E$ ) data over the specified time interval. Depending on the number of iterations completed and the metrics coefficients, the more appropriate filling-in method for the case study was selected for precipitation and temperature, respectively. Fig. 12 shows a summary of the adopted procedure.

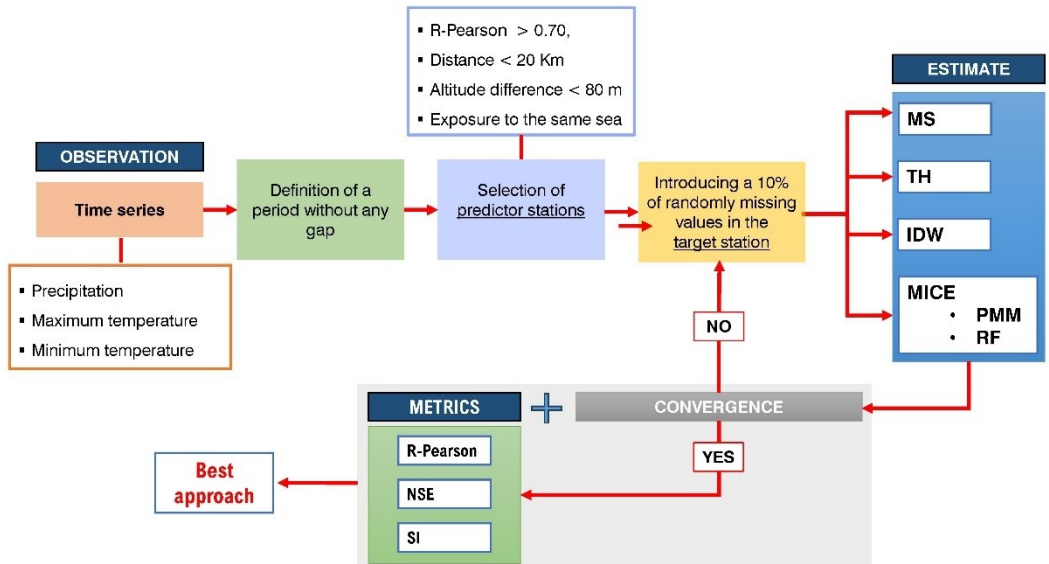


Fig. 12 - Filling-in missing values procedure.

## 2.2 Filling-in methods and adopted metrics

The following paragraphs briefly describe the five imputation methods and the three metrics adopted.

### 2.2.1 Mean Substitution

The Mean Substitution method (MS) is the most straightforward imputation method. It replaces the daily missing value at the target station with the corresponding mean of the values recorded in the selected reference stations. Ben Aissia et al. (2017) suggested that the MS technique effectively applies when missing values are less than 10%. The estimation of missing data is achieved by calculating the arithmetic average of the observations from the selected reference stations, as shown below:

$$E_i = \frac{\sum_{i=1}^N O_i}{N} \quad (4)$$

where  $E_i$  is the estimated value of the missing data,  $O_i$  is the observation at  $i^{th}$  nearest weather station, and  $N$  is the number of the closest stations.

### 2.2.2 Thiessen polygons

The Thiessen polygons (TH) is a commonly used method in hydrology. It assumes a weight factor for each meteorological station in proportion to the area it represents. It assumes that meteorological parameters are constant in each Thiessen polygon and match the reference station measurement. Here, missing values correspond to the values from the polygon of the reference station in which the target station occurs. The Thiessen polygons are defined by excluding the target meteorological station.

### 2.2.3 Inverse Distance Weighting

In the Inverse Distance Weighting (IDW) method, missing values at the target station correspond to a weighted average of the data at the reference stations. The weighting parameters inversely depend on the spatial distances from the target station. This method is appropriate for highly correlated data. The estimated missing value is given by:

$$E_i = \frac{\sum_{i=1}^n (O_i/d_i)}{\sum_{i=1}^n (1/d_i)} \quad (5)$$

where  $d_i$  is the distance between target and predictor stations.

### 2.2.4 Multiple Imputation by Chained Equations

Multiple Imputation by Chained Equations (MICE) is an advanced method which accounts for multiple sets of plausible imputed values through a prediction model built on already available data in the reference stations (Zhang, 2016). It is based on the MAR assumption.

According to van Buuren & Groothuis-Oudshoorn (2011), this procedure consists of three main steps. The first step is to replace missing data at the target station with values drawn from a distribution fitted to the reference time series. The second step is the esteem and pool (following Rubin, 1987) of the linear regression parameters of each group of the imputed dataset. The third involves the evaluation of the final linear regression parameters with which missing values are calculated. This study employed

two MICE approaches: Predictive Mean Matching (MICE\_pmm, Aguilera et al., 2020) and Random Forest (MICE\_rf, Jing et al., 2022).

MICE\_pmm works by first estimating  $k$  set of plausible values using a predictive model, typically a regression model, between each imputation and observation  $I_i(X_{missing}, X_{obs})$ . The predicted values are then matched to observed values from the dataset, preserving the distributional characteristics of the observed data and obtaining  $k$  different estimation results. Matching predicted values to observed values generates imputations consistent with the existing data, producing more realistic and accurate results than other imputation techniques. Finally, the set of imputed values is combined according to Rubin (1987):

$$E_i = \frac{\sum_{i=1}^k I_i(X_{missing}, X_{obs})}{k} \quad (6)$$

MICE\_rf is a non-parametric imputation technique based on the principle of random forest. It works iteratively, continuously updating the variably imputed matrix and evaluating its performance between iterations. The algorithm stops when the difference between two successive imputation results reaches a set threshold or the maximum number of iterations the user specified. The multiple sets of estimated values are then combined.

### 2.3 Metric coefficients

The metric coefficients used to analyse and evaluate the performance of each filling-in missing value method are briefly described below. In the following  $E$  is the estimated value,  $O$  is the observed value for the target station, and  $n$  represents the length of the considered time series.

#### 2.3.1 Pearson's correlation coefficient

The Pearson correlation coefficient (1896) is a parametric method for measuring the linear dependence between two series. It ranges between  $-1$  (inverse relationship, i.e., when one variable increases, the other decreases) and  $+1$  (direct

relationship, i.e., both variables move in the same direction). A value of zero indicates the absence of correlation. The Pearson's correlation coefficient is expressed as:

$$R = \frac{\sum_{i=1}^n (O_i - \bar{O})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 (E_i - \bar{E})^2}} \quad (7)$$

### 2.3.2 Nash-Sutcliffe Efficiency

The efficiency value NSE (1970) ranges from  $-\infty$  to 1. A positive efficiency of up to 1 indicates higher accuracy in the model predictions. An efficiency value of 0 indicates that the model predictions are as accurate as simply using the mean of the observed data. Conversely, when the metric is less than zero ( $NSE < 0$ ), it implies that the estimated mean is a less accurate predictor than the mean of the observed data. The NSE is calculated as follows:

$$NSE = 1 - \frac{(E - \bar{O})^2}{(O - \bar{O})^2} \quad (8)$$

### 2.3.3 Similarity Index

The Similarity index indicates the percentage of agreement between the estimated and observed values. This metric ranges between 0 and 1 to denote absent to perfect agreement, respectively (Willmott, 1981). The SI formula is given by:

$$SI = \frac{\sum_{i=1}^n (E - O_i)^2}{\sum_{i=1}^n (|O_i - \bar{O}| + |E_i - \bar{O}|)^2} \quad (9)$$

### **3. DATA-DRIVEN METHODS FOR QUANTITATIVE CHARACTERISATION**

#### **3.1 *Correlation indices***

Drought is a complex natural phenomenon that affects society and its environment in many ways (Loukas et al., 2003). In the scientific literature, it is common to distinguish it into four categories (Van Loon et al., 2012; Chang et al. 2018; Wanders et al., 2010): (i) meteorological drought, which refers to a lack of precipitation in a large area and over a long period of time; (ii) soil moisture drought (also called agricultural drought), which relates to a deficiency of soil moisture, usually in the root zone; (iii) hydrological drought, associated to negative anomalies in surface and sub-surface water; (iv) socio-economic drought, linked to the failure of water resources systems to meet water demands and consequent ecological or health-related impacts. When these typologies coincide in a specific area, they could generate groundwater drought, i.e., lack of groundwater expressed in terms of recharge, storage, or heads over a particular period (Van Loon, 2015). Human activities also play a significant role in influencing the occurrence of this natural hazard. Climate change can lead to an increased demand for water, which, in turn, drives aquifer mining for agricultural or industrial purposes, ultimately depleting groundwater resources. Land use changes can alter evapotranspiration patterns, depriving aquifers of sufficient recharge during dry periods.

To evaluate the potential impacts of climate change on future water resources availability, it is crucial to understand the hydrological response of aquifers to climate variability and specifically to drought periods. This aspect is particularly significant for arid and semi-arid regions, such as the Mediterranean area (Green et al., 2011), where reduction of future precipitation and increased evapotranspiration may easily lead to critical water shortages in several regions (Lionello & Scarascia, 2018).

A comprehensive understanding of how aquifers respond to external conditions is crucial for evaluating the potential impact of climate changes on groundwater recharge,



storage, and overall availability. Furthermore, it can aid water utilities and policymakers in formulating the most appropriate management strategies and mitigation measures. Numerous indices have been set up to investigate various aspects of drought, including its duration, timing, severity, and spatial extent. These indices help in interpreting the principal hydrological drought processes and their impacts. The selection of a specific index depends on various factors, with the availability of extensive meteorological time series being a critical issue. Several studies correlate these meteorological indices with groundwater levels to investigate aquifer hydrodynamics. Bloomfield & Marchant (2013) introduced the Standardized Groundwater Level Index (SGI) to characterise groundwater droughts. They discovered a correlation between the SGI and the Standardised Precipitation Index (SPI), with the strongest correlation observed during a site-specific period of precipitation accumulation. Babre et al. (2022) examined the relationship between the SGI and various meteorological and hydrological drought indices in the Baltic region. The findings indicated that meteorological drought indices, such as the SPI, Standardised Precipitation and Evapotranspiration Index (SPEI), and Reconnaissance Drought Index (RDI), exhibited the most significant correlation with groundwater conditions. Ndehedehe et al. (2023) conducted a global assessment of spatial correlations between SPEI and GWLs. Their study revealed positive and negative correlations worldwide, with a predominance of positive values at the 12-month aggregation scale, indicating the potential impact of climate change on groundwater. Negative correlations, on the other hand, may result from extensive water extraction or complex geological conditions and aridity, indicating a delayed response. Secci et al. (2021) investigated the correlation between SGI, SPI and SPEI to infer the effects of climate change on groundwater in Tuscany. Using the relationships between groundwater and meteorological indices established in the historical period, they projected future SGI values using climate variables provided by an ensemble of Regional Climate Models (RCMs) under different emission scenarios.

The correlations between hydrological and hydrogeological time series were adopted, assuming the existence of a potential correlation between meteorological indices and groundwater level. Specifically, meteorological drought indicators were used as input

signals, while groundwater levels were used as output. SPI and SPEI have been selected among all the drought indices because of their simple computation based on precipitation and temperature time series. Considering at least 30 years of precipitation and temperature data, SPI and SPEI allow for identifying the frequency and intensity of wet and dry periods.

Pearson's, Spearman's and Kendall's correlation coefficients were selected since they are the most used indicators of monotone association, with the latter two usually suggested for non-normally distributed data. These three correlation coefficients can be described as the differently weighted mean values of the same concordance factors. Chon (2010) investigated the intrinsic ability of Pearson's, Spearman's, and Kendall's correlation coefficients to affect the statistical power of tests for monotone association in continuous data. The superiority of Pearson's correlation stems from the fact that it better reflects the degree of concordance and discordance of pairs of observations for some types of distributions. On the other hand, it is largely sensitive to outliers, especially for large sample sizes with a higher probability of datasets with outliers.

The following paragraphs briefly describe Spearman's and Kendall's correlation coefficients and meteorological indices used in this study. See 2.3.1 for Pearson's correlation coefficient.

### 3.1.1 Kendall's tau correlation coefficient

The Kendall's correlation coefficient (1938) is a nonparametric indicator for evaluating the strength and direction of association between two variables measured on at least an ordinal scale. It is the nonparametric alternative to Pearson's and Spearman correlation coefficients. This coefficient requires examining two assumptions before its use: (i) the variables should be measured in an ordinal or continuous scale, and (ii) it is suitable that the two variables follow a monotonic relationship.

For the presence of ties, the  $\tau_B$  statistic was calculated (Agresti, 2012). It ranges from  $-1$  (negative association) to  $+1$  (positive association) with a value of zero indicating the absence of association. The Kendall  $\tau_B$  coefficient is defined as:

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \quad (10)$$

$$n_0 = n(n - 1)/2 \quad (11)$$

$$n_1 = \sum_i t_i(t_i - 1)/2 \quad (12)$$

$$n_2 = \sum_i u_j(u_j - 1)/2 \quad (13)$$

in which  $n_c$  and  $n_d$  are the number of concordant and discordant pairs, respectively,  $t_i$  is the number of tied values in the  $i^{th}$  group of ties for the first quantity, and  $u_j$  is the number of tied values in the  $j^{th}$  group of ties for the second quantity.

Kendall  $\tau_B$  coefficient was calculated using the package *DescTools* of R-Studio software (R Core Team, 2019)

### 3.1.2 Spearman’s rank-order correlation coefficient

Spearman’s rank-order correlation coefficient (1904) is a nonparametric indicator of the statistical dependence between the rankings of two variables. It permits to define if a monotonic function can represent the relationship between the two variables. The mathematical structure of the Spearman coefficient is defined as the Pearson one substituting the values of the two variables with their ranks (Myers & Well, 2013). Thus, the Spearman coefficient assesses monotonic relationships (whether linear or not). The sign of the estimated indicator defines the direction of association between the independent and the dependent variables: the positive coefficient indicates that both variables increase. In contrast, the negative value indicates that the dependent variable increases when the independent one decreases. A Spearman coefficient of zero denotes no tendency between the two variables. Its magnitude ranges from -1 to +1.

Spearman rank correlation coefficient was calculated using *stats* package of R-Studio software (R Core Team, 2019).

### 3.1.3 Meteorological indices

The SPI (McKee et al., 1993) and SPEI (Vicente-Serrano et al., 2010; Beguería et al., 2013) are meteorological drought indices developed to characterise wetness or

dryness conditions. The first is computed using monthly precipitation as input data, while the second uses monthly accumulated climatic water balance anomalies, defined as the difference between precipitation and potential evapotranspiration (PET).

They are normalised indices representing the probability of occurrence of precipitation and water balance anomalies compared with the ones over the long reference period for climate. Positive SPI or SPEI denote wet conditions, whereas negative values refer to dry conditions (Table 4).

Table 4 - SPI's drought category according to Mckee et al. (1993).

<b>SPI Value</b>	<b>Drought Category</b>
$SPI > 0$	no drought
$0 \geq SPI > -1$	minor drought
$-1.0 \geq SPI > -1.5$	moderate drought
$-1.5 \geq SPI > -2$	severe drought
$SPI \leq -2$	extreme drought

The indices can be evaluated at different time scales, ranging from 1 to 48 months. Thus, their interpretations depend on the time scales selected (Khan et al., 2008). The 1 to 6-months indicates short-term conditions: it reflects seasonal variations of precipitation and soil-moisture, defining agricultural drought. The 9- to 12-months shows long-term conditions in precipitation patterns, highlighting hydrological and even groundwater drought. Finally, the 48-months account for socio-economic impacts (Potop et al., 2014).

The calculation of the SPI and SPEI indices requires three steps. The cumulated precipitation and water balance anomalies at different time scales are computed, fitted to an appropriate probability distribution, and then transformed into a standard normal distribution. To consider the zero precipitation, a piecewise probability distribution was used:

$$p(x) = \begin{cases} p_0 + (1 - p_0)G(x, \gamma); & x > 0 \\ p_0 = \frac{n_{p=0} + 1}{2(n + 1)}; & x = 0 \end{cases} \quad (14)$$

where  $p$  is the probability distribution,  $p_0$  is the zero-precipitation probability,  $n_{p=0}$  is the number of zero precipitation in the whole time series,  $G(x, \gamma)$  is the Gamma distribution with  $\gamma$  parameters, and  $x$  is a precipitation element. Finally, the inverse normal cumulative distribution function (with mean zero and variance one) yields the time series of SPI values.

Regarding the SPEI, the potential evapotranspiration, i.e., the amount of evaporation and transpiration that would occur if a sufficient water source was available, can be calculated with three approaches: (i) the Thornthwaite method (Thornthwaite, 1948), (ii) the Hargreaves method (Hargreaves, 1994), and (iii) the Penman–Monteith method (Allen et al., 1998). The first method is the simplest because it only needs monthly mean temperature values and the site's latitude. The second one computes monthly reference evapotranspiration (ET<sub>0</sub>) of grass crops and requires the site minimum and maximum temperature values and latitude. The last method is considered the most accurate but most complex and data demanding by far, compared to the previous two methods, because it calculates ET<sub>0</sub> of a hypothetical reference crop, based on minimum and maximum temperature values and time series of monthly mean daily external radiation, monthly mean daily wind speed at 2 m height, monthly mean daily bright sunshine hours and monthly mean cloud cover in percentage. The Food and Agriculture Organization of the United Nations (FAO) and the American Society of Civil Engineers (ASCE) advised the use of the Penman–Monteith formulation (Walter et al., 2000; Droogers & Allen, 2002). Even with the known limits of accuracy, the Thornthwaite method was adopted to estimate the monthly PET due to the limited requirements and, in parallel, the lack of the input data required by the more advanced Penman-Monteith formulation.

### **3.2 Seasonal and trend decomposition**

Time series data on groundwater level fluctuation can provide valuable insights into various aspects, including the nature of the aquifer, the influence of climatic conditions, and significant abstractions. The analysis of such data often requires more advanced methods to discern the impact of inter-annual, annual, and short-term variations. Nevertheless, time series decomposition is a valid technique for disentangling different factors within a time series. The technical literature features numerous studies applying Seasonal and Trend decomposition using the Loess (STL) procedure to analyse groundwater patterns and dynamics. Lafare et al. (2016) employed the STL technique to assess daily groundwater responses in the Eden Valley, UK. Colyer et al. (2022) investigated the influence of meteorological and geological heterogeneity on the same aquifer. In another study, Fan et al. (2023) utilised STL for monitoring groundwater dynamics and employed spatial interpolation to detect spatio-temporal patterns of groundwater level and salinity in the Yellow River Delta. Niranjannaik et al. (2022) applied STL to correlate GWL patterns with rainfall and groundwater abstraction variability for the Betwa River in India. Moreover, Duy et al. (2021) explored GWL trends, memory effects in alluvial aquifers, and response times between surface water and groundwater in the Vietnamese Mekong Delta.

STL is a flexible and reliable approach for decomposing time series data. It offers several advantages when dealing with monthly and quarterly data with various forms of seasonality. Furthermore, it effectively handles changes in both the seasonal component rate and the trend cycle's smoothness. It also addresses outliers effectively to avoid influencing the trend or seasonal component (Hyndman & Athanasopoulos, 2018). This statistical technique helps isolate different information in the time series, i.e., it decomposes it into three constituent parts: seasonality, trend, and residual. In this way, it is possible to identify distinct contributions, which are conversely undetectable using the whole time series because the individual information overlaps. As a component, seasonality represents regular and predictable patterns that recur at fixed intervals. For instance, it identifies wet and dry periods in the context of rainfall patterns. The trend component provides a general direction of the overall data, indicating whether

there is an upward or downward trend. Lastly, the residual part, often called noise, captures the random fluctuations of the background signal. These unpredictable variations depend on local conditions over time. In the case of rainfall patterns, the residual component can highlight significant and extreme wet or dry events.

On the other hand, Loess is a technique that uses locally weighted regression to fit a smooth curve through points in a sequence. This method is beneficial for estimating nonlinear relationships (Cleveland et al., 1990).

The advantages of STL include its simplicity, speed of computation, robustness of results, flexibility, and excellent data visual analytics. It is widely used in many sectors, including natural and environmental science (Lafare et al., 2016; Xia et al., 2019, Niranjannaik et al., 2022).

The STL method can be classified as an additive approach for decomposing the time series according to the following formula:

$$Y_t = T_t + S_t + R_t \tag{15}$$

where  $Y_t$  is the GWL,  $T_t$  is the trend,  $S_t$  is the seasonal, and  $R_t$  is the remainder component.

However, it is necessary that the acquisition interval of the dataset to be decomposed with STL is constant and that any missing data are filled with linear interpolation.

Following the methodology of Lafare et al. (2016) and Colyer et al. (2022), the results of the STL analysis can also be investigated using the ratio of the variance of each component to the variance of the original GWL:

$$Variance\ ratio_{component} = \frac{Variance_{component}}{Variance_{original}} \tag{16}$$

The closer the variance ratio of a component approaches 1, the more influential the same component is in the studied time series.

### **3.3 Time series analyses**

Examining a karst aquifer on a regional scale poses significant difficulties due to its considerable variety and heterogeneity. The term “karst aquifer” encompasses a broad range of limestone formations, varying from less developed and dispersed flow aquifers to highly concentrated and channelised flow aquifers. Although many karst aquifers typically display a blend of these flow types, accurately evaluating the degree of karstification within a particular aquifer and discerning the importance of each flow component constitutes a challenging task. In the field of hydrology, a commonly employed approach for this purpose is time series analysis. Among the various techniques, correlation and spectral analyses are frequently applied as they are straightforward to implement and yield valuable insights into the behavior of an aquifer. The fundamental premise underlying both correlation and spectral analyses is rooted in a system-based approach, wherein statistical functions are employed to establish relationships between input factors and resultant outcomes. Mangin (1984) was the first to investigate input-output relationships in the karst system and provide a classification based on time series analyses. Specifically, the karst aquifer is viewed as a filter that modifies, retains, or attenuates the input signal, represented by the precipitation, into an output signal (Larocque et al., 1998), as the spring discharge (Benavente & Bosch, 1985; Larocque et al., 1998; Angelini, 1997; Amraoui et al., 2003; Fiorillo & Doglioni, 2010; An et al., 2019), the groundwater level (Imagawa et al., 2013; Delbart et al., 2014; Cai & Offerdinger, 2016; Chiaudani et al., 2017), the river flow rate (Bailly-Comte et al., 2008; Chiaudani et al., 2017) or other physical or chemical parameters characterising the aquifer (Bailly-Comte et al., 2011). The extent to which the input signal is transformed provides valuable information about the flow dynamics within the system. Time series studies involves both univariate and bivariate analysis in the time and frequency domains (Pulido-Bosch, 2021).

#### **3.3.1 Time Domain Analysis**

The Autocorrelation Function (ACF) analyses the existence of a linear tendency of successive values of a single series, giving a measure about the memory effect that the previous values of a time series exert on subsequent data (Chiaudani et al., 2017).



In general terms, ACF estimates the time necessary to forget the initial conditions of each time series (Larocque et al. 1998; Box et al., 2013; Chiaudani et al., 2017):

$$ACF(k) = \frac{C(k)}{C(0)} \tag{17}$$

$$C(k) = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{18}$$

where  $k$  is the time lag (from  $k = 0$  to  $m$ ),  $m$  is the cutting point,  $n$  is the length of the time series  $x_t$ , and  $\bar{x}$  represents its mean,  $C(0)$  is the covariance at value 0 and  $C(k)$  is the covariance at value  $k$ . Mangin (1984) indicates an  $m$  value less than  $n/3$ , not based on theoretical concepts but on his empirical research carried out to define the best possible estimate for truncation of  $m$ . He verified that correlograms were skewed for a truncation point between  $n/3$  and  $n/2$ .

The shape of the correlogram provides useful information on a karst system; a gentle slope shows the resilience of the aquifer in terms of infiltration time and/or major groundwater storage. In contrast, a steep slope highlights a more rapid flow through the aquifer or a poor storage capacity (Padilla & Pulido-Bosch, 1995; Panagopoulos & Lambrakis, 2006; Duvert et al., 2015; Chiaudani et al., 2017). The time required before the ACF reaches a value of 0.2 is known as memory effect (Mangin, 1984). For a random variable, the ACF decreases quickly and reaches zero for very short time lags (Delbart et al., 2016); instead, a slow decrease shows an idle nature of the karst, characterised by large storage (Sađir et al., 2020).

The Cross-Correlation Function (CCF) is useful to understand the impulse response of the karst system to the recharge process since rainfall events are normally known as a random process. Defining the two discretised chronological series of length  $n$ ,  $x_t$ , and  $y_t$  so that the first causes the second, the CCF with a truncation point  $m$  ( $k = 0, 1, 2, \dots m$ ) is not symmetrical.

$$CCF(k) = \frac{C_{xy}(k)}{\sqrt{C_x^2(0)C_y^2(0)}}, k > 0 \tag{19}$$

$$C_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) \quad (20)$$

$$CCF(k) = \frac{C_{yx}(k)}{\sqrt{C_x^2(0)C_y^2(0)}}, k < 0 \quad (21)$$

$$C_{yx}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(x_{t+k} - \bar{x}) \quad (22)$$

where  $C_{xy}(k)$  is the covariance between input and output, and  $C_x(0)$  and  $C_y(0)$  are their respective standard deviations. A significant correlation between input and output time-series at 95% confidence interval is reached when the CCF exhibits a correlation coefficient greater than the standard error  $2/n^{0.5}$  (Diggle, 1990, Lee et al., 2006, Cai et al., 2016).

The CCF indicates a random process with an impulse response: when the CCF shows a maximum for  $k > 0$  it suggests an impact of the input signal on the output signal; on the contrary (for  $k < 0$ ), the output influences the input (Larocque et al., 1998). The delay calculated from a time lag equal to 0, to a time lag corresponding to the maximum value of CCF represents the transfer velocity of the aquifer, which is known as response time (Cai et al., 2016).

### 3.3.2 Frequency Domain Analysis

Spectral analysis allows to break down the total variance of a variable based on the frequency of the events that compose it and therefore to highlight its structure. The spectral approach compared to the correlation approach is twofold; it expresses in the frequency domain what the correlogram reflects in the time domain. Mangin (1984) pointed out that the analysis of a phenomenon in the frequency domain can often facilitate its interpretation. The spectral analysis allows detecting the randomness of a time series since the absence of peaks of the spectral density function characterises a purely random phenomenon for characteristic frequencies. In the frequency domain, we could obtain different functions to characterise aquifer behaviour, particularly with complex systems where the quick flow superimposes on the baseflow. A short sign of the

primary expressions used for conducting the univariate and bivariate analyses in the frequency domain is presented in the following, based on a few studies (Padilla & Pulido-Bosch, 1995, Laroque et al., 1998, Panagopoulos & Lambrakis, 2006).

Since the long-term trend of each time series must be eliminated to satisfy the assumption of stationarity, a linear detrending was applied (Molénat et al., 1999).

The spectral density function is the ACF in the frequency domain through a Fourier transformation:

$$S(f) = \frac{1}{2\pi} \left[ 1 + 2 \sum_{k=1}^m ACF \cos(2\pi fk) \right] \tag{23}$$

The above expression is known as raw spectral estimate (Chow, 1969) because it is not a consistent evaluation of the spectral density (Molénat et al., 1999). Thus, a smoothing procedure is performed to determine the smoothed spectral estimate (Chow, 1969). This procedure permits to give more importance to the periodogram ordinates in correspondence to the interested frequencies, and progressively less weight to the periodogram ordinates at increased frequencies (Chatfield & Xing, 2013). The Daniell or “rectangular” window (Priestley, 1981, Molénat et al., 1999) was applied as smoothing procedure. The spectral density allows defining the duration of the impulse response of the system (Laroque et al, 1998) in terms of regulation time, which is a measure of the inertia of the system; it gives a sign of the length of the impulse response of the system. Differently from the memory effect, it is less sensitive both to the sampling interval of the time series and the correlation between distant precipitation events (Zhang, 2013). It is expressed as the period corresponding to the half of the maximum spectral intensity as frequency goes to zero (Meng et al., 2021):

$$T_r = \frac{S(f = 0)}{2} \tag{24}$$

The Cross–Spectral Density Function  $S_{xy}(f)$  coincides with the Fourier transformation of the CCF and, due to its asymmetry, it is necessary to express it using the complex expression:

$$S_{xy}(f) = |\alpha_{xy}(f)| \exp[-i\phi_{xy}(f)] \quad (25)$$

where  $\alpha_{xy}(f)$  represents the Cross-Amplitude Function (CAF) and  $\phi_{xy}(f)$  the Phase Function (PHF).

The CAF indicates the duration of the impulse response function and the filtering of the periodic components of input data, enabling the characterisation of the modulating effect of the aquifer in the short, medium, and long term. It provides a decomposition of the total covariance between input and output, as a function of frequency, and indicates how the system transforms the input function.

The PHF estimates the phase shift or mean delay between input and output signals for different frequencies (Meng et al., 2021). It generally varies in the interval  $-\pi$  to  $\pi$ . According to Padilla and Pulido-Bosch (1995) the average delay over the range of considered frequencies is equal to  $d = A/2\pi$  where  $A$  is the slope of the regression line on the points of the PHF. It indicates the mean delay between precipitation and GWL response.

From the spectral density functions of the input and output signals and the CAF, two other useful expressions could be obtained, the Coherency (COF) and the Gain (GAF) Functions that are formulated as follows:

$$COF = \frac{(\alpha_{xy}(f))^2}{S_x(f)S_y(f)} \quad (26)$$

$$GAF = \frac{\alpha_{xy}(f)}{\sqrt{S_x(f)}} \quad (27)$$

*COF* indicates whether certain variations in the output signal can be attributed to the input one. A loss of coherency at lower frequencies indicates a scarcely karstified aquifer where baseflow dominates, whereas a high *COF* also to greater frequencies, denotes a highly karstified system. *GAF* identifies the type of flow and expresses the amplification/attenuation of the input signal: the frequency associated to (i)  $GAF = 1$  represents the duration of baseflow, (ii)  $GAF = 0.4$  indicates the duration of quick flow, and (iii)  $0.4 < GAF < 1$  characterises an intermediate flow.

### 3.4 Wavelet Analysis

The Fourier transform decomposes a signal into sine waves of different frequencies, filtering certain information that can be discerned in the time domain. When examining the Fourier transform of a time series, it becomes impossible to pinpoint the exact moment when a specific event occurred. Many natural phenomena, especially meteorological events, exhibit non-stationary signals. Non-stationary characteristics such as drifts, trends, or abrupt changes often contain the most crucial aspects of a time series, and Fourier analysis fails to accentuate them. In contrast, the wavelet method offers an intriguing alternative by preserving both time and frequency information during the transformation. Wavelet analysis entails the use of variable-sized regions or windows, enabling the utilisation of longer time intervals to extract more information about low-frequency content and shorter periods to scrutinize high-frequency components. One notable advantage of wavelet analysis lies in its capacity for localized examination, concentrating on specific segments of an extended signal and effectively identifying even the tiniest discontinuities that might otherwise remain elusive. This statistical method decomposes the signal into various iterations of the mother wavelet, which are shifted and scaled. The objective is to advance or retard the mother wavelet in time while stretching or compressing it, thereby identifying the precise duration for each frequency window (Holman et al., 2011). Wavelets can be distinguished into Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). In contrast to DWT, which uses a discrete subset of all possible values, the latter is the sum of the signal multiplied by the shifted and scaled version of the wavelet function over the entire time interval.

The wavelet function  $\psi_0(\eta)$  located in both time and frequency domains, presents a zero mean (Farge 1992, Percival & Walden 2000, Grinsted et al., 2004). Technical literature reports several mother wavelets. The well-known Morlet wavelet was used in the following because it assures a good balance between time and frequency localisation (Grinsted et al., 2004; Holman et al., 2001).

$$\psi_0(\eta) = \pi^{-1/4} e^{-i\omega_0\eta} e^{-\frac{1}{2}\eta^2} \tag{28}$$

where  $\omega_0$  is the dimensionless frequency and  $\eta$  is the dimensionless time.

CWT is a suitable method to identify anomalies in the signal. However, sometimes it is likewise remarkable to overlap two different signals for identifying similarities. Considering a time series  $(x_n, n=1, \dots, N)$  with the uniform time steps  $\delta t$ , CWT is defined as the convolution of  $x_n$  with a scaled and translated mother wavelet of  $\psi_0(\eta)$  (Torrence & Compo, 1998; Grinsted et al., 2004):

$$W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=0}^N x_{n'} \psi_0^* \left[ (n' - n) \frac{\delta t}{s} \right] \quad (29)$$

where  $\sqrt{\frac{\delta t}{s}}$  is the normalisation factor and the asterisk represents the complex conjugate. The complex part of  $W_n^X(s)$  is the local phase and  $|W_n^X(s)|^2$  is the wavelet power. However, CWT presents artefacts at the edges because the wavelet is not entirely localised in time. Since these are time series of finite length, errors will occur at the beginning and end of the wavelet power spectrum. An approach is to fill the time series' tails with zeroes before applying the wavelet transform and remove them afterwards (Meyers et al. 1993). This restriction on the width of the wavelet leads to a cut-off of the coefficients at the ends of the scalogram, usually represented with a black line delimiting the Cone of Influence (COI). Values outside the COI are not used for calculating the average spectrum because they may be unreliable due to the stated reasons. The COI is defined as the area in which the wavelet power caused by an edge discontinuity has fallen to  $e^{-2}$  times the edge value (Torrence and Compo, 1998; Grinsted et al., 2004).

Considering two times series X and Y, the Cross Wavelet Transform (XWT) counts for two CWTs, estimating their common power and relative phase in the time-frequency domain. It can be expressed as:

$$W^{XY}(s) = W^X(s) \cdot W^{Y*}(s) \quad (30)$$

in which the asterisk is the complex conjugate. Further, the cross-wavelet power is defined as  $|W^{XY}|$  while the complex argument is the locale relative phase between X and Y.

Wavelet Coherence (WTC) is the correlation coefficient in frequency and time domains that quantify the relationship between two-time series (Cazelles et al., 2008) in the time and frequency domains and can be estimated as:

$$R_n^2(s) = \frac{|S(s^{-1}W_n^{XY}(s))|^2}{S(s^{-1}|W_n^X(s)|^2) \cdot S(s^{-1}|W_n^Y(s)|^2)} \quad (31)$$

in which  $S$  is the smoothing operator (Torrence and Compo, 1998; Grinsted et al., 2004).  $R_n^2(s)$  ranges between 0 and 1; the closer the value is to 1, the stronger the correlation between the two-time series (Liu et al., 2018).

## 4. DATA-DRIVEN METHODS FOR QUALITATIVE CHARACTERISATION

### 4.1 *Multivariate statistical analysis*

Multivariate statistical analysis (MVSA) is a valuable approach that offers a means to manage multiple geochemical and physical parameters, treating them as variables, and grouping water samples with similar characteristics into distinct clusters (Güler et al., 2002; Ghesquière et al., 2015; Machiwal et al., 2018). This method has the potential to reveal spatial and temporal variations in groundwater quality, identify key hydrochemical processes, and assess how they change over time, as evidenced by various studies (Güler et al., 2002; Papatheodorou et al., 2007; Pacheco Castro et al., 2018; Bahrami et al., 2020; Prusty & Farooq, 2020). Among the different multivariate techniques, the Hierarchical Cluster Analysis (HCA) and the Factor Analysis (FA) were selected to recognise the spatial and temporal dynamics of groundwater salinisation in the Salento aquifer. Prior to applying these methods, it is crucial to conduct preliminary investigations to ensure the quality and reliability of the datasets. Therefore, after selecting chemical analyses (sampling survey periods and validation of related groundwater analyses) and compiling the dataset, the MVSA techniques were applied to the compiled dataset. The Q-mode HCA was used to cluster water samples based on the similarities of their parameters, identifying groups of samples that share similar characteristics. R-mode FA was applied to analyse the chemical analyses of water samples, allowing the identification of the underlying factors which can reveal the key chemical processes affecting groundwater. In the following, the steps involved in the statistical analyses are briefly described.

#### 4.1.1 Preliminary investigation

The water chemistry dataset goes through data screening using predetermined criteria. One such condition involved excluding water samples that did not meet a specified Charge Balance Error (CBE). CBE is a standard measure to judge the validity and



quality of water analyses by evaluating that the total sum of all the positive charges (cations) must equal the total sum of all negative charges (anions):

$$CBE = \frac{\sum cations - |\sum anions|}{\sum cations + |\sum anions|} * 100 \quad (32)$$

Acceptable water analyses usually have *CBE* less than  $\pm 5\%$ . Considering that the chemical dataset collected for the study area comprised numerous samples with high salinity, water samples with a *CBE* inside the range of  $\pm 10\%$  were considered for further evaluation (as recommended in Güler et al., 2002). Furthermore, groundwater samples were organised into different surveys to compare water quality over time. One approach excludes wells not sampled in all the surveys, thus ensuring meaningful comparisons between different periods. In cases where available data are limited or sampling surveys do not precisely align in time, one can consider using all accessible analyses, provided that the distribution of wells remains consistent relative to the extent of the study area. The water chemistry dataset was then examined regarding values below the detection limit and missing data values since such datasets are improper for MVSA applications (Farnham et al., 2002). To replace concentration values recorded as “less than”, the technique suggested by Sanford et al. (1993) recommends multiplying by 0.55 the lower detection limit of the instrument. Parameters for which the database had over 15% of non-numerical and missing values were excluded.

In the second step, the statistical distribution of parameters was checked using the Kolmogorov-Smirnov test (Kolmogorov, 1933; Smirnov, 1948). Typically, MVSA requires the normalisation and standardisation of parameters under investigation (Alther, 1979; Romesburg, 1984; Reimann & Filzmoser, 2000). Therefore, for parameters exhibiting a non-normal distribution, we applied the Box-Cox transformation method (Box & Cox, 1964, 1982). In cases where a parameter remained non-normally distributed even after the Box-Cox transformation, excluding it from subsequent statistical analyses was recommended. However, recent studies applying MVSA for the geochemical characterisation of groundwater do not provide detailed specifications regarding the normalisation of the investigated variable but instead focus on its standardisation (Arumugam et al., 2023; Karangoda & Nanayakkara, 2023; Solano et al., 2023). Thus, in the

final step, all parameters were standardised (by calculating their standard z-scores) to ensure that each parameter was weighted equally (Johnson & Wichern, 1992). The standardised values of the final parameters to be analysed represent the input dataset for MVSA.

#### 4.1.2 Q-mode Hierarchical Cluster Analysis

The Q-Mode HCA is a statistical technique used to identify and group objects or samples based on their similarity or dissimilarity. Various distance metrics can be used depending on the nature of the data and the research question. (i.e., Euclidean distance, Manhattan distance, or correlation distance). It can be either agglomerative (bottom-up) or divisive (top-down). In the first case, clustering starts with each object as an individual cluster and successively merges the closest groups until a single set remains. Divisive clustering begins with all objects in one group and recursively splits clusters into smaller ones until each object forms its group. This technique results in a dendrogram (Davis, 1986), a tree-like diagram that illustrates the hierarchical relationships among the samples. The height or distance on the dendrogram indicates the dissimilarity between clusters, and a horizontal line (a.k.a. phenon line) drawn across the dendrogram branches identifies the number of groups. In Q-mode HCA, Ward's method (Ward, 1963) was used as the linkage scheme to assess the similarities among members of water groups, while the Euclidean metric was selected to account for distance. The obtained clusters were interpreted and analysed by means of boxplots representing the major anions and cations for each group, as well as maps associated with each monitoring survey. These maps visually displayed the identified water clusters using distinct symbols.

#### 4.1.3 R-mode Factor Analysis

R-mode FA is a statistical technique used to establish the correlation structure among the observed (original) variables and to extract unobserved (latent) variables named factors (Dalton & Upchurch, 1978; Basilevsky, 1994). It aims to identify fewer unobserved factors that can explain the relationships among a set of observed variables.

The sampling adequacy test of Kaiser-Meyer-Olkin (KMO) was applied to determine the suitability of the compiled database for FA application (Kaiser, 1974, 1981). KMO values are considered acceptable if higher than 0.5, whereas values close to unity define the best FA application related to the dataset. The number of factors summarising the key processes related to the dataset was selected according to the “Explained variance criterion”. In this case, an orthogonal varimax rotation was applied to maximise the loading related to one factor and minimise the loading related to the others (Davis, 1986). Then, the factor loading matrix was examined to define variable-factor associations and describe the hydrogeochemical process associated with each latent factor, where the higher the factor loading (close to  $\pm 1$ ), the stronger the influence of variables on the respective factor. Conversely, factor loadings close to zero identify a weak correlation between the variables and factors. The maps displaying the spatial distribution of the factor scores (FSs), which are related to the intensity of the process underlying each factor (Dalton & Upchurch, 1978), help to understand the spatial evolution of the identified factors.

## 4.2 *Hydrogeochemical Facies Evolution-Diagram*

The HFE-D proposed by Giménez-Forcada (2010, 2014) allows exploring the hydrogeochemical variations that occur over time by groundwater freshening and salinisation processes (available at <https://hidrologia.usal.es/HFE-D.htm>). The diagram includes four main “heteropic” facies ( $\text{NaHCO}_3$ ,  $\text{NaCl}$ ,  $\text{CaHCO}_3$ , and  $\text{CaCl}$ ), which are defined by coupling sodium or calcium and chloride or bicarbonate percentages when higher than 50 % (Fig. 13). Furthermore, it includes mix facies in which the named ion is higher than any other ion, but is less than 50 % (Giménez-Forcada, 2010, 2014; Giménez-Forcada & Sánchez San Román, 2015). The abscissae of the HFE-D separately mark the percentages of  $\text{Na}^+$  and  $\text{Ca}^{2+}$ , aiming to recognise heteropic hydrochemical facies (HFs - as  $\text{NaHCO}_3$  and  $\text{CaCl}$ ) and mix facies that occur in coastal aquifers under the effects of base-exchange reactions triggered by mixing of fresh- and saltwaters. The percentage of chloride in the ordinate tracks groundwater salinisation, while the percentage of bicarbonate or sulphate (depending on the dominant anion in freshwater) typifies the HFs typical of groundwater recharge.

The Conservative (or non-reactive) Mixing Line (CML), built with a freshwater end-member and a saltwater water end-member (usually seawater), separates (i) the freshening phases and facies, which are located to the left above the CML, and (ii) the intrusion phases and facies, which are to the right under the CML. The position of the CML in the HFE-D is crucial for the water sample classification. In this work, since the groundwater salinisation is primarily due to saltwater upconing caused by intense groundwater exploitation (Tadolini et al., 1982; Tulipano & Fidelibus, 2002), the concentrations of the major ions of the saltwater end-member were defined as the average of chemical analyses associated with saltwater samples collected in past monitoring programs (1986) in deep wells reaching saltwater beneath freshwater in the study area (Fidelibus et al., 2011 – data is available from the MEDSAL Observatory at <https://medsal.eu/observatory/> and reported in Table 5). Regarding the freshwater end-member, the HFE-D model automatically selects the chemical composition of the freshest groundwater collected during the analysed sampling surveys.

Table 5 - Saltwater end-member chemical parameters set for CML definition expressed in mg/L.

	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Na <sup>+</sup>	K <sup>+</sup>	HCO <sub>3</sub> <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	Cl <sup>-</sup>
<b>Seawater</b>	712	1,090	11,805	391	138	3,577	20,000

Different sub-stages related to freshwater and intrusion phases can be identified in accordance with the chemical composition of sampled waters. Both freshening sub-stages (from the beginning to the end of the freshening process are f1, f2, f3, f4 and FW) and the intrusion sub-stages (from the beginning to the end of the intrusion process are i1, i2, i3, i4 and SW) help to identify the salinisation dynamics of groundwater samples (Giménez-Forcada, 2014, 2019). The HFs distribution was finally plotted for each survey to determine the temporal evolution of the salinisation process. These maps follow the methodological approach proposed by Giménez-Forcada (2010, 2014, 2019).

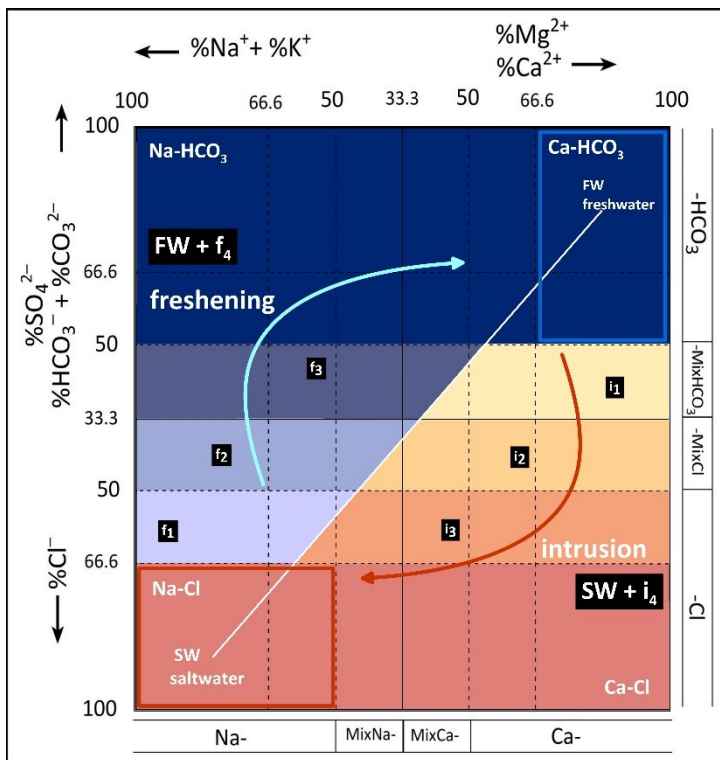


Fig. 13 - Hydrogeochemical Facies Evolution-Diagram (HFE-D) proposed by Giménez-Forcada (2010, 2014).

### 4.3 *Geostatistic*

Geostatistics is a branch of applied statistics recognised as a helpful tool for the spatial analysis of countless variables for more than 50 years (Matheron, 1963). It is commonly used in geology, environmental science, mining, hydrology, agriculture, and more, where observed points have geographic coordinates or are otherwise related to space.

Among the geostatistical interpolation methods, the Kriging technique, which originated in mining geology (Krige, 1951), differs from more straightforward techniques (i.e., Inverse Distance Weighted Interpolation or Linear Regression) by considering spatial correlations between empirical observation points to interpolate values in the spatial field. Furthermore, Kriging provides estimates of uncertainty for each interpolated value. The primary goal is to examine the data's spatial patterns and variability, making predictions and estimates at unsampled locations and considering the spatial correlation of data values. Therefore, data points close to each other are more likely to have similar values than distant points apart.

Assuming that data are part of a realisation of an intrinsic random function, Kriging can be conceptualised as a two-step procedure: firstly, it involves determining the spatial covariance structure of the sampled points by fitting a variogram  $\gamma(h)$ . Subsequently, the derived weights from this covariance structure are utilised to interpolate values for unsampled points or blocks across the entire spatial field.

The variogram is a graphical representation showing how the variance between data points changes with distance or lag, providing information about the spatial structure of data. The “experimental” variogram plots observed values, while the “theoretical” variogram is the distributional model that best fits the data.

Three parameters are essential components of the variogram model (Fig. 14):

1. Nugget represents the discontinuity in the variogram at the origin (lag distance equals zero). It reflects the variance of the data at very short distances, typically at the scale of measurement error or sampling error.
2. Sill defines the upper limit of uncertainty when making predictions at unsampled locations, i.e., the variance of the data as the lag distance increases.

3. Range measures the spatial continuity beyond which the spatial autocorrelation becomes negligible. It is crucial to determine the influence of a data point as distance increases. A short range indicates that nearby points strongly influence each other, while an extended range means a smoother variation over longer distances.

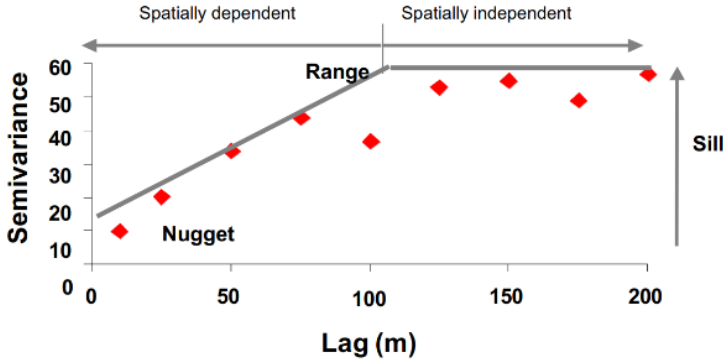


Fig. 14 - Experimental variogram and its main parameters.

After computing the experimental variogram directly from the data, it is possible to fit a theoretical function from those available in the scientific literature. Some of the most widely used theoretical variograms are shown in Fig. 15 and summarised in Table 6.

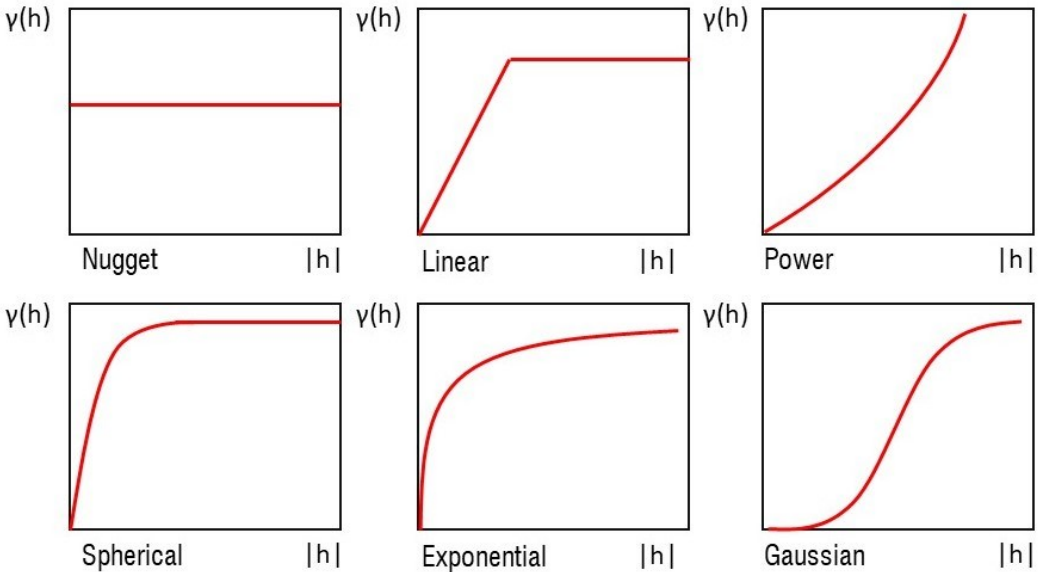


Fig. 15 - Example of theoretical variograms.

Table 6 - Theoretical variogram functions.

Model	Function $\gamma(h)$	Parameters
Nugget	$\begin{cases} C_0, & h > 0 \\ 0, & h = 0 \end{cases}$	$C_0 > 0$
Linear	$\theta \cdot h$	$\theta > 0$
Power	$\theta \cdot h^s$	$\theta > 0, 2 > s > 0$
Spherical	$\begin{cases} \left( \frac{3h}{2\alpha} - \frac{1h^3}{2\alpha^3} \right) \sigma^2, & 0 \leq h \leq \alpha \\ \sigma^2, & h > \alpha \end{cases}$	$\sigma^2 > 0, \alpha > 0$
Exponential	$\sigma^2 \left( 1 - \exp \left( -\frac{h}{l} \right) \right)$	$\sigma^2 > 0, l > 0$
Gaussian	$\sigma^2 \left( 1 - \exp \left( -\frac{h^2}{L^2} \right) \right)$	$\sigma^2 > 0, L > 0$

The two main assumptions for Kriging to provide the best linear unbiased prediction are (i) stationarity and (ii) isotropy. Stationarity considers that the joint probability distribution remains constant. Consequently, the same variogram model is valid and applicable throughout the entire study domain, since parameters like the overall mean of the values, range and sill of the variogram do not exhibit spatial variations. Isotropy assumes uniformity in all directions, although it is possible to consider anisotropy by fitting variograms in different directions.

Several Kriging techniques include Ordinary Kriging (OK), Simple Kriging (SK), Universal Kriging (UK), Indicator Kriging (IK) and more. In general, the accuracy of these interpolation methods depends on the number of sample points and their spatial distribution. Moreover, it is reasonably sensitive to the selected variogram since its parameters are user-defined. In the environmental field, the stationarity hypothesis can be challenging. In the following paragraphs, OK and IK are briefly described.

#### 4.3.1 Ordinary Kriging

Ordinary Kriging (OK) is the most widely used spatial prediction method. It is used to estimate a value in an unobserved point  $x_0$  of a region based on  $n$  spatial neighbourhood observations, assuming that the mean and variance of the values are constant across the spatial field.



The predicted value  $Z^*(x_0)$  is expressed as:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \tag{33}$$

where  $\lambda_i$  is the weighting coefficient calculated on the  $n$  positions within the neighbourhood points. The optimal selection of  $\lambda_i$  should minimize the bias between the predicted  $Z^*(x)$  and the real (unknown) value  $Z(x)$ .

The weighting coefficients are selected based on the requirements of (i) unbiasedness, i.e., the expected value of the estimation error  $Z^*(x_0) - Z(x_0)$  must be zero, and (ii) the best or minimum variance estimation, i.e., the expected value of the square error must be as small as possible. These conditions result in a system of  $n + 1$  linear equations with  $n + 1$  unknowns, notorious as the Kriging system:

$$\begin{aligned} - \sum_{j=1}^n \lambda_j \gamma(X_i - X_j) + \nu &= -\gamma(X_i - X_0), i = 1, 2, \dots, n \\ &= \sum_{j=1}^n \lambda_j = 1 \end{aligned} \tag{34}$$

where  $\nu$  is a Lagrange multiplier. The mean square error of estimation is given by the formula:

$$\sigma^2_o = E[(Z^*(X_0) - Z(X_0))^2] = -\nu + \sum_{j=1}^n \lambda_j \gamma(X_i - X_j) \tag{35}$$

In the global neighbourhood approach, all observations are included. In contrast, the local neighbourhood approach comprises only a subset of known points, set by the user as (i) a fixed number of observations closest to the point  $X_0$ , (ii) all observations within a specific radius or (iii) a combination of these strategies.

This research used OK to spatially plot the outcomes obtained from FA and HFE methods in each survey and compare them in time.

### 4.3.2 Indicator Kriging

Traditional geostatistical methods may provide a misleading depiction of the phenomenon under study when dealing with variables with high variability like nitrates. In the case of asymmetric distributions and the presence of outliers, classical geostatistical techniques become less reliable for predicting values at unknown locations within the spatial domain.

To overcome this issue and prevent overestimation or underestimation, a more sophisticated approach is required for selecting the mean of the values, as suggested by Glacken & Blackney (1998). In these cases, using a representation of the phenomenon based on the probability of exceeding fixed thresholds is more beneficial. To this end, one of the most suitable methods is the Indicator Kriging (IK) introduced by Journel (1983). It is a spatial interpolation technique to estimate a variable's conditional cumulative distribution function at an unsampled location. Unlike SK and OK, IK stands out as a nonparametric method. It employs new indicator variables to generate predictions instead of utilising the estimated variable's actual values.

Thresholds are established based on the distribution of the data and initial statistical analysis. These limits are then used to construct the cumulative distribution function. IK converts the original data into binary form through a binary transformation: one if the value exceeds the threshold and zero if it does not. Once obtained the binary conversion, the procedure aligns with that of OK, including the subsequent definition of the variogram. The prediction made by IK is interpreted as the probability of exceeding the threshold value at an unknown point. This is an estimate of the proportion of values in the vicinity of a point that is higher than the indicator or threshold value. The essence of the IK method is the definition of an indicator function  $i(x; z_k)$  which can respectively assume the 1 or 0 value depending on the set threshold  $z_k$ :

$$i(x; z_k) = \begin{cases} 1 & \text{if } z(x) \leq z_k \\ 0 & \text{if } z(x) > z_k \end{cases} \quad (36)$$

IK does not estimate the indicator variable but provides a method for least-squares estimation of the cumulative distribution function at the threshold  $z_k$ :

$$\hat{I}(x_0; z_k) = E\{x_0; z_k | (n)\} \tag{37}$$

where  $(n)$  represents the proximity of other values at the point  $x_0$ . Due to the way the random indicator function  $I(x; z)$  is constructed, the expected value of  $I(x; z)$  corresponds to the cumulative distribution function of the random function  $z_x$ . In fact:

$$E [I(x; z)] = 1 * Prob\{Z(x) \leq z\} + 0 * Prob\{Z(x) > z\} \tag{38}$$

$$E [I(x; z)] = Prob\{Z(x) \leq z\} = F(z) \tag{39}$$

Therefore, the estimation of the indicator variable at threshold  $z$  becomes:

$$\hat{I}(x_0, z_k) = E\{x_0, z_k | (n)\} = Prob\{Z(x_0) \leq z | (n)\} \tag{40}$$

$$\hat{I}(x_0, z_k) = \sum_{i=1}^n w(x_i; z) I(x_i; z) + \left[ 1 - \sum_{i=1}^n w(x_i; z) \right] F(z) \tag{41}$$

in this case, the  $w(x_i; z)$  are the weights assigned to the threshold  $z$ .

IK was applied to a regional database consisting of chemical analyses conducted over the past 25 years, with measurements ranging from 1995 to 2021 on static and pumping wells included in regional monitoring networks and consortia. The laboratory analyses were carried out by the Regional Agency for Environmental Prevention and Protection (ARPA, <https://www.arpa.puglia.it/>), and details of the analytical procedures can be found at [https://www.arpa.puglia.it/pagina3128\\_qualit.html](https://www.arpa.puglia.it/pagina3128_qualit.html).

Unfortunately, the measurements in the dataset vary in terms of sampling frequency, techniques and instruments, water sample storage, transport, and laboratory protocols. Recognizing the non-uniformity of the data and considering the three-dimensional vulnerability of the study area, we employed the collected dataset to assess the qualitative changes in nitrate pollution in both space and time within the Salento aquifer. The set thresholds follow the 2000/60/EC and 2006/118/EC European Water and Groundwater Framework Directives, respectively, and are 20 (good quality), 40 (intermediate quality), and 50 mg L<sup>-1</sup> (poor quality), as widely used in the scientific literature (Chica-Olmo et al., 2014; Piccini et al., 2012).

## 5. CLIMATE CHANGE MODELS

### 5.1 *Climate Models and uncertainty related to BC methods*

Climate Change Models, also known as Earth System Models, are computer-based tools used to simulate and project future climate conditions. These models are based on complex mathematical equations representing the physical, chemical, and biological processes within the Earth's climate system. They were developed by incorporating various components of the Earth's system, including the atmosphere, oceans, land surface, ice, and vegetation. These components are represented as grids or cells, with equations describing how they interact and exchange energy, momentum, and mass with each other. There are different types of climate models, including Earth System Models (ESMs), General Circulation Models (GCMs), and Regional Climate Models (RCMs).

ESMs are more comprehensive models that integrate the physical components of the climate system and biogeochemical cycles, such as carbon and nitrogen cycles, capturing interactions between the atmosphere, oceans, land surface, and ecosystems and providing a complete representation of the Earth system.

GCMs are large-scale models that simulate the Earth's climate system globally. They are designed to simulate past climate conditions, understand current climate processes, and project future scenarios under different greenhouse gas emission scenarios on a global scale. However, they are often affected by uncertainty predominantly due to low resolutions (approximately 100–250 Km) that inevitably lack regional details (Randall et al., 2007). To this end, several downscaling methods have been developed to transform the large-scale information of GCMs to finer scales (25–50 Km), resulting in Regional Climate Models (RCMs) (Teutschbein & Seibert, 2012; Maraun, 2016).

RCMs are particularly useful for studying climate change impacts at a regional scale, assessing regional climate variability, and providing more detailed information for climate adaptation and decision-making at the local level. Notwithstanding, RCMs show several limitations (Christensen et al., 1998, Teutschbein & Seibert, 2012, Varis et al.,

2004) since, depending on the study area, they could provide inaccurate seasonal precipitation patterns, overestimate wet days, or incorrectly yield extreme temperatures. For this purpose, many bias-correction (BC) methods were developed to overcome the significant bias in RCMs. They can adapt RCMs to local observations regarding mean and variance (scaling methods) or distribution probabilities. However, using BC methods may introduce uncertainty in climate risk assessments due to the potential for various algorithms to yield varying impact results (Iizumi et al., 2017). Further uncertainty may be introduced by the accuracy of observation data (Kim et al., 2015) and the historical period used as a reference (Chen et al., 2015; Gampe et al., 2019). Noto et al. (2023b) argued that the intricate sequence of modelling procedures, i.e., emission scenarios, climate models, downscaling and/or bias-correction techniques, and hydrological models, involves a certain level of uncertainty that spreads throughout the entire process, resulting in divergent and occasionally unexpected outcomes.

Through uncertainty analysis, Senatore et al. (2022) demonstrated in a catchment located in Southern Italy that the primary source of uncertainty is introduced by GCMs, followed by RCMs and applied BC methods. Thus, an ensemble of RCM simulations and field observation data may be considered together with different bias correction methods to simultaneously evaluate the uncertainties of each simulated dataset and the performance of every technique (Fantini et al., 2018). This suggestion is particularly relevant when the hydrological variables are used to investigate future impacts in areas highly vulnerable to climate change, such as the Mediterranean basin (IPCC, 2014; Giorgi & Lionello, 2008).

Climate models need to be representative of the local climatic conditions; therefore, they are evaluated and validated using historical climate data to ensure they accurately reproduce observed climate patterns. Once validated, they are used to project future climate scenarios under different greenhouse gas emission scenarios, providing valuable insights into the potential impacts of climate change on various sectors, ecosystems, and regions. To discuss the future hydrological impacts of climate change on the Salento aquifer, simulated daily precipitation and minimum and maximum temperature

data were selected from an ensemble of different RCMs belonging to the EURO-CORDEX domain (Jacob et al., 2014).

Several BC methods were applied at daily time steps to improve the RCMs' simulations in terms of local climate variability representation, and their performance was evaluated by comparing results with monthly and annual precipitation and temperature (minimum and maximum) observations in the historical period (1971-2005). After selecting two-time windows for near to medium (2031-2060) and distant (2071-2100) future, trend analyses were performed on monthly and annual time steps, comparing the projected with the corresponding precipitation and temperature data of 1971-2005.

For this purpose, two different approaches may be adopted to analyse the performance of BC methods in future climate scenarios. In climatology, the projected bias-corrected data over a future period are usually compared to the original incorrect simulations as the main focus is to investigate how BC algorithms influence historical and projected climate data (Dieng et al., 2022); in contrast, from a hydrogeological impact assessment point of view, these are typically compared with historical observations (Pfeifer et al., 2015; Arampatzis et al., 2018; Sperna Weiland et al., 2021). Finally, the potential climate change in precipitation and temperature over the study area were qualitatively discussed.

### 5.1.1 Climate projections

The CORDEX (Coordinated Regional Climate Downscaling Experiment) is a global initiative to enhance regional climate change understanding and its impacts worldwide. It involves different continents, including Africa, Asia, Europe, the Americas, and the Middle East. Specifically, the EURO-CORDEX domain used in this research is the regional climate modelling initiative focusing on the European region. For this initiative, several GCMs, resulting from the CMIP5 project (Coupled Model Intercomparison Phase 5; Taylor et al., 2012) were dynamically downscaled and provided with a resolution of  $0.11^\circ$ . Despite CMIP6 having significantly advanced the assessment of climate change impacts by integrating RCP scenarios with Shared Socioeconomic Pathways (SSPs), i.e. climate change scenarios of projected socioeconomic global changes, its database has not yet made regional simulations publicly available. The progress in

producing such data shows that out of a total of 420 datasets, only 32 have been completed.

The selected ensemble of 12 RCMs from CMIP5 project is listed in Table 7, in which each model is renamed with an acronym comprising two letters referring to the GCM name and two to the RCM name. They have been commonly applied in climate change impact assessments in the Mediterranean region. Furthermore, this selection represents a sufficiently varied range of RCM and GCM combinations, as it encompasses 7 RCMs driven by 4 distinct GCMs.

Table 7 - List and acronyms of used GCM-RCM combinations.

<b>RCM</b>	<b>GCM</b>	<b>Model acronym</b>
ALADIN53 <sup>1R</sup>	CNRM-CERFACS-CNRM-CM5 <sup>1G</sup>	AL_CN
	CNRM-CERFACS-CNRM-CM5	CC_CN
CCLM4-8-17 <sup>2R</sup>	ICHEC-EC-EARTH <sup>2G</sup>	CC_IC
	MPI-M-MPI-ESM-LR <sup>3G</sup>	CC_MP
HIRHAM5 <sup>3R</sup>	ICHEC-EC-EARTH	HI_IC
RACMO22 <sup>4R</sup>	ICHEC-EC-EARTH	RA_IC
	CNRM-CERFACS-CNRM-CM5	RC_CN
RCA4 <sup>5R</sup>	ICHEC-EC-EARTH	RC_IC
	IPSL-IPSL-CM5A-MR <sup>4G</sup>	RC_IP
	MPI-M-MPI-ESM-LR	RC_MP
REMO2009 <sup>6R</sup>	MPI-M-MPI-ESM-LR	RE_MP
WRF331F <sup>7R</sup>	IPSL-IPSL-CM5A-MR	WF_IP

<sup>1R</sup> Colin et al., 2010; Herrmann et al., 2011; <sup>2R</sup> Rockel et al., 2008; <sup>3R</sup> Christensen et al., 1998; <sup>4R</sup> van Meijgaard et al., 2008; <sup>5R</sup> Samuelsson et al., 2011, Kupiainen et al., 2011; <sup>6R</sup> Jacob et al., 2012; <sup>7R</sup> Skamarock et al., 2008; <sup>1G</sup> Voltaire et al., 2013; <sup>2G</sup> Hazeleger et al., 2010; <sup>3G</sup> Giorgetta et al., 2013; <sup>4G</sup> Dufresne et al., 2013.

Only the RCP4.5 intermediate scenario for 1960-2005 and 2006-2100 was considered, as it refers to a stabilisation scenario assuming the invocation of climate policies to limit emissions and radiative forcing (Thomson et al., 2011). The more plausible outcomes are reflected by RCP4.5 considering the current (and, to a certain degree, pledged) policies (Hausfather & Peters, 2020). In contrast, RCP8.5 is a high emissions scenario, representative of the wide range of non-climate policy scenarios (van Vuuren et al., 2011).

### 5.1.2 Historical dataset

The Civil Protection Service of the Apulia Government provided daily historical climate observations, as described in 1.2.1. Specifically, 19 precipitation and 11 temperature stations were selected due to their consistent time series with less than 7% missing values. The reference period ranges from 1960 to 2005 (except SAL\_MS6 and SAL\_MS14, which date from 1971 to 2005).

Fig. 16 illustrates the spatial distribution of the selected meteorological stations in the study area. Before applying BC techniques, precipitation and temperature datasets were subjected to the iterative procedure of missing values infilling, as described in Chapter 2.

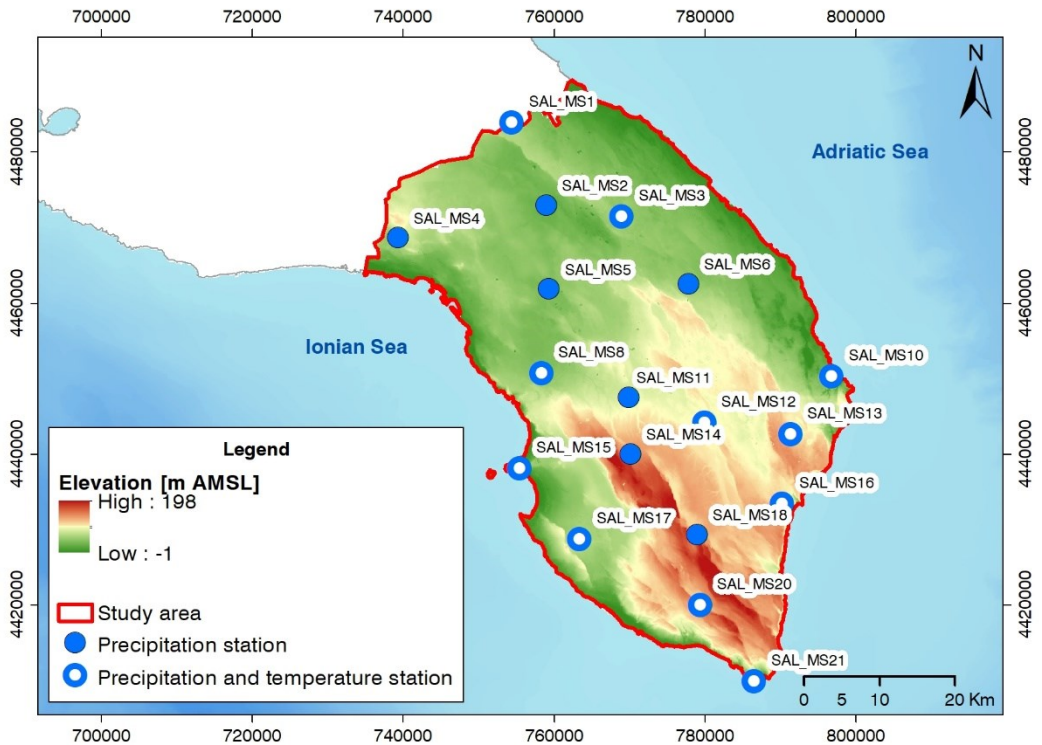


Fig. 16 - Meteorological station selected over the study area.



### 5.2 Bias-correction methods

Eight BC methods were applied at daily time steps, and their results were assessed on monthly and annual time steps, using the daily observations for the period 1971-2005 (Table 8).

The BC methods are generally distinguished into (i) scaling and (ii) distribution adjustment techniques. Examples of scaling methods among those selected in this study are the Linear Scaling method (LS), the Local Intensity scaling of precipitation (LOCI), the Power Transformation of Precipitation (PTR), and the Variance scaling of temperature (VAR). In contrast, Empirical Quantile Mapping (EQM), Parametric Quantile Mapping (PQM), Generalised Quantile Mapping (GPQM), Detrended Quantile Matching (DQM), and Quantile Delta Mapping (QDM) belong to the second category. While scaling methods account for the bias in the mean and/or variance, the quantile mapping BC methods commonly correct bias between simulated and observed data equating cumulative distribution functions (CDFs) of both datasets.

BC methods were applied using the well-developed and tested climate4R (Ilturbide et al., 2019) and downscaleR (Marsh et al., 2018) libraries for R software. The correction factor for daily data in all methods is based on a sliding window (Smitha et al., 2018), referring to 31 days. The sliding window approach consists of selecting a minimum of 100 rainy day values for each year in the reference period, ensuring enough data points for generating probability distribution plots. These chosen rainy day values were then utilised in various bias correction methods to compute the daily correction factors.

Table 8 - Lists and acronyms of the implemented BC methods.

<b>BC methods</b>	<b>Method acronym</b>	<b>Variables</b>
Empirical Quantile Mapping	EQM	Pr, $T_{max}$ , $T_{min}$
Linear Scaling	LS	Pr, $T_{max}$ , $T_{min}$
Parametric Quantile Mapping	PQM	Pr, $T_{max}$ , $T_{min}$
Generalised Quantile Mapping	GPQM	Pr, $T_{max}$ , $T_{min}$
Local intensity scaling	LOCI	Pr
Power transformation of precipitation	PTR	Pr
Detrended quantile matching	DQM	Pr, $T_{max}$ , $T_{min}$
Quantile delta mapping	QDM	Pr, $T_{max}$ , $T_{min}$
Variance scaling of temperature	VAR	$T_{max}$ , $T_{min}$

### 5.2.1 Linear Scaling (LS)

The LS method consists of scaling simulated data with a multiplicative or additive factor calculated as the difference/ratio between the observed and the simulated mean in the reference period (Lender & Buishand, 2007). The multiplicative factor is typically used with variables with a lower bound (e.g., precipitation or wind speed), while the additive is preferably applicable to unbounded variables (e.g., temperature):

$$P_{BC,m,d} = P_{raw,m,d} \times \frac{\bar{P}_{obs,m}}{\bar{P}_{raw,m}} \quad (42)$$

$$T_{BC,m,d} = T_{raw,m,d} \times (\bar{T}_{obs,m} - \bar{T}_{raw,m}) \quad (43)$$

Where  $P_{BC,m,d}$  and  $T_{BC,m,d}$  are the bias-corrected precipitation and temperature on the  $d$ th day of  $m^{\text{th}}$  month,  $P_{raw,m,d}$  and  $T_{raw,m,d}$  are the raw precipitation and temperature on the  $d$ th day of  $m^{\text{th}}$  month.  $\bar{P}_{obs,m}$  and  $\bar{T}_{obs,m}$  and  $\bar{P}_{raw,m}$  and  $\bar{T}_{raw,m}$  represent the mean values of observed and simulated precipitation and temperature at a given month  $m$ .

### 5.2.2 Local Intensity scaling of precipitation (LOCI)

The LOCI technique (Schmidli et al., 2006) is an improvement of the LS approach because it involves the adjustment of the mean as well as both wet day frequencies and intensities of rainfall time series, by setting a common precipitation threshold for all considered stations ( $P_{thres} = 0.20$  mm for the Salento case study) such that the number of simulated days exceeding this threshold matches the number of observed days. The bias-corrected daily precipitation values are calculated as follows:

$$P_{BC,m,d} = \begin{cases} 0, & \text{if } P_{raw,m,d} < P_{thres} \\ P_{raw,m,d} \times \frac{(\bar{P}_{obs,m} | P_{obs,m,d} > P_{thres})}{(\bar{P}_{raw,m} | P_{raw,m,d} > P_{thres})}, & \text{otherwise} \end{cases} \quad (44)$$

### 5.2.3 Power Transformation of Precipitation (PTR)

The PTR introduced by Leander & Buishand (2007) uses an exponential function permitting changes in the mean and the variance of the dataset. It is designed only for precipitation data due to the power function applied.

Firstly, the optimal exponent  $b_m$  for the  $m^{th}$  month is determined as the function that minimises Eq. 45. Then, the corrected precipitation series are obtained based on Eq. 46.

$$f(b_m) = \frac{\sigma(P_{obs,m})}{\bar{P}_{obs,m}} - \frac{\sigma(P_{raw,m}^{b_m})}{\bar{P}_{raw,m}^{b_m}} \tag{45}$$

$$P_{BC,m,d} = P_{raw,m,d} \times \frac{\bar{P}_{obs,m}}{\bar{P}_{raw,m}^{b_m}} \tag{46}$$

where  $\sigma(.)$  indicates the standard deviation operator.

### 5.2.4 Variance scaling of temperature (VAR)

The VAR is the complementary approach of PTR for temperature data. It adjusts the mean and variance of normally distributed variables (Chen et al., 2011a, Chen et al., 2011b). Temperature is typically corrected using VAR method with Eq. 47.

$$T_{BC,m,d} = (T_{raw,m,d} - \bar{T}_{raw,m}) \times \frac{\sigma(T_{obs,m})}{\sigma(T_{raw,m})} + \bar{T}_{obs,m} \tag{47}$$

### 5.2.5 Empirical Quantile Mapping (EQM)

The EQM method fits the CDF of simulated data to that of the observations, effectively correcting bias in the mean, standard deviation, and quantiles. This technique also adjusts the overestimation of wet or dry day frequency (defined as days with precipitation above or below  $P_{thres}$  in the observation dataset) following Themeßl et al. (2012) and Wilcke et al. (2013). The correction using EQM can be expressed in terms of the empirical cumulative distribution function (*ecdf*) and its inverse (*ecdf<sup>-1</sup>*):

$$P_{BC,m,d} = ecdf_{obs,m}^{-1}(ecdf_{raw,m}(P_{raw,m,d})) \quad (48)$$

$$T_{BC,m,d} = ecdf_{obs,m}^{-1}(ecdf_{raw,m}(T_{raw,m,d})) \quad (49)$$

### 5.2.6 Parametric Quantile Mapping (PQM)

The PQM algorithm uses a theoretical distribution calibrated over the training period. Usually, the Gamma distribution applies to precipitation (Piani et al., 2009), while the Gaussian distribution is appropriate for temperature data (Collados-Lara et al., 2018). As for EQM, the overestimation of wet or dry day frequency is assumed considering the above-cited  $P_{thres}$ .

$$P_{BC,m,d} = F_G^{-1}(F_G(P_{raw,m,d}|\alpha_{raw,m}, \beta_{raw,m})|\alpha_{obs,m}, \beta_{obs,m}) \quad (50)$$

$$T_{BC,m,d} = F_N^{-1}(F_N(T_{raw,m,d}|\mu_{raw,m}, \sigma_{raw,m})|\mu_{obs,m}, \sigma_{obs,m}) \quad (51)$$

where  $F_G(\cdot)$  and  $F_G^{-1}(\cdot)$  are the Gamma CDF and its inverse,  $F_N(\cdot)$  and  $F_N^{-1}(\cdot)$  are the Gaussian CDF and its inverse.  $\alpha_{raw,m}$  and  $\beta_{raw,m}$  are the fitted Gamma parameters for the raw precipitation at a given month  $m$ , and  $\alpha_{obs,m}$  and  $\beta_{obs,m}$  are those for observation.  $\mu_{raw,m}$  and  $\sigma_{raw,m}$  are the fitted Gaussian parameters for the raw precipitation at a given month  $m$ , and  $\mu_{obs,m}$  and  $\sigma_{obs,m}$  are those for observations.

### 5.2.7 Generalised Quantile Mapping (GPQM)

Gutjahr & Heinemann (2013) proposed the GPQM technique, which uses two theoretical distributions, i.e., the Gamma distribution to values under the threshold given by the 95<sup>th</sup> percentile and a general Pareto distribution to values above the threshold. This threshold is the 95<sup>th</sup> percentile of the observed and predicted wet-day distribution. For temperature data, the general Pareto distribution relates to values below the 5<sup>th</sup> percentile and the Normal distribution to the rest of the values. The wet-day frequency adjustment is also considered. This method aimed to adjust the extreme and non-extreme values separately.

### 5.2.8 Detrended Quantile Matching (DQM)

The DQM algorithm applies the EQM to the detrended series and then reapplies the mean trend to the bias-adjusted series. It preserves the long-term mean signal in a climate change context. More details and differences between QDM and DQM are explicitly provided by Cannon et al. (2015).

### 5.2.9 Quantile Delta Mapping (QDM)

The QDM method (Cannon et al. 2015) preserves the change signal in the simulated quantiles of variables considering a bias-corrected value term obtained from the observations dataset and the relative change term (delta) obtained from the simulated data. Therefore, model projections are firstly detrended per quantile, and quantile mapping is applied to correct systematic distributional biases compared to the observations. Secondly, the removed projected trends are restored to the bias-corrected quantiles.

### 5.3 BC Performance evaluation

The performance of the raw and bias-corrected climate model simulations was evaluated for each method on a monthly and annual basis by comparing bias-corrected simulated data with the corresponding precipitation and temperature historical observation from 1971 to 2005. Taylor diagrams (Taylor, 2001) were used to analyse and compare the results of each BC method since they incorporate correlation coefficient, root mean square difference and standard deviation into a single graph.

The adopted metrics are Root Mean Square error (RMSE), Spearman correlation coefficient (RHO), Nash-Sutcliffe efficiency (NSE), and Percent bias error (BIAS).

$$RMSE = \sqrt{\frac{\sum(S_i - O_i)^2}{N}} \quad (52)$$

$$RHO = \frac{cov(R(O)R(S))}{\sigma_{R(X)}\sigma_{R(Y)}} \quad (53)$$

$$NSE = 1 - \frac{(S - O)^2}{(O - \bar{O})^2} \quad (54)$$

$$PBIAS = 100 * \frac{1}{N} \sum (S_i - O) \quad (55)$$

where  $S_i$  indicates the  $i^{th}$  simulated raw data,  $O_i$  the corresponding  $i^{th}$  historical observation,  $R(O)$   $R(S)$  their ranks and  $N$  the length of the time series.

Choosing the most effective BC method is challenging because different statistical metrics may lead to inconsistent results (Gado et al., 2021). In this framework, the compromise programming proposed by Zeleny (1973) was applied to determine the most effective BC methods and evaluate the future scenario. It consists of measuring the distance of each method from the ideal value of the selected metrics and opting for the minimum one as the best method. The distance  $L_{cp}$  is estimated as follows:

$$L_{cp} = \left[ \sum_{j=1}^J |f_j - f^*_j|^p \right]^{\frac{1}{p}} \quad (56)$$

where  $J$  is the number of the metrics used,  $f_j$  is the normalised value of metric  $j$  obtained for a given method and  $f^*_j$  is the ideal value of the metric;  $p$  represents the maximal deviation and is equal to 1 for linear and 2 for Euclidean distance measure. In this study, a linear scale is used.

To evaluate the effect of the selected methods on future projections, the mean annual amount and the change signal were calculated for each RCM in the two 30-years periods from the observed reference. Specifically, the climate change signal refers to the percentage difference between future raw and bias-corrected RCMs output for precipitation, and deviations for temperature, from the reference period.

## 6. RESULTS AND DISCUSSION

### 6.1 *Missing values production*

Gaps or missing values within meteorological datasets continue to pose a recurring challenge, and their accurate filling presents a formidable task, especially when the available meteorological stations are limited compared to the size of the geographical regions under consideration (Aguilera et al., 2020).

In the Salento case study context, five different approaches were employed to impute missing values within the hydrological time series dataset. Their performance was assessed based on criteria such as the number of iterations required to reach predefined threshold values during two consecutive stages of data filling, as well as crucial metrics including R, NSE, and SI. After a preliminary analysis of the precipitation and temperature datasets to determine the extent of missing values, it was observed that the proportion of missing data points was relatively low, with precipitation reaching a maximum of 7% and temperature having a maximum of 5% missing values. Once selecting a period without missing data, the iterative procedure was tested for all methods and variables. The methodologies were applied to daily values in the reference period and monthly evaluated on the monthly records.

In the case of precipitation, characterised by its random nature, the estimated daily values exhibited variations compared to the observed ones, particularly on days with higher precipitation levels. However, when assessing the data on a monthly aggregation scale, the results were notably accurate, a conclusion further corroborated by the robustness of the acquired metrics. Fig. 17 illustrates the monthly comparison between the observed and estimated amount of precipitation and the maximum and minimum temperatures for each data interpolation method used during the reference period for the Lecce meteorological station, indicatively.



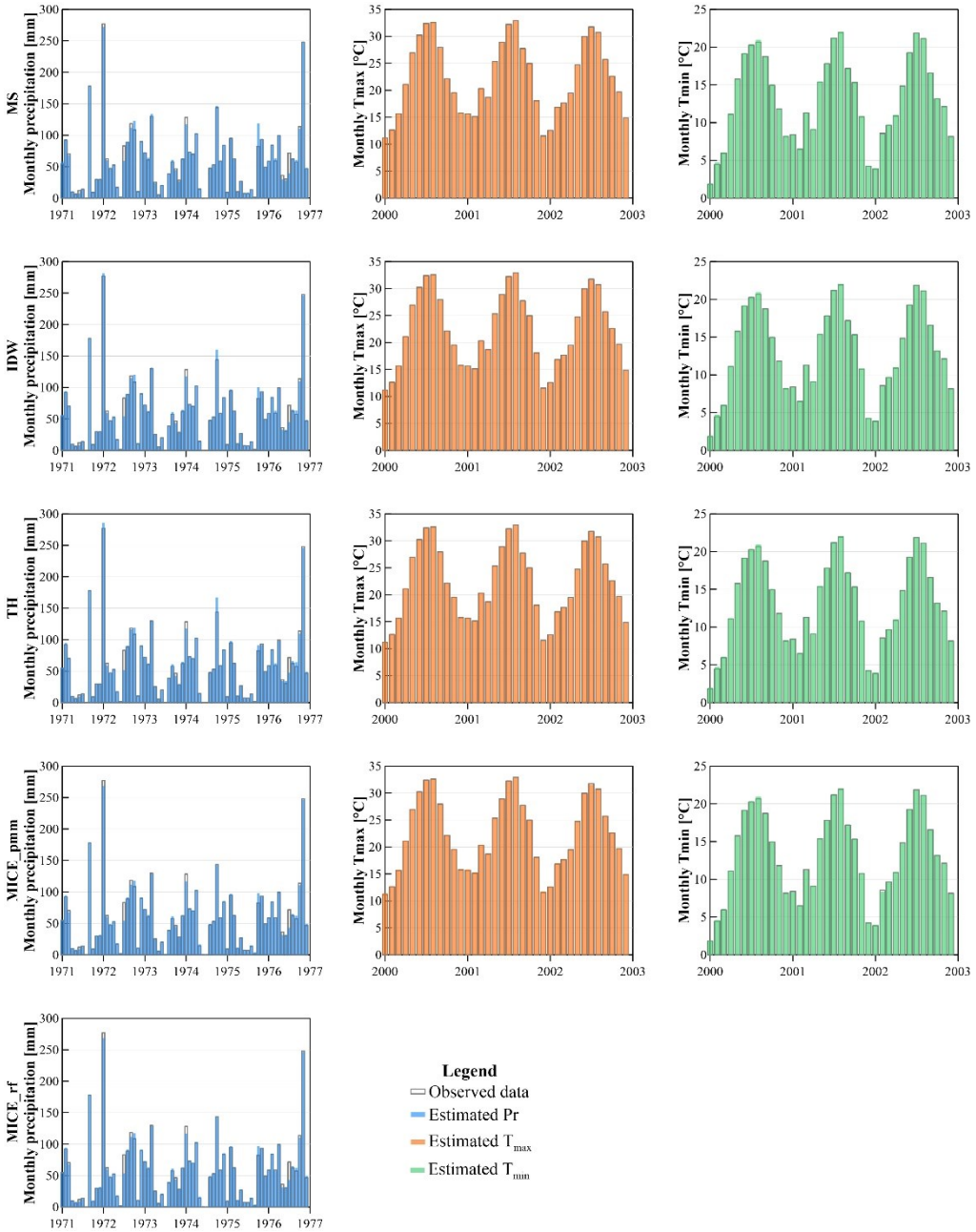


Fig. 17 - Monthly precipitation (left-hand side column), maximum (middle side column) and minimum (right-hand side column) temperature comparison between observed and estimated values obtained with the five selected infilling methods. The reference period for precipitation is 1971-1976, while for temperature is 2000-2002.

It is worth noting that all methods demonstrated satisfactory performance in filling in missing data, as evidenced by the substantial R and SI coefficients computed, in addition to the yielded NSE coefficients (Fig. 18). Furthermore, the number of iterations required to reach the convergence threshold of 0.01 mm for precipitation and 0.01°C for temperature was notably minimal, particularly for temperature. However, specific methods, such as Thiessen and IDW, in the case of rainfall, necessitated more steps to achieve convergence. Based on a comprehensive assessment that considered both the metrics and the number of iterations needed to reach convergence, the MICE\_pmm method was chosen for imputing missing data in the precipitation daily time series. In contrast, the MS method was selected for the temperature daily time series.

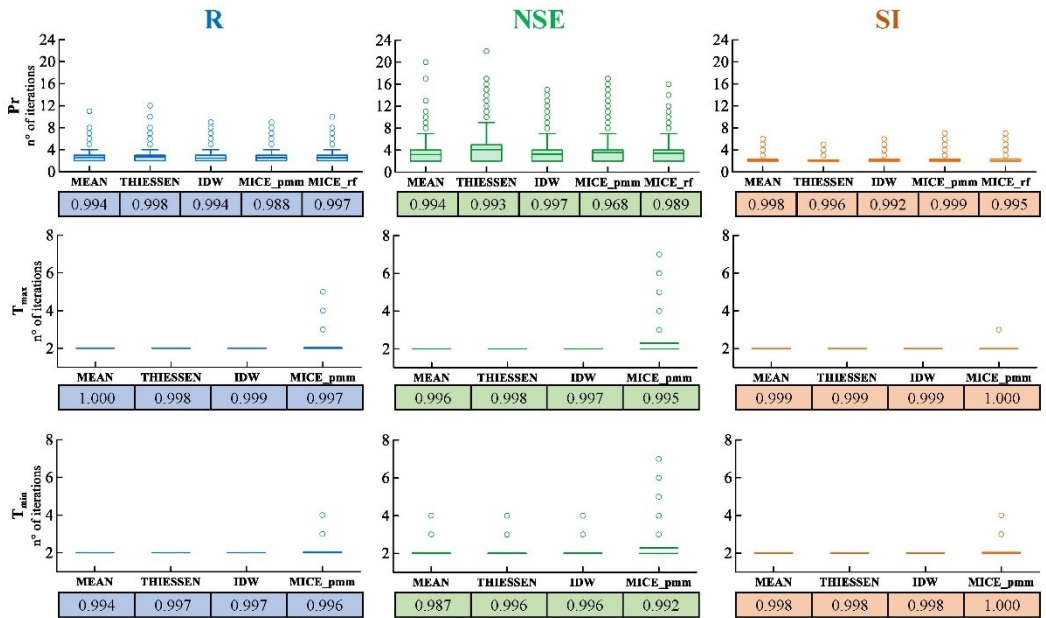


Fig. 18 - Number of iteration boxplots and final R, NSE, and SI coefficients for each filling method referring to Pr (top side), Tmax (middle side) and Tmin (bottom side), respectively.

The satisfactory outcomes observed can likely be related to the careful pre-screening employed when selecting the predictor stations. This process involved criteria such as a higher correlation coefficient, a reasonable distance, elevation difference, and shared proximity to the same sea compared to the station containing missing values.

Consequently, the criteria for selecting reference stations in this study can serve as a valuable reference for imputing precipitation and air temperature data in regions with similar hydroclimatic conditions.

In the scientific literature, MICE has proven to be a suitable method for addressing missing hydrological data. For instance, Turrado et al. (2014) obtained excellent results in estimating daily solar radiation in the Galicia catchment (Spain), outperforming other methods like Ordinary Kriging. Wesonga (2015) applied this technique to handle an incomplete time series of wind speed registered at the Entebbe International Airport (Uganda). Abdullah et al. (2020) conducted a study on extreme temperature and rainfall events in Bangladesh, using the MICE technique and implementing the predictive mean matching algorithm to fill in missing daily values with promising results. In the case of temperature, the low percentage of missing values in the initial dataset and the non-random nature allow less sophisticated methods to obtain reasonable results.

While recognizing the importance of assessing data-filling techniques in terms of their capability to capture average values over time and their proficiency in managing extreme events, it is worth noting that the analysis of extreme events has received limited attention. The primary emphasis was comprehending larger-scale future trends at the monthly and annual levels. Furthermore, it is essential to acknowledge that a substantial volume of data is typically necessary to analyse extreme events precisely. In this case, the timeframe used for evaluating the filling-in method performance (1971-1976 for precipitation and 2000-2002 for temperature) is insufficient for extreme events assessment.

## **6.2 Correlation between GWL and meteorological indicators**

The characterisation of the hydrodynamic mechanism of an aquifer system and the knowledge of groundwater response to climate change are paramount for strategic water resources planning. However, the absence of complete and enduring monitoring of groundwater levels prevents the application of some methodologies, which require long time series.

Analysing climate indices to describe the groundwater level variation is a possible approach under data scarcity. In a study by Balacco et al. (2022a), the response of the Salento aquifer to precipitation variability was evaluated through correlations between SPI and SPEI and groundwater levels (GWLs) for nine monitoring wells, covering the period from July 2007 to December 2011. Three different correlation factors were employed in this analysis. To provide a comprehensive climatic description of the study area and to obtain data for comparison within the specified time frame for GWL analyses, the two climate indices ranging from 6 to 48 months were initially calculated. This calculation was done using monthly precipitation data and monthly average air temperatures from 1951 to 2020. Subsequently, these indices were coupled with the nearest rainfall station (Table 9). Five rain gauge stations (Copertino, Galatina, Lecce, Maglie, and Ruffano) were identified as potential candidates based on the available monitoring wells. The effectiveness of this approach was subjected to further validation through an examination of the spatial distribution of correlation coefficients between GWLs and SPI and SPEI time series. This analysis revealed that, overall, most of the GWLs exhibited a positive correlation with the majority of meteorological stations in the Salento region.

Fig. 19 exclusively presents SPI and SPEI data for the Lecce station for brevity. Likewise, SPI and SPEI time series also show minor differences for the other stations. Notably, the severity of drought events reflected by the SPI tended to be more pronounced than that indicated by the SPEI (Pei et al., 2020). As the accumulation period for SPI and SPEI lengthened, drought events became more prolonged over time and occurred less frequently. In other words, dry events at shorter intervals in SPI and SPEI-6 months

tended to combine into a single extended event when transitioning to a higher accumulation period.

Table 9 - Wells and associated weather stations.

<b>Well</b>	<b>Weather stations</b>	<b>Available Monitoring Period</b>
LE_19/II S	Copertino	Jun 2007–Dec 2011
LE_12/IIIS	Lecce	Jun 2007–Dec 2011
LE_NC4	Copertino	Jun 2007–Dec 2011
LE_LS21LE	Ruffano	Oct 2007–Dec 2011
LE_PS24LE	Ruffano	Jun 2007–Dec 2011
LE_PS17LE	Galatina	Jun 2007–Dec 2011
LE_4/II S	Lecce	Jun 2007–Dec 2011
LE_1/LR	Copertino	Jun 2007–Dec 2011
LE_2/BS	Maglie	Jun 2007–Dec 2011

The time series data obtained from all the rain gauge stations corroborate the findings of Alfio et al. (2020), particularly concerning the occurrence of an extreme drought in the late 1990s and a series of moderate to severe droughts, with a few instances of extreme droughts, in the 2000s. Notably, when examining the SPI and SPEI data for Lecce, there is evidence of an extreme drought from 2016 to 2020. In contrast, the data from the other rain gauge stations reveal dry events ranging from moderate to severe during the same time frame.

These recent patterns of prolonged dry conditions are alarming, as they can potentially impact groundwater levels significantly. Such impacts can affect the quality and quantity characteristics of an aquifer, posing a substantial threat to its overall status. This calls for attention to the urgency of addressing and mitigating the consequences of these dry patterns on the regional groundwater resources.

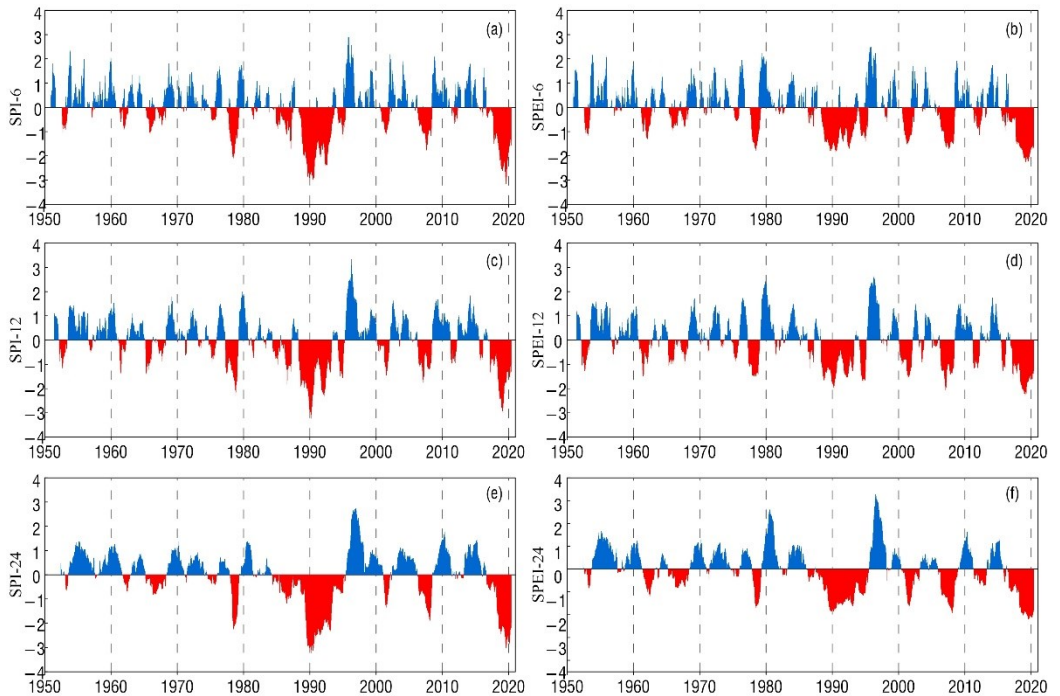


Fig. 19 - Temporal evolution of SPI and SPEI, respectively at the representative time scales of (a, b) 6-months, (c, d) 12-months, (e, f) 24-months for the Lecce rain gauge station from 1951 to 2020. The year marks are centred on the 1st of July of each year.

The relationships between SPIs/SPEIs from 6- to 48-months and GWLs were evaluated through Pearson's, Kendall's, and Spearman's correlation coefficients. Results highlight a statistically significant positive correlation with different time scales for each monitoring well. Fig. 20 and Fig. 21 illustrate the resulting correlation coefficient curves. Notably, all curves exhibit a consistent pattern for both SPI and SPEI: they initially increase until reaching the peak of correlation, after which they gradually decline to no longer statistically significant values. Pearson's and Spearman's coefficients show similar trends, whereas Kendall's coefficients are consistently lower. At any given time scale, the maximum correlation values calculated for the nine monitoring wells exceeded 0.60 for Pearson's and Spearman's methods. Some wells, such as LE\_LS21LE, LE\_4/IIS, and LE\_2/BS, reached a correlation peak of approximately 0.85 for both SPI and SPEI.

Furthermore, different responses were detected with respect to the time scales of SPI and SPEI. All wells showed the highest correlation between 16 to 23 months except for well LE\_NC4, which exhibited the correlation peak at nine months. All three methodologies adopted recognise these distinct response patterns. Additionally, no significant differences were found in the correlation between GWLs and SPEI compared to those obtained with SPI. In fact, the same range of maximum correlation and time responses could be identified for SPEI.

The observation of a long-time scale response in the Salento aquifer suggests that, despite its complexity, the aquifer tends to linearly react to variations in precipitation and temperature over the long term. This behaviour is analogous to that of a low-pass filter, characterised by significant transmissivity and storage capacity inertia, as discussed in earlier research (Molénat et al., 1999; Imagawa et al., 2013).

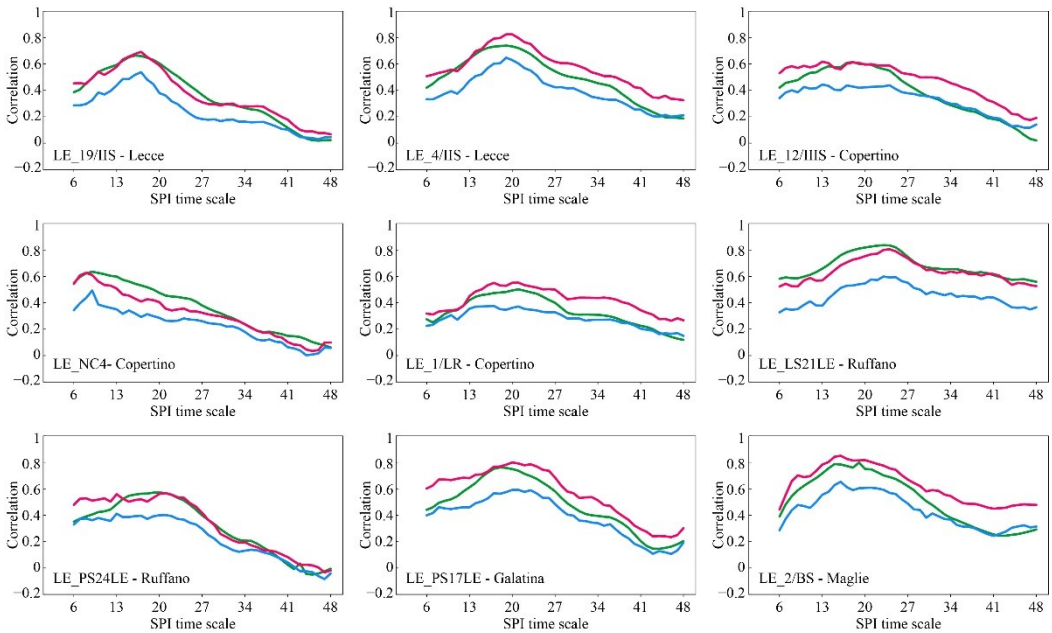


Fig. 20 - Correlation coefficients between GWLs and the corresponding SPI from 6- to 48-months calculated with Pearson's (green lines), Kendall's (blue lines), and Spearman's (purple lines) methods.

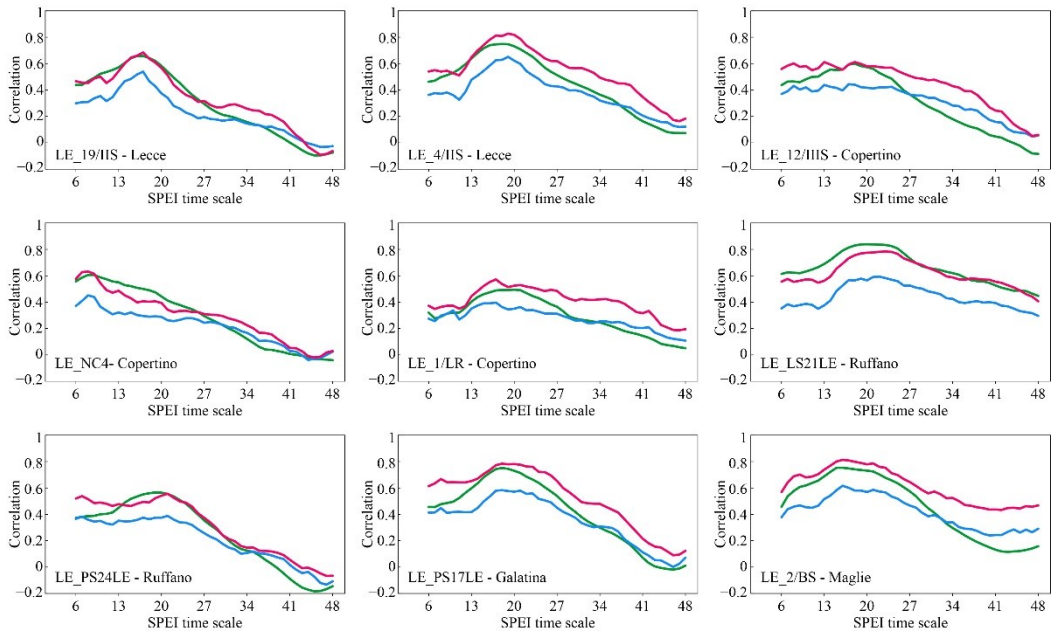


Fig. 21 - Correlation coefficients between GWLs and the corresponding SPEI from 6- to 48-months calculated with Pearson's (green lines), Kendall's (blue lines), and Spearman's (purple lines) methods.

Given the observed extended response between groundwater levels (GWLs) and climatic indices in most cases, a representative time scale of 18 months was chosen for analysis. The exception was LE\_NC4 well, with a more representative time scale of nine months. Consequently, scatter plots were generated to illustrate the relationships between GWLs and the corresponding SPI at the selected time scale (18 months for all monitoring wells, except nine months for LE-NC4), as shown in Fig. 22.

The linear regression models fitted to these datasets yielded relatively low R-squared values. This indicates that the data points are somewhat dispersed, particularly in cases where SPI values are positive. However, the trends in GWLs consistently align with dry periods, resulting in decreasing groundwater levels. This observation holds significance, especially from a management perspective, as the primary focus often revolves around predicting groundwater droughts. Forecasting the aquifer response to a period of meteorological drought is vital for identifying priorities in terms of water resource utilization and allocation.



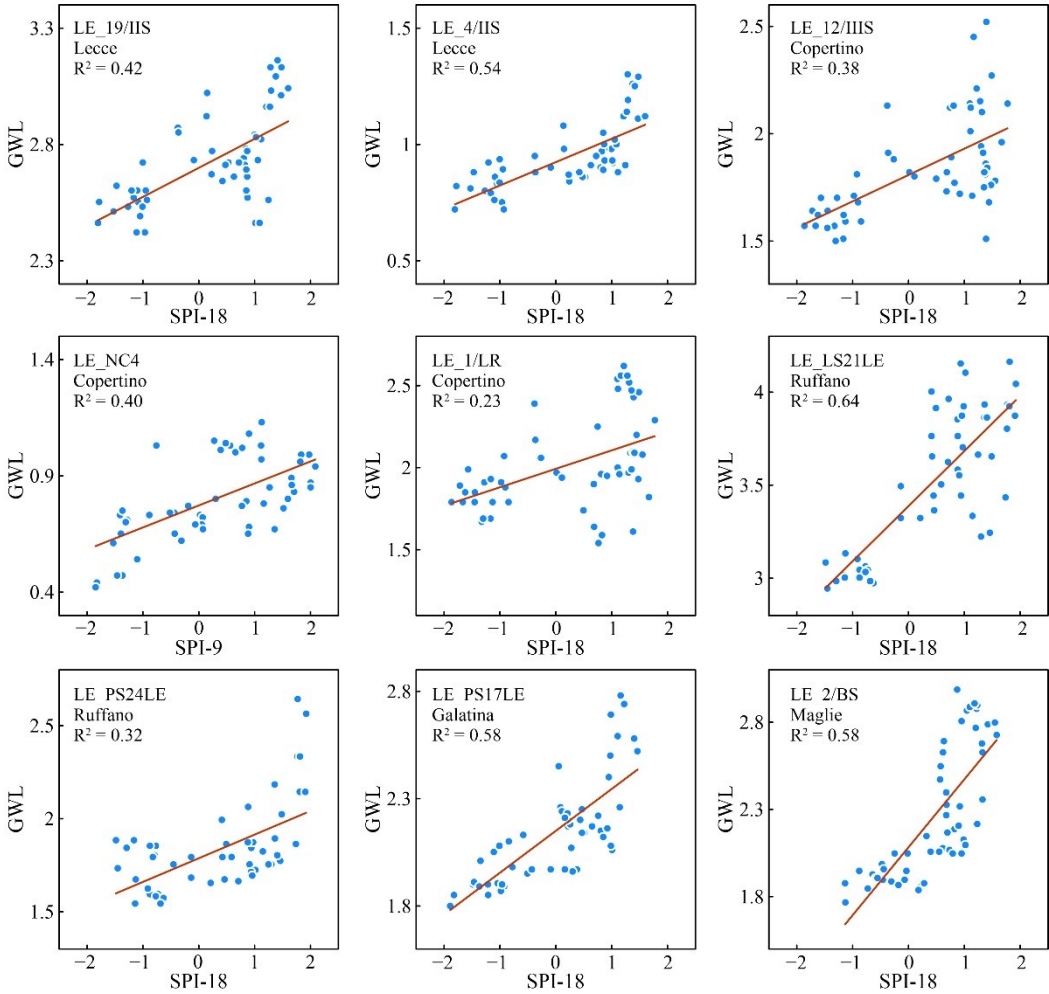


Fig. 22 - Scatterplots and R-squared between GWLs and the corresponding SPI for the more correlated time scale (18 months for all monitoring wells except 9-months for LE-NC4).

The study of the relationship between GWLs and meteorological indicators was improved with D’Oria et al. (2023) extending the dataset (the initial water level dataset was completed using sporadic data, mainly collected during spring and autumn, from 2013 to 2018) and introducing various time delays (or lags) between meteorological drivers (SPIs and SPEIs) and GWLs. The linear regression models were also applied considering a subset of the data collected during late winter-early spring (specifically in February, March, and April). The selection of this subset is motivated by the reduced impact of anthropogenic influences on groundwater conditions during this period.

Notably, irrigation withdrawals, which play a significant role in the area, are low during this time, thus providing a more reliable representation of natural groundwater dynamics. Having identified the time windows and lag times that exhibited the highest correlation between SPIs and groundwater levels (GWLs) or between SPEIs and GWLs, a linear regression model was constructed for each well. This approach assumes that the observed linear relationships detected in the historical data will remain consistent in the future. Based on these models, projections for GWLs were generated for the short-, medium-, and long-term. To accomplish this, SPIs and SPEIs were calculated using precipitation and temperature data sourced from an ensemble of 17 climate models as part of the EURO-CORDEX initiative (<https://www.eurocordex.net>, accessed on June 29, 2023). The data covered the period from 1976 to 2095 and were subjected to bias-correction at a monthly scale using a linear-scaling approach (as described by D’Oria et al., 2018). Two Representative Concentration Pathways were taken into consideration: the intermediate scenario (RCP4.5) and the most severe scenario (RCP8.5). The same correction factors that were employed during the historical period were applied to rectify bias in the scenario period.

Fig. 23 shows the results of the correlation analysis using SPIs and SPEIs, respectively. The findings highlight that the accumulation periods in meteorological indices showing the highest correlations are specific to each well. Considering a lag-time can enhance the correlation, leading to improved alignment between meteorological indices and groundwater levels.

Compared to Balacco et al. (2022a), analysing the subset generally improves the maximum correlations (excluding well LE\_LR). This can be attributed to the reduced influence of anthropogenic factors during these months (February-April), resulting in a more reliable representation of natural groundwater dynamics. Except for well LE\_LR, all wells demonstrate a maximum correlation occurring between 15 and 22 months, accompanied by Pearson correlation coefficients exceeding 0.7, indicating a strong correlation. The lag-times range from 0 to 3 months, indicating a delay between meteorological input and the aquifer response for some wells.

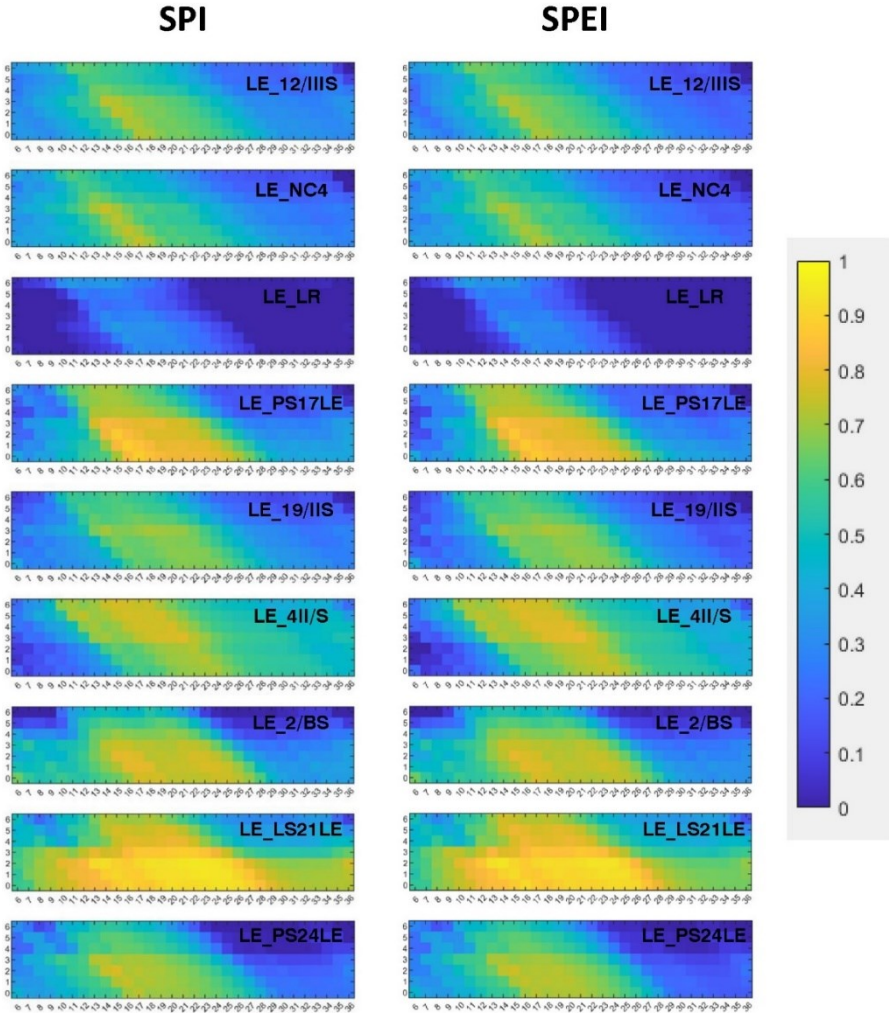


Fig. 23 - Heat maps of the Pearson correlation coefficients between GWLs and SPIs (left) and GWLs and SPEIs (right) for all wells and different time windows and lags using the subset of data collected during late winter-early spring.

Once the correlation between SPIs and GWLs and between SPEIs and GWLs was established, linear regression models were identified between groundwater levels recorded at the eight (LE\_LR was excluded due to its low correlation coefficient) remaining monitoring wells and the meteorological indices.

Table 10 and Table 11 list the obtained correlation coefficients, the lag at which they were obtained and the resulting slopes and intercepts of the regression lines. These

linear relationships serve as forecasting models used to assess the impacts of climate change on the GWLs of the Salento aquifer.

Table 10 - Wells, associated meteorological gauging stations, maximum correlation coefficients between February-April GWLs and SPIs and their accumulation periods and lags in months. Slopes and intercepts (m a.s.l.) of the linear regressions between GWLs and SPIs.

Well	Gauge	Correl.	SPI	Lag	Slope	Interc.
LE_12IIS		0.74	15	2	0.18	1.84
LE_NC4	Copertino	0.75	14	3	0.10	0.79
LE_LR		0.36	11	6	0.07	2.07
LE_PS17LE	Galatina	0.89	17	0	0.26	2.18
LE_19IIS	Lecce	0.71	18	3	0.13	2.74
LE_4IIS		0.78	21	3	0.12	0.93
LE_2BS	Maglie	0.80	15	2	0.37	2.14
LE_LS21LE	Ruffano	0.95	22	2	0.35	3.53
LE_PS24LE		0.77	15	2	0.19	1.75

Table 11 - Wells, associated meteorological gauging stations, maximum correlation coefficients between February-April GWLs and SPEIs and their accumulation periods and lags in months. Slopes and intercepts (m a.s.l.) of the linear regressions between GWLs and SPEIs.

Well	Gauge	Correl.	SPEI	Lag	Slope	Interc.
LE_12IIS		0.72	15	2	0.16	1.88
LE_NC4	Copertino	0.72	14	3	0.09	0.81
LE_LR		0.36	11	6	0.07	2.09
LE_PS17LE	Galatina	0.88	17	0	0.23	2.26
LE_19IIS	Lecce	0.72	18	3	0.12	2.76
LE_4IIS		0.80	21	3	0.11	0.95
LE_2BS	Maglie	0.80	15	2	0.33	2.27
LE_LS21LE	Ruffano	0.94	14	2	0.44	3.57
LE_PS24LE		0.75	15	2	0.17	1.82

Fig. 24 depicts the results for the eight wells in 1986-2095 according to the SPI-GWL relationships. GWLs are averaged in February-April and represented in terms of GWL

anomalies relative to the average GWLs in the reference period, inferred from the RCMs. According to the RCM ensemble median and emission pathways, the analysis indicates that no evident changes in the GWLs are expected in the future for all wells; however, the inter-model variability between the 17 RCMs is different for the eight wells. Wells LE\_PS17LE, LE\_2BS and LE\_LS21LE have the highest uncertainty; the remaining wells show less variability among the RCMs. The highest mean gradient is expected for the LE\_LS21LE well (-3.0 cm/decade).

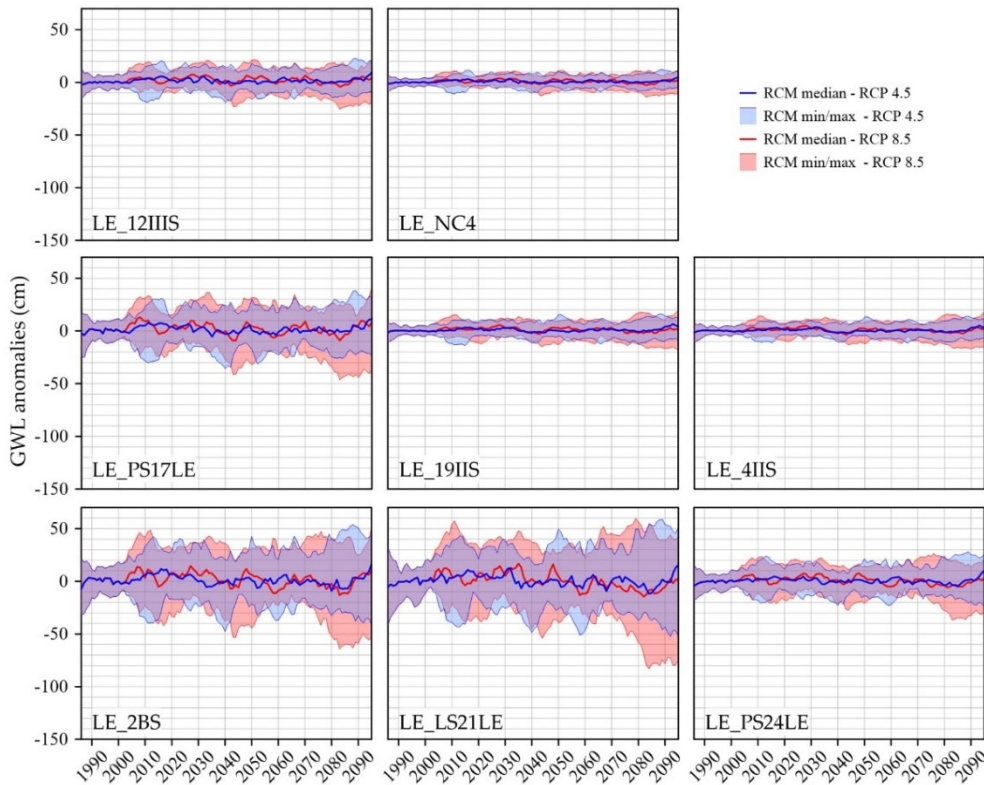


Fig. 24 - Groundwater level anomalies, relative to 1986-2095, in terms of 10-year moving average for the eight wells according to the SPI-GWL relationships and the RCP4.5 and RCP8.5 scenarios. The GWLs are averaged in February-April and represent the late winter-early spring period.

The future GWL anomalies evaluated using the SPEI-GWL relationships are shown in Fig. 25. A general decline in groundwater levels is expected for all wells, which is more severe for the RCP8.5 scenario; the inter-model variability is comparable to that obtained with the SPI analysis.

Under the RCP4.5 scenario, most RCMs indicate significant negative trends in GWLs, with only a few exceptions. However, when considering the more severe RCP8.5 scenario, all RCMs consistently project substantial declines in GWLs across all monitoring wells. Notably, the mean trend gradients calculated for RCP8.5 are higher than those obtained based on SPEI, while the standard deviations of these trend gradients are comparable. The LE\_LS21LE well is expected to experience the most pronounced decrease in GWLs in both scenarios. Specifically, under the RCP4.5 scenario, it is projected to have a change rate of -3.5 cm per decade, while under the more severe RCP8.5 scenario, the projected change rate is even more substantial at -11.6 cm per decade.

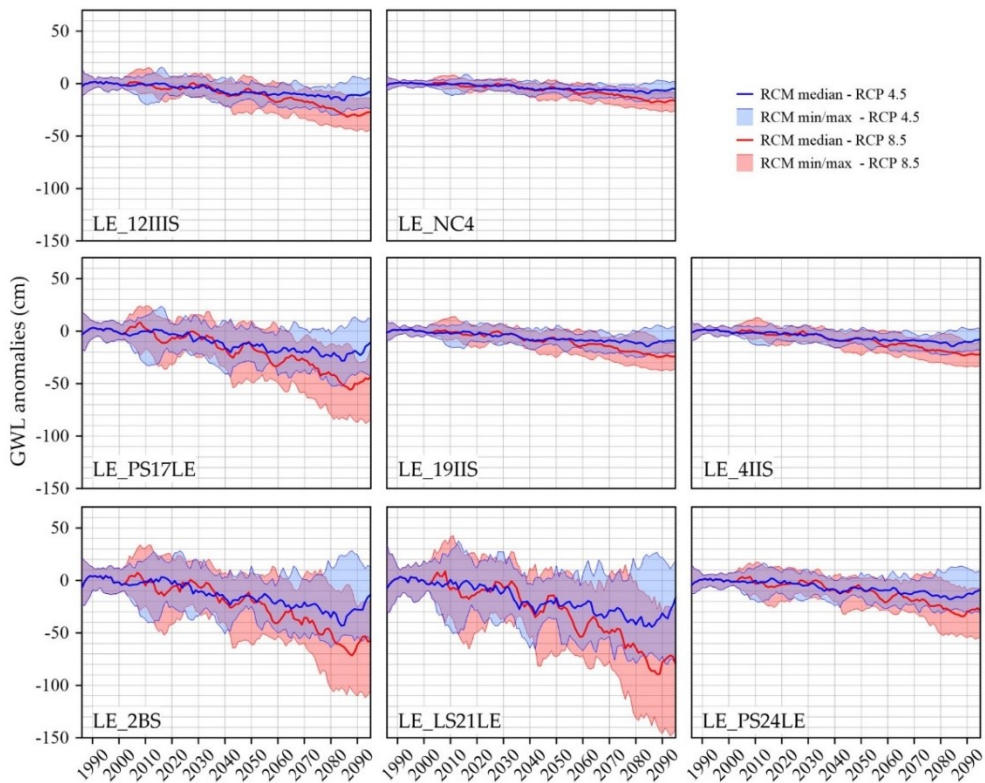


Fig. 25 - Groundwater level anomalies, relative to 1986-2095, in terms of 10-year moving average for the eight wells according to the SPEI-GWL relationships and the RCP4.5 and RCP8.5 scenarios. The GWLs are averaged in February-April and, therefore, represent the late winter-early spring period.

The response of the groundwater system to the climate signal, as inferred from the extended time windows of SPIs and SPEIs and the lag-time that exhibits the highest correlations with GWLs, suggests that the climate variables within the delay period have limited influence on the aquifer response. This results in a delayed reaction of the aquifer to changes in climate conditions. This improvement validates the preliminary results presented by Balacco et al. (2022b), which concluded that the Salento aquifer exhibits characteristics akin to a low-pass filter in its response to meteorological events and is marked by great inertia and storage capacity.

The responsiveness of the groundwater system depends on various factors, primarily localised conditions such as geology, surface morphology, unsaturated zone permeability, precipitation patterns, the size of the aquifer under investigation, water withdrawals, and regional flow patterns. These factors collectively influence how meteorological droughts propagate to affect groundwater levels (Kumar et al., 2016; Uddameri et al., 2019), resulting in a time lag between precipitation events and fluctuations in groundwater. This time lag can impact the linear correlation between these variables, explaining the dispersion observed in data points around the linear regression lines relating SPIs/SPEIs to GWLs. However, the dispersion is lower for negative SPIs and SPEIs concerning the positive ones, indicating that the regression model is more reliable during dry periods. This reinforces the credibility of the proposed methodology, as one of its key objectives is to enhance the ability to predict groundwater droughts, a critical aspect of adequate water resources management. The variations observed among wells in terms of accumulation periods for meteorological indicators, lag times, and correlation coefficients underscore the complex nature of the studied aquifer, characterised by anisotropy and heterogeneity.

### 6.3 Analyses of GWL and precipitation in time and frequency domain

Based on the results achieved with the previous applications, further statistical techniques have been applied to deepen the hydrodynamic characterisation of the Salento aquifer. In detail, this section presents the outcomes of time series analyses in the time and frequency domain investigating daily groundwater levels and precipitation time series separately and combined.

Specifically, the analysis used daily precipitation as input and GWL as output, considering eleven monitoring wells for a minimum of two to four hydrological years, depending on the well. Based on the available hydrological years, the GWL dataset was divided into Group 1 (4 years), Group 2 (3 years) and Group 3 (two years). In the time domain, ACF and CCF were calculated at all wells and associated rain gauge stations, fixing a step of one day and a truncation point at 125 days for each series (Mangin, 1984).

The ACF of daily precipitation at each rain gauge station exhibits a rapid decline within two days, as characteristically illustrated for Ruffano station (Fig. 26). It remains in a defined confidence interval (-0.05 to +0.05) for all stations and lacks oscillation due to the varying timing of minimum and maximum rainfall occurrences. Despite the presence of a seasonal pattern in precipitation, this behaviour confirms that precipitation events are uncorrelated random processes (Angelini, 1997; Panagopoulos & Lambrakis, 2006).

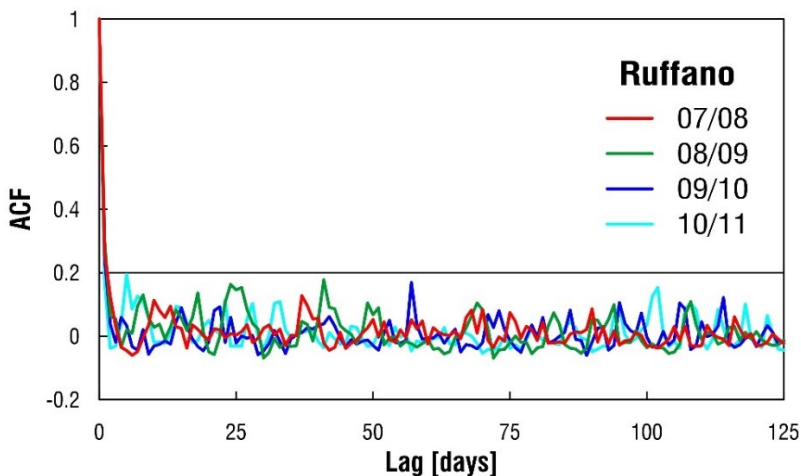


Fig. 26 - Autocorrelation function of Ruffano precipitation time series for each considered hydrological year.



In contrast, the ACFs of GWLs exhibit a gradual decrease and indicate a different response of certain wells when comparing a hydrological year with lower rainfall to a much wetter one, such as 2008/09, which exceeds the average annual precipitation (Fig. 27). The ACFs of most wells in different hydrological years show negative values, indicating a yearly cycle that varies from 30 to 125 days (Larocque et al., 1998).

Table 12 summarises the memory effect for each hydrological year and well, where the assessment of time lag corresponds to an ACF value of 0.2, as Mangin (1984) outlined. Based on the long memory effect detected, baseflow can be considered the primary water transfer process of the Salento aquifer due to the limited connectivity between fissures and fractures within the rock matrix. However, when precipitation exceeds the average of 638 mm/year (Portoghese et al., 2013), the correlogram initially displays a steep slope, indicating the rapid flow through karst conduits and large fractures, while the baseflow process becomes apparent only later, favoured by minor fractures and discontinuities that drain the unsaturated zone (Panagopoulos & Lambrakis, 2006). In contrast, for certain wells (LE\_NC4 and LE\_2/BS), the baseflow is the sole water transfer process, regardless of rainfall intensity and frequency. This observation suggests that certain parts of the aquifer exhibit a much higher level of inertness, indicating limited responsiveness to changes in precipitation.

CCF between precipitation and GWL for each well and hydrological year (Fig. 28) is dissimilar case-by-case, with variable amplification and response time (Table 12). The degree of correlation is generally low, with values never over 0.24. Overall, the CCFs decrease very slowly, confirming the inertial behaviour of the Salento aquifer. Some wells (LE\_P1TAU, LE\_4/IIS, and LE\_1/LR) show a negative response time, indicating that precipitation does not directly influence GWL (Cai & Ofterdinger, 2016).

During the driest year (2007-08), wells of Group 1 show a fast response time ranging from 1 to 14 days, except for LE\_NC4 (56 days). During the wettest hydrological year (2008-09), the wells of Groups 1 and 2 present different behaviour in terms of response time. LE\_LS21LE and LE\_P1TAU wells exhibit a very fast response time compared to other wells, equal to 3 and 28 days, respectively; during the same year, LE\_19/IIS, LE\_12/IIIS, and LE\_SG3 highlight a time lag of about 56-57 days, while LE\_NC4,

LE\_PS24LE, and LE\_PS17LE have a response time lag of 71 days. For 2009-10 and 2010-11, closer to the average hydrological year described by Portoghese et al. (2013), the response time varies depending on the well and hydrological year considered.

Wells of Group 3 exhibit different behaviour: LE\_4/IIS and LE\_1/LR show a response time of 28 days for 2009-10, the wettest of the two considered hydrological years, and 103 and 113 days, respectively, for the second year. On the contrary, LE\_2/BS displays an opposite response, i.e., 69 days for 2009-10 and 8 days for 2010-11. This faster response time during a drier year could be associated with a concentrated period of rainfall that occurred from October to November 2010, i.e., at the beginning of the 2010-11 hydrological year. Furthermore, it reveals a bimodal hydrodynamic behaviour (Panagopoulos & Lambrakis, 2006), where certain wells exhibit a series of peaks following the initial peak. This phenomenon aligns with the observations of Padilla & Pulido-Bosch (1995) in attributing it to additional flow components entering the aquifer after the initial peak.

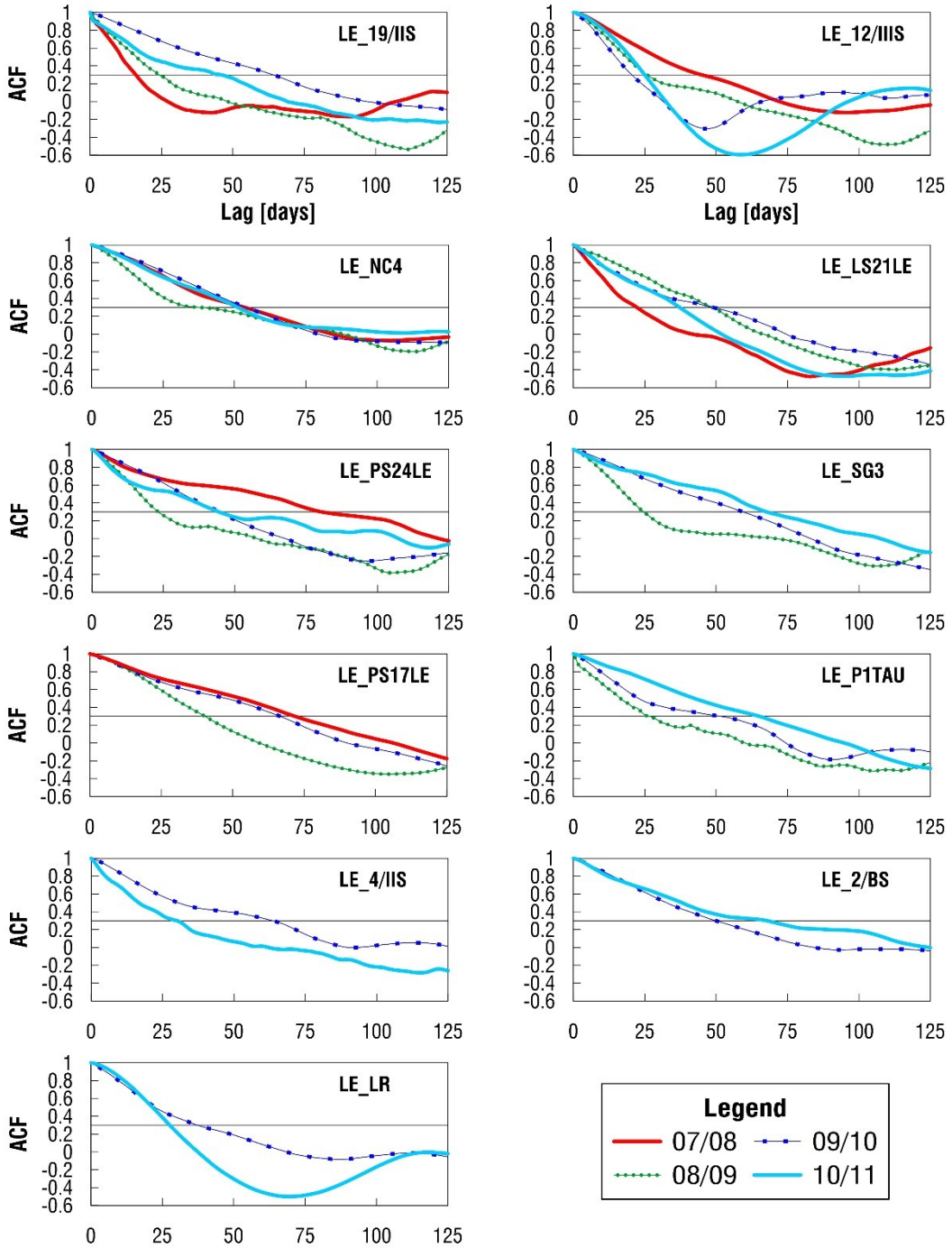


Fig. 27 - Autocorrelation functions of the short-term analyses conducted for each well.

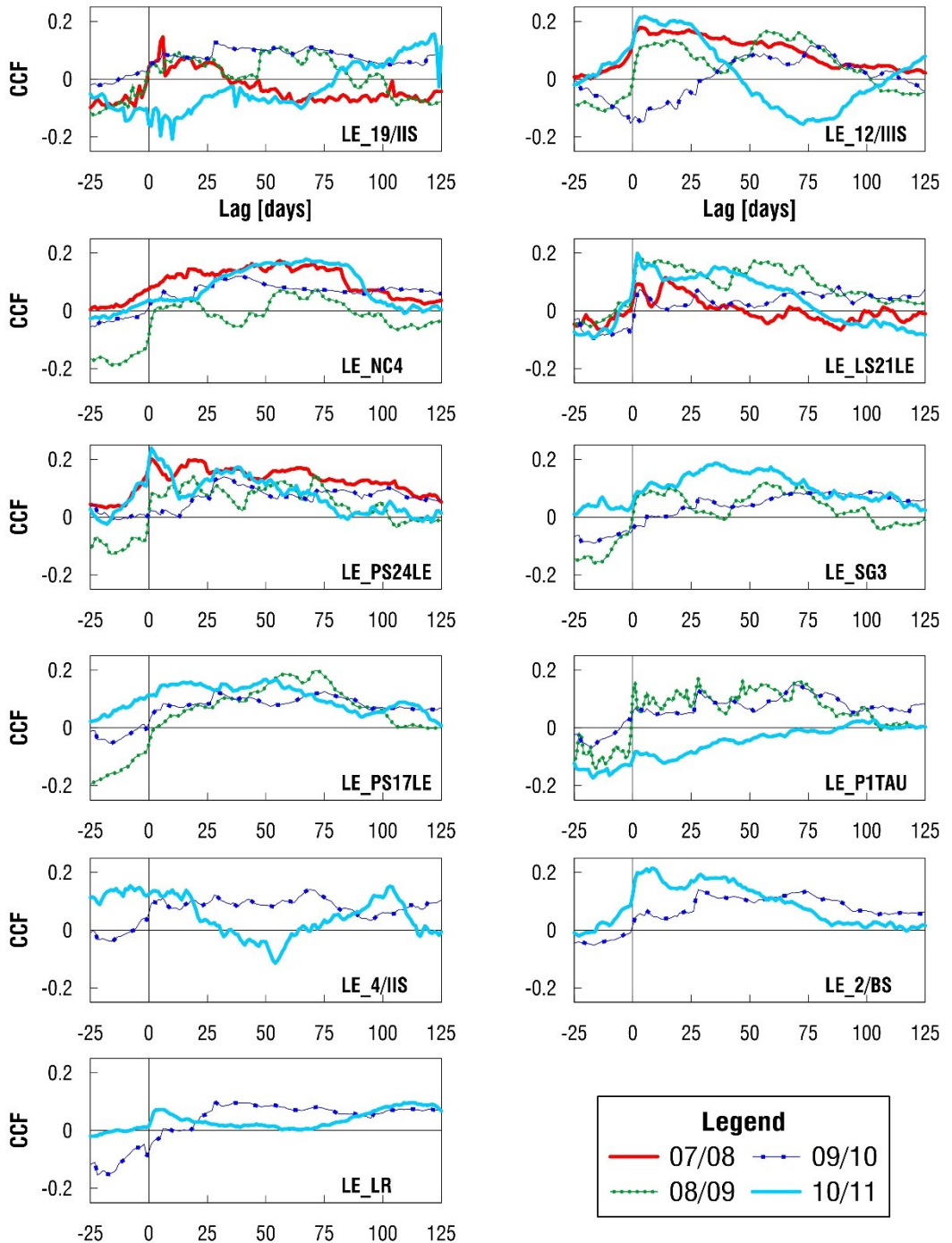


Fig. 28 - Cross-correlation functions of the short-term analyses conducted for each well.

The analyses of ACF and CCF were finally repeated on the entire dataset to investigate the long-term response of precipitation and GWLs (Fig. 29). The outcomes of the ACF reveal that in most cases, the reaction of GWLs in terms of memory effect is longer than 100 days, with few exceptions (Fig. 29a). The common behaviour of the long-term ACFs outlines the significant role of the rock matrix in slowly releasing infiltration water, so providing a large buffering capacity. As to the long-term CCF (Fig. 29b), the absence of statistically significant correlation between precipitation and groundwater levels related to LE\_SG3 and LE\_1/LR, and the low cross-correlation detected for the remaining wells, suggest that the Salento aquifer, contrary to most karst aquifers where exploitation occurs by the uptake of springs, may be influenced by additional factors, as the heavy exploitation of groundwater by wells.

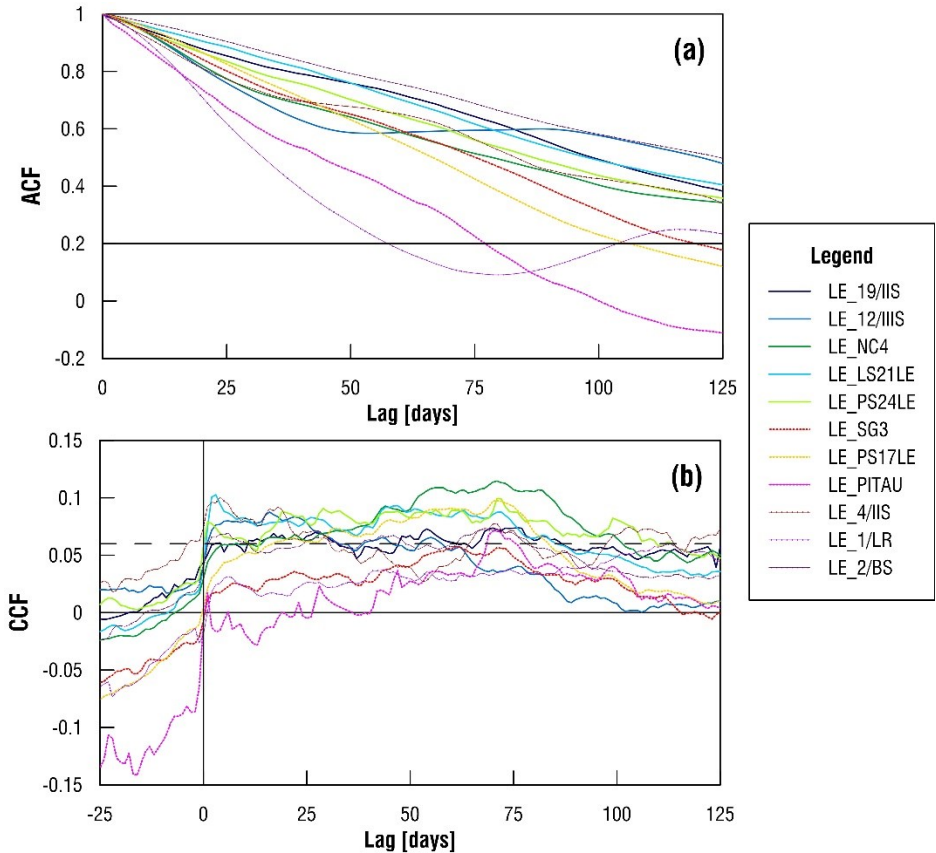


Fig. 29 - a) Autocorrelation and (b) Cross-correlation functions of the long-term analyses conducted for each well on the entire available time series.

Considering the frequency domain, the spectral density, the cross amplitude, the phase, the coherence, and the gain functions were determined for each well and hydrological year. The Spectral Density Function highlights aspects of the output signal that ACF cannot clearly distinguish. Moreover, it allows the calculation of the regulation time, which ranges from 50 to 100 days, emphasising the noticeable regulation capacity of the system. For the hydrological year 2007-08, which is the driest among those considered, the CAF reaches a maximum of 1. In contrast, the remaining studied years display highly variable amplitudes more significant than 1. However, these high amplitudes occur at frequencies ranging from 0.05 to 0.1, corresponding to the extended period identified in the time domain. The comparison of CAF with precipitation spectral density demonstrates the role of the studied aquifer as a low-pass filter since the prominent peaks of precipitation signal reflect in those of GWL at the same frequencies but with a considerably lower amplitude. Fig. 30 presents a comparison between the CAFs of the LE\_P1TAU well during the wettest (Fig. 30a) and driest (Fig. 30b) hydrological years, alongside the corresponding spectral density of precipitation from the Ruffano rain gauge station.

The results of PHF, COF and GAF are not reported for brevity. All of them confirm the strong influence of the hydrological year on the aquifer response, showing how the input signal is much attenuated when compared with GWLs. They ensure the role of the rock matrix in damping the input signal, causing poor correlation and coherency between the two analysed time series.

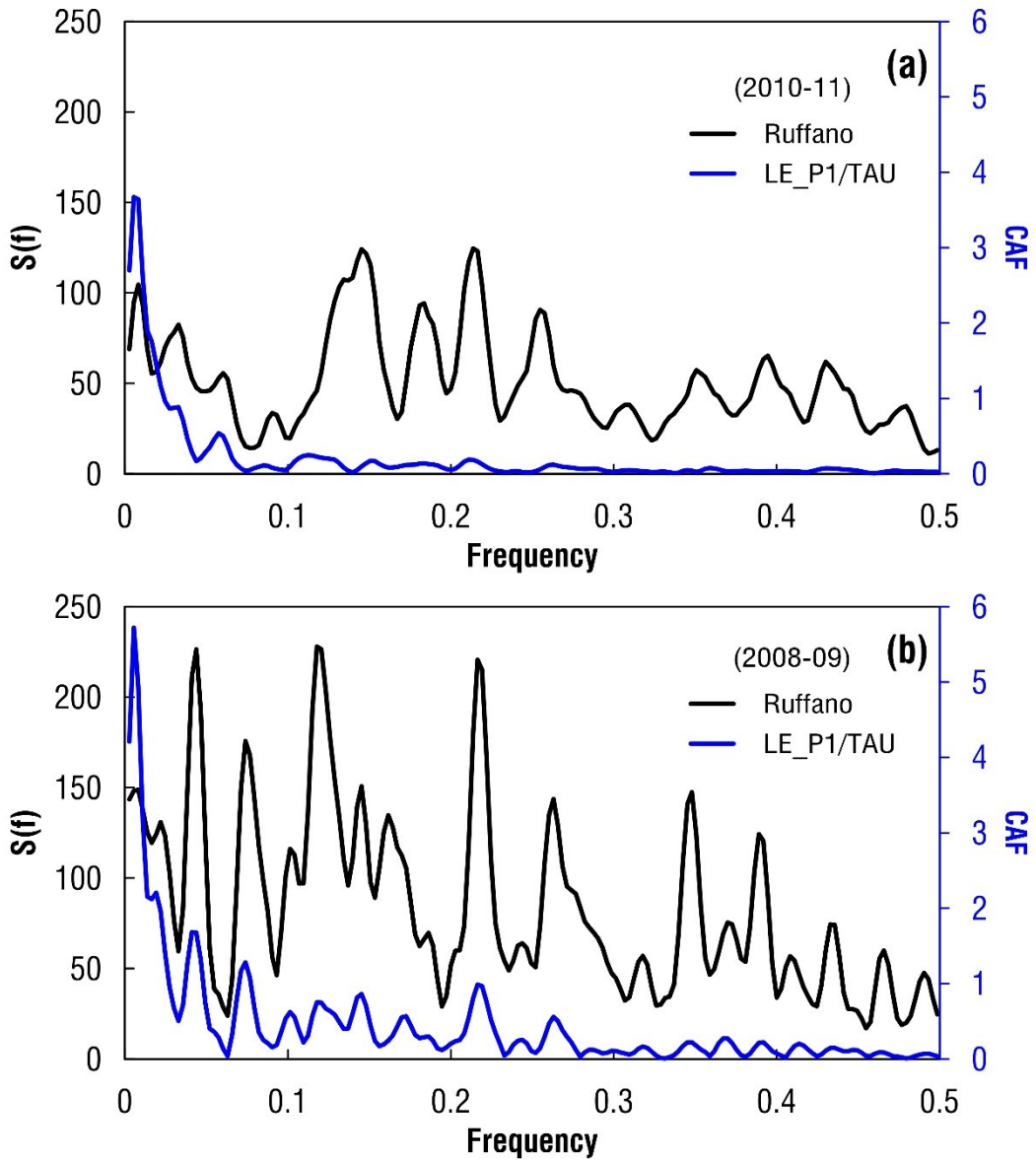


Fig. 30 - Cross Amplitude and precipitation spectral density functions for Ruffano station and LE\_P1TAU of (a) 2008-09 and (b) 2010-11 hydrological cycle.

Table 12 - Calculated hydrological parameters for the Salento aquifer.

<b>Group 1</b>	<b>Year</b>	<b>Memory Effect [days]</b>	<b>Regulation Time [days]</b>	<b>CCF coeff [-]</b>	<b>Response Time [days]</b>	<b>Mean Delay [days]</b>
LE_19/IIS	2007/08	19	50	0.146	6	ND
	2008/09	26	89	0.113	57	5.7
	2009/10	72	150	0.127	29	1.3
	2010/11	54	143	0.156	122	ND
LE_12/IIIS	2007/08	56	65	0.178	3	0.8
	2008/09	32	118	0.165	57	0.9
	2009/10	24	76	0.117	81	ND
	2010/11	27	72	0.218	5	2.2
LE_NC4	2007/08	64	79	0.173	56	ND
	2008/09	56	69	0.08	71	0.4
	2009/10	62	133	0.120	40	2.4
	2010/11	58	87	0.179	67	ND
LE_LS21LE	2007/08	27	78	0.115	14	6.8
	2008/09	54	143	0.192	3	0.8
	2009/10	60	148	0.082	88	1.8
	2010/11	42	111	0.199	2	1.3
LE_PS24LE	2007/08	104	42	0.202	1	1.3
	2008/09	28	123	0.147	71	ND
	2009/10	51	138	0.140	32	ND
	2010/11	70	60	0.239	1	ND
<b>Group 2</b>	<b>Year</b>	<b>Memory Effect [days]</b>	<b>Regulation Time [days]</b>	<b>CCF coeff [-]</b>	<b>Response Time [days]</b>	<b>Mean Delay [days]</b>
LE_SG3	2008/09	29	111	0.120	56	0.6
	2009/10	68	150	0.087	103	1.5
	2010/11	80	85	0.187	35	ND
LE_PS17LE	2008/09	46	133	0.199	71	0.6
	2009/10	74	151	0.129	28	1.9
	2010/11	82	147	0.168	50	ND
LE_P1TAU	2008/09	34	132	0.171	28	ND
	2009/10	63	145	0.149	69	8.4
	2010/11	75	113	-0.173	17	ND
<b>Group 3</b>	<b>Year</b>	<b>Memory Effect [days]</b>	<b>Regulation Time [days]</b>	<b>CCF coeff [-]</b>	<b>Response Time [days]</b>	<b>Mean Delay [days]</b>
LE_4/IIS	2009/10	72	136	0.140	28	1.8
	2010/11	35	120	0.152	103	ND
LE_1/LR	2009/10	49	133	0.101	28	ND
	2010/11	31	79	0.097	113	2.9
LE_2/BS	2009/10	60	144	0.140	69	0.3
	2010/11	89	55	0.214	8	ND



#### **6.4 *GWL and precipitation decomposition***

Groundwater levels, differently from precipitation, are classified as additive time series, wherein fluctuations are influenced by multiple drivers (Shapoori et al., 2015). Seasonal and trend decomposition analysis is a valuable approach to isolate each signal component, such as seasonal or inter-annual variations, long-term trends, and random factors, allowing an appropriate conceptual understanding of an aquifer system. This technique was applied to the eleven selected monitoring wells at a daily scale to investigate the influence of seasonal, trend and noise components on GWL.

The visual classification of the 11 boreholes based on the shape and amplitude of the STL results led to the identification of three groups (Fig. 31). In the first Group (Fig. 31a), three peaks can be detected for each available hydrological year, concentrated in wet period (mainly from December to April) except for LE\_2/BS and LE\_P1TAU wells which reach June. The dry period, instead, corresponds to September, except for LE\_NC4 (August) and LE\_P1TAU (October). The seasonal fluctuations are limited to around 0.20 m, while the trend component is predominantly negative, displaying an increasing pattern during 2009-10 hydrological year. The second Group exhibits a trend pattern like the previous case for all wells. Regarding the seasonal component, seasonal fluctuations are also limited but higher, approximately 0.50 m (Fig. 31d). Two peaks were identified for these wells, principally in February and from May to July; the dry period corresponds to September. The third Group is characterised by one seasonal peak in February (LE\_PS17LE) and March (LE\_LS21LE), while the dry period concentrates in September (Fig. 31g). The ranges of the seasonal fluctuations are more similar to Group 2. The trend pattern agrees with the previous results.

In all groups, the seasonal and the remainder components have a less significant influence on the GWL signals compared to the trend ones, as demonstrated by the variance ratio values obtained following the methodology of Lafare et al. (2016) and Colyer et al. (2022) reported in Table 13. LE\_LR and LE\_P1TAU are the unique wells for which the variance ratio is equally distributed in all three components, preventing the identification of a prevailing pattern.

Table 13 - Statistical summary of STL results for Salento boreholes.

<b>Well</b>	-	<b>Observation</b>	<b>Seasonal</b>	<b>Trend</b>	<b>Reminder</b>
LE_19IIS	var.	0.036	0.005	0.022	0.004
	var. ratio	-	0.14	<b>0.60</b>	0.12
LE_12IIS	var.	0.058	0.003	0.034	0.012
	var. ratio	-	0.05	<b>0.59</b>	0.22
LE_NC4	var.	0.031	0.005	0.017	0.005
	var. ratio	-	0.15	<b>0.53</b>	0.17
LE_LS21LE	var.	0.146	0.016	0.090	0.026
	var. ratio	-	0.11	<b>0.62</b>	0.18
LE_PS17LE	var.	0.060	0.008	0.030	0.006
	var. ratio	-	0.14	<b>0.51</b>	0.10
LE_PS24LE	var.	0.053	0.009	0.027	0.007
	var. ratio	-	0.16	<b>0.50</b>	0.14
LE_4IIS	var.	0.017	0.002	0.008	0.003
	var. ratio	-	0.13	<b>0.47</b>	0.19
LE_2BS	var.	0.122	0.006	0.087	0.007
	var. ratio	-	0.05	<b>0.71</b>	0.06
LE_LR	var.	0.087	0.018	0.019	0.029
	var. ratio	-	0.20	0.21	0.33
LE_SG3	var.	0.067	0.014	0.034	0.008
	var. ratio	-	0.21	<b>0.51</b>	0.12
LE_P1TAU	var.	0.029	0.011	0.004	0.010
	var. ratio	-	0.39	0.14	0.34

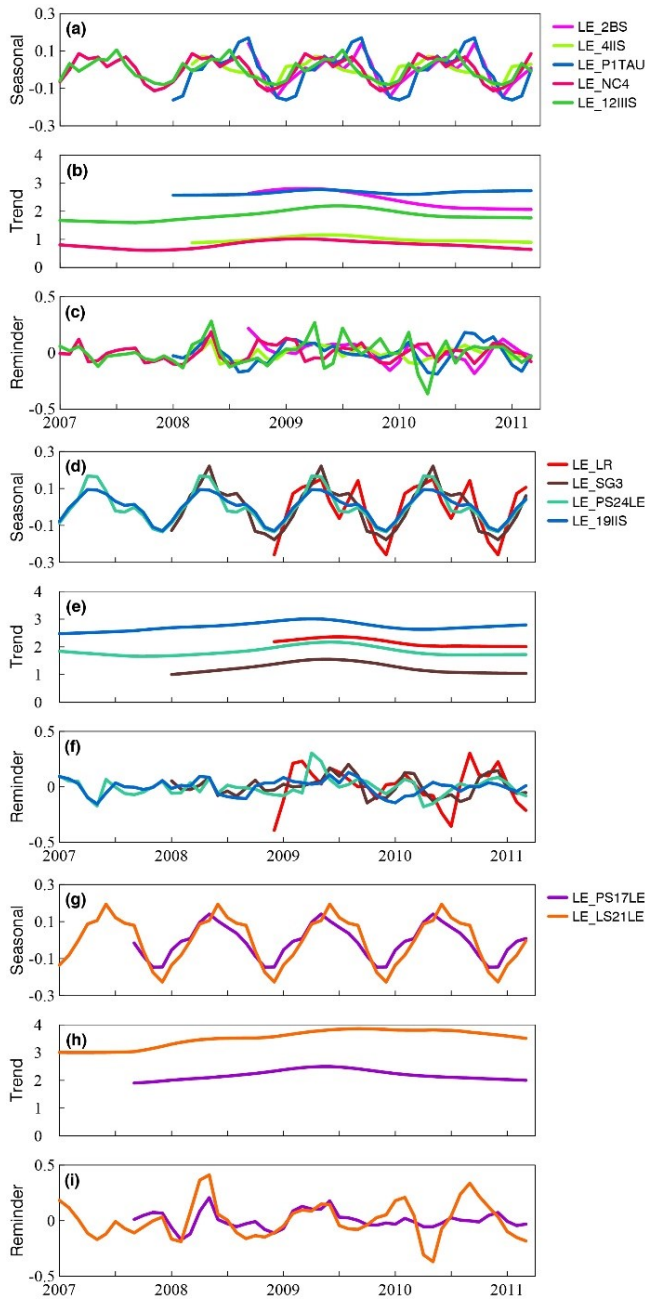


Fig. 31 - GwL-STL decomposition classified into Group 1 (a, b, c), Group 2 (d, e, f) and Group 3 (g, h, i). The decomposition reports the Seasonal component as the first graph for every category, the trend in the middle and the remainder as the last. Ticks are centred in October.

## 6.5 Wavelet results

The wavelet analysis was applied to the daily meteorological and GWL time series. Fig. 32 displays the time-frequency spectrum of the daily precipitation registered by Ruffano station, whilst Fig. 33 shows the time-frequency spectra for the daily GWL time series from 2007-2011. The thick black solid line designates a 5% significance level in both figures. Additionally, the COI is shown by lighter shades. Warmer colours (e.g., red) indicate the peaks, while colder colours (e.g., blue) designate the domain of the power density.

The wavelet precipitation analyses indicate moderate power regions with 4-16-day (in red) reoccurrence patterns and a second considerable power band (in orange) that occurs from 2008-2010 for periods higher than 256 days. As for Ruffano, another significant power region occurred in all rain gauge stations in 2010, with a power band from 64 to 256 days. This region has been more extensive in Lecce station since 2009 (Fig. 32b).

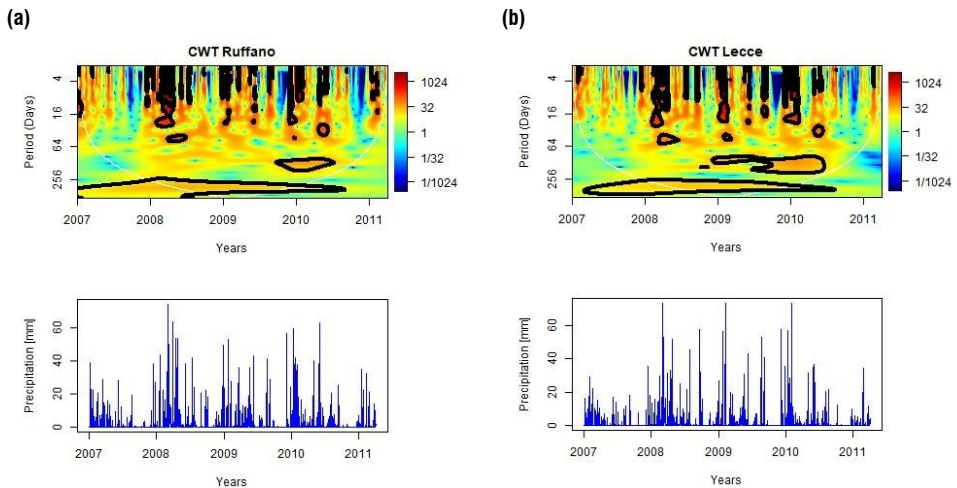
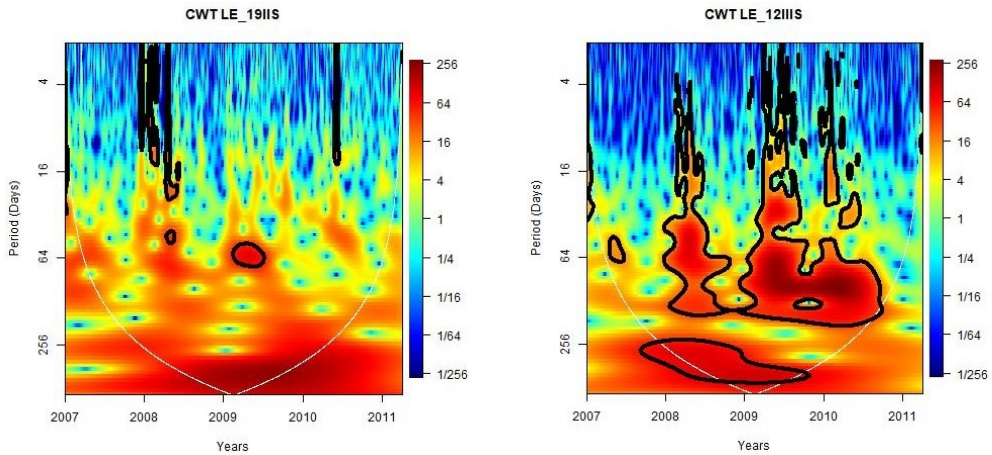
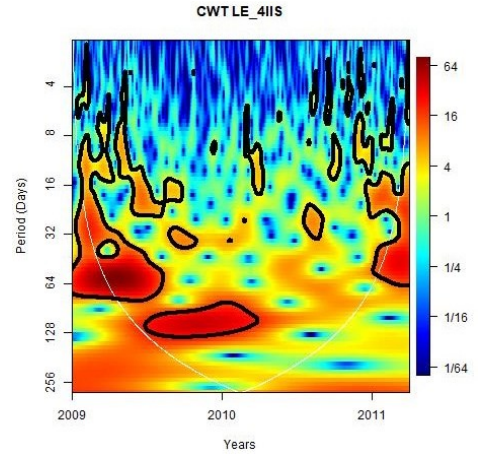
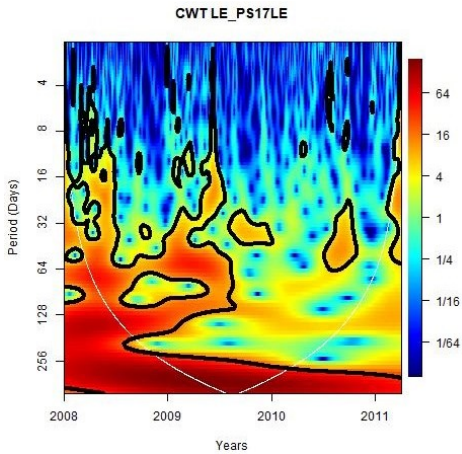
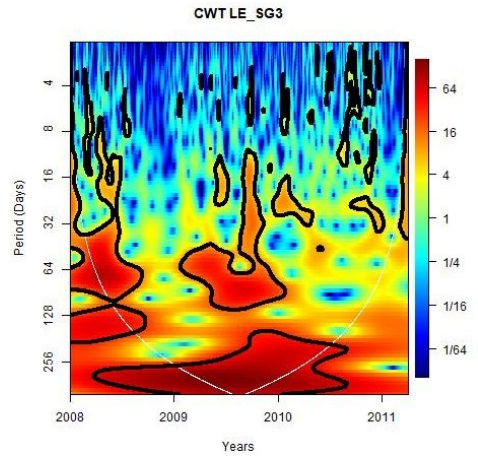
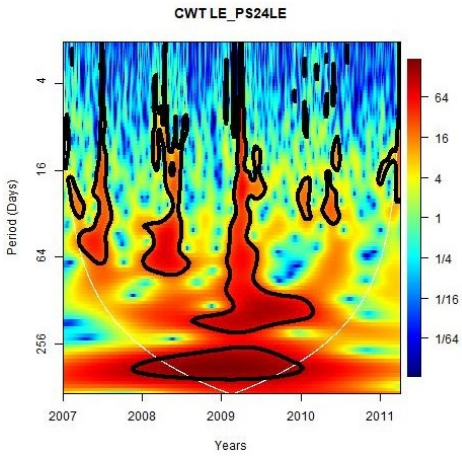
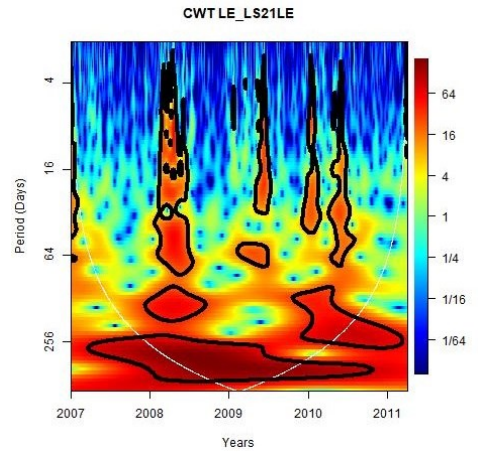
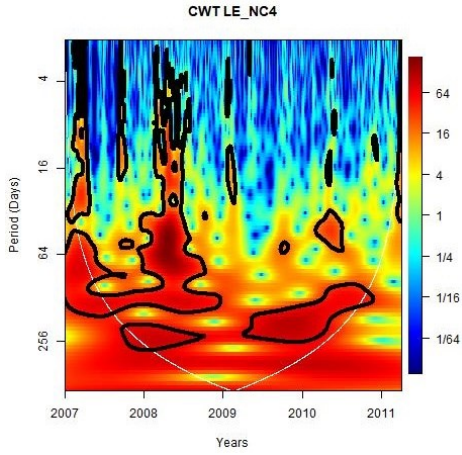


Fig. 32 - Continuous wavelet transform (CWT) of daily rainfall series for (a) Ruffano and (b) Lecce stations. The area covered by a white screen represents the cone of influence (COI). The thick black contour represents the significant region at a 95 % confidence level.

The time-frequency spectra for the daily GWLs shown in Fig. 33 highlight relatively distinct patterns. In contrast to precipitation, many statistically significant periodicities can be detected in intervals of less than a year. At the 5% significance level, notable power is observed in the 4- to 64-day periods, corresponding to the autumn and

summer seasons. This is particularly evident in the LE\_12/IIS, LE\_LS21LE, and LE\_PS24LE wells. In contrast, no significant seasonal periodicity is detected in the LE\_19/IIS, LE\_PS17LE, LE4/IIS, LE\_2/BS, and LE\_LR wells, as no robust and meaningful power regions can be identified in these cases. As detected in precipitation, most analysed GWLs exhibit continuous periodicity exceeding 256 days. These findings emphasise that the aquifer acts as a filter, transforming the precipitation signal into groundwater signals, as discussed by Wang et al. (2023). The apparent annual periodicity observed in most monitoring wells demonstrates the extensive storage capacity of the investigated system. However, in this coastal aquifer, when infiltration waters penetrate the saturated zone, the rise in water level disrupts the freshwater-saltwater equilibrium, causing local changes in the elevation of the transition zone. Concurrently, groundwater undergoes spatial redistribution over time on a regional scale due to buoyancy effects, which could potentially govern the extent of GWL and its oscillations, possibly hiding the effects of overexploitation. Different results were reached by Alfio et al. (2023) for the Pinios River basin (Greece), which extension is marginal compared to that of Salento. Applying this technique to fourteen boreholes and two rain gauge stations, they discovered a notable seasonal periodicity in most of the analysed wells in the power band from 4 to 64 days, attributable to climate and human pressure. In the long period, the common frequencies reflect those obtained for the associated precipitation stations, while in the short term, they correspond to groundwater abstraction since this basin is highly stressed by withdrawal for irrigation.





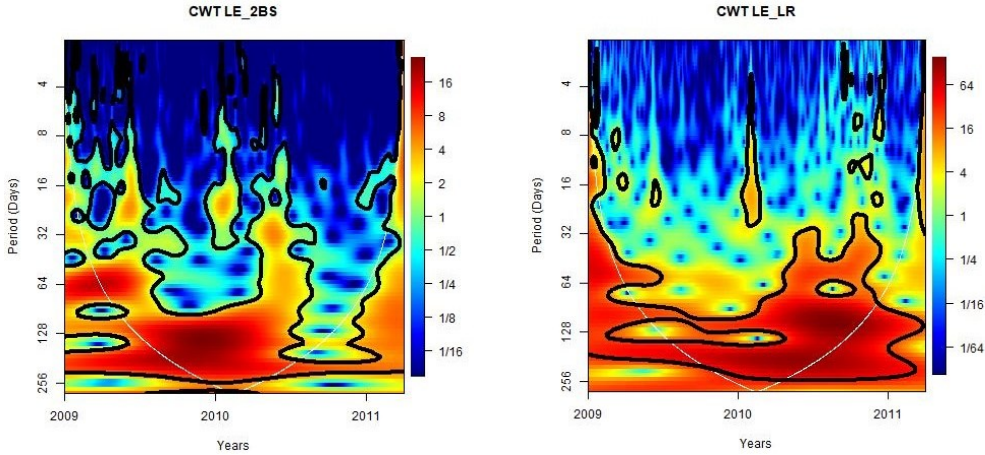
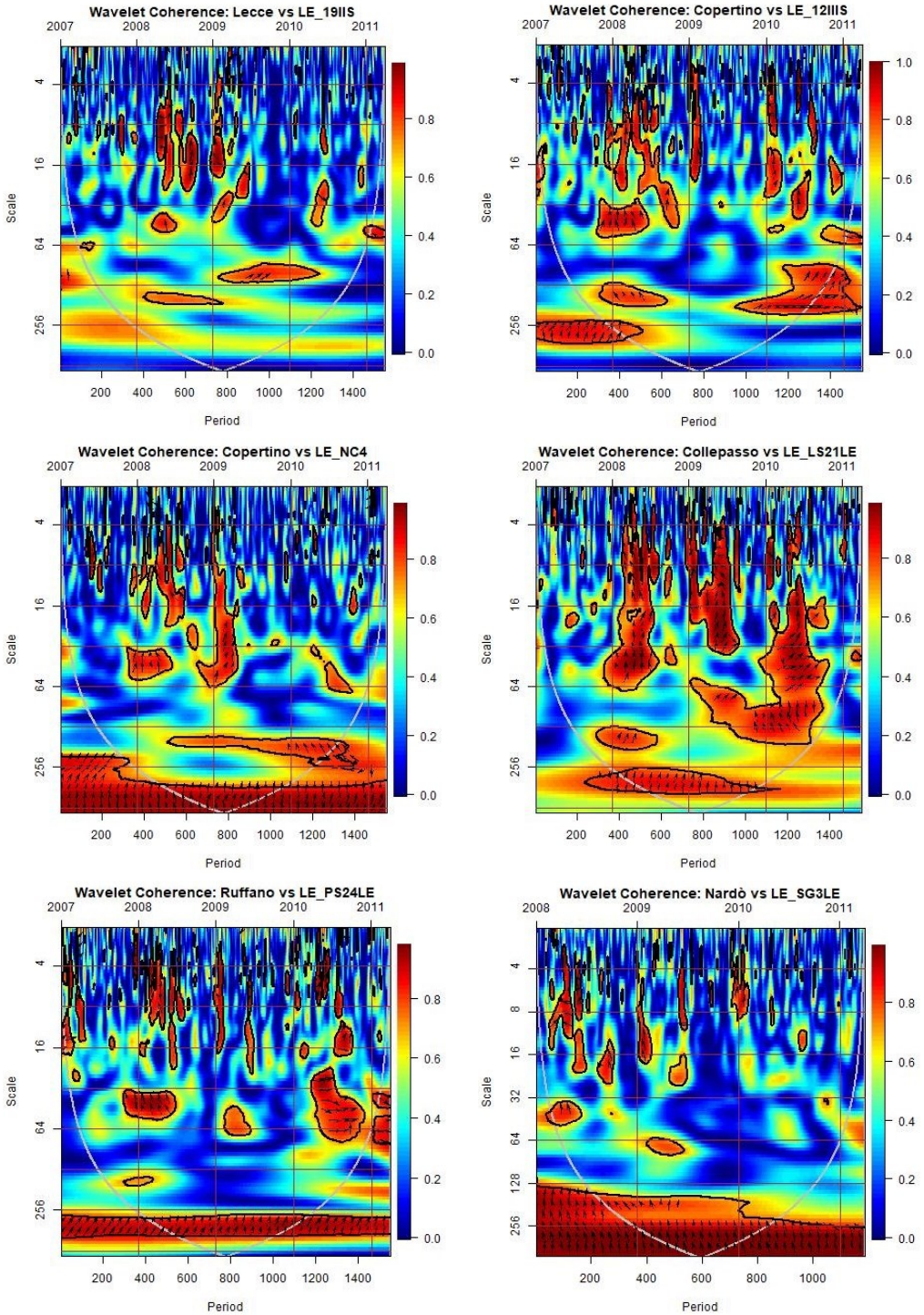


Fig. 33 - Continuous wavelet transform (CWT) of daily GWL series for the Salento wells. The area covered by a white screen represents the cone of influence (COI). The thick contour represents the significant region at a 95 % confidence level.

Wavelet coherence analysis was employed for each pair of time series to evaluate the correlation between precipitation and GWLs (Fig. 34). The coloured shading on the graphs corresponds to the coherence magnitude, as depicted in the colour bar on the right side. It varies from 0 (indicating no correlation) to 1 (representing a perfect correlation) and reflects the temporal variation in the correlation between the two-time series. Generally, the regions of significance are those where the coherence values exceed 0.75 (highlighted in red) and fall within the COI. The areas outside the COI must not be interpreted. Furthermore, the black arrows define the relative phase relationships. Arrows pointing to the right signify an in-phase relationship (positive correlation), while arrows pointing to the left represent an anti-phase relationship (negative correlation). Arrows in other directions indicate time lags between precipitation and GWL. Distinct patterns are observable in each well. In most cases, substantial and uninterrupted coherence regions are consistently present, which are statistically significant at the 95% confidence level, particularly in the higher-than-256-day band. Conversely, correlations at shorter time scales (4 to 64 days) tend to be sporadic and in an anti-phase relationship. These findings validate the influence of rainfall on GWLs over the long term and underscore the delayed response of the aquifer to seasonal climate variations.





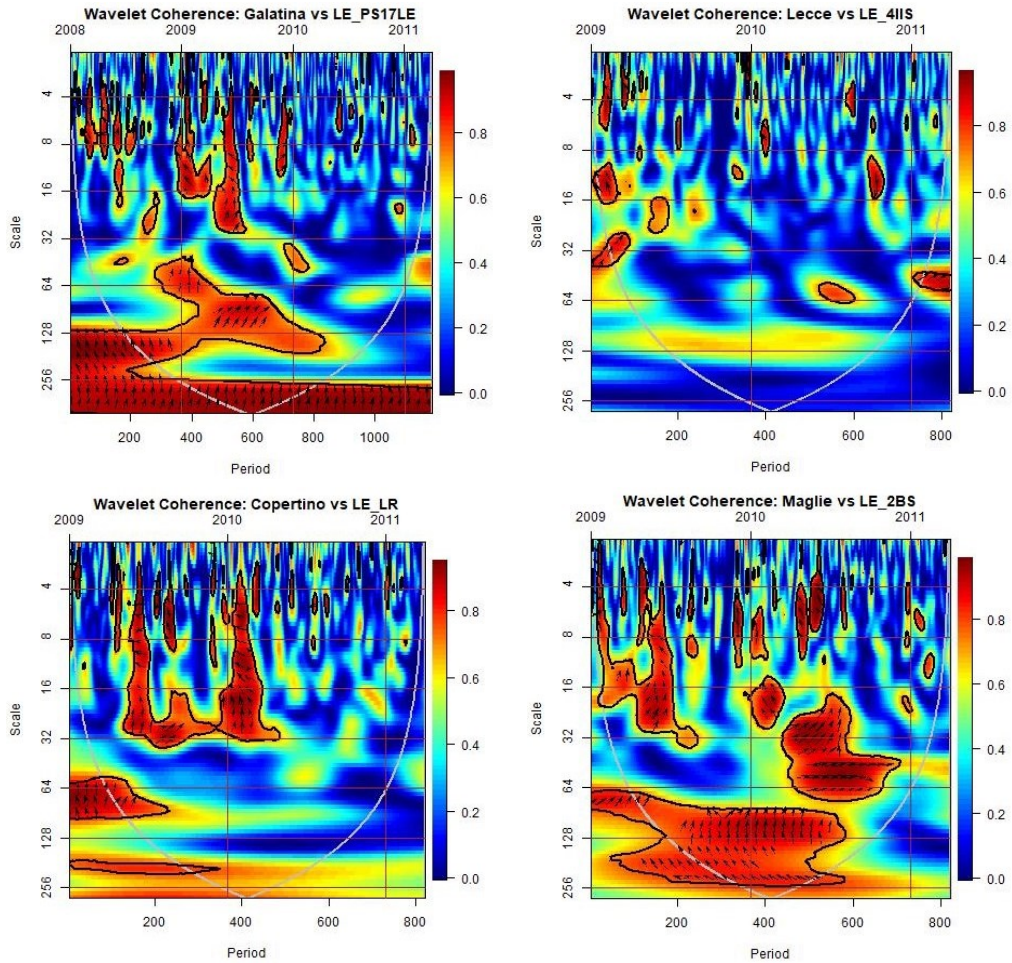


Fig. 34 - Wavelet coherence between GWLs time series and precipitation stations.

## 6.6 *MVSA and HFE comparison*

### 6.6.1 Tiziano Project database

The Salento database included 449 samples of chemical and physical parameters collected from 2007 to 2011, of which 268 were considered for Multivariate Statistical Analyses (HCA and FA) and Hydrogeochemical Facies Evolution (HFE), following the multi-criteria selection process described in section 4.1.1.

The dataset was divided into seven campaigns, with sampling at the end of each wet season (from April to September) and at the end of each dry season (from October to March). Unfortunately, each sampling campaign did not include the same set of wells. Consequently, there are instances where the study area is comprehensively sampled, and there are periods during which certain areas lack consistent and uniform sampling coverage.

MVSA applications for the Salento aquifer considered the physical and chemical variables, including Hardness (HR), EC,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{HCO}_3^-$ ,  $\text{NO}_3^-$ , TOC, SAR, Br, F, Fe. Table 14 summarises the statistical results of the considered parameters. According to the Kolmogorov-Smirnov test, most showed a non-normal distribution ( $p\text{-value} < 0.05$ ). Therefore, these parameters were subjected to a data normalisation step using the Box-Cox transformation. Finally,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ , TOC, SAR, Br, F, and Fe nearly matched the normal distribution after Box-Cox transformation.

Fig. 35 displays the dendrogram from HCA applied to all the sampling periods and the selected variables. Within this analysis, three clusters of groundwater samples were determined by choosing a dissimilarity value of approximately 25. This selection was pointed out by the visual inspection of the dendrogram, where the "phenon line" was established.

To further illustrate the variations between these identified clusters, boxplots of all major ions were generated for each group derived from the HCA (Fig. 36). Additionally, Table 15 provides the median values for all variables, offering a summarised view of the dataset's characteristics within these distinct groupings.

Table 14 - Statistical summary of parameters considered in MVSA applications.

Param.	Min	Max	Mean	Median	Std. Dev	p-value	BC $\lambda$	BC p_value
<b>HR</b>	8.60	583.80	45.2	36.3	53.8	<0.05	-0.57	0.46
<b>EC</b>	315.00	44,600	1755.3	978	3982.5	<0.05	-0.65	0.96
<b>HCO<sub>3</sub><sup>-</sup></b>	127.00	625	314.7	317	65.7	0.49	NA	NA
<b>Ca<sup>2+</sup></b>	9.18	722	96	82.2	70	<0.05	-0.14	<0.05
<b>Cl<sup>-</sup></b>	17.80	18,400	482.9	169.5	1748.7	<0.05	-0.19	0.55
<b>Mg<sup>2+</sup></b>	3.05	980	51.5	36.8	92.2	<0.05	-0.08	<0.05
<b>K<sup>+</sup></b>	0.20	445	13	6.2	38.1	<0.05	-0.05	0.22
<b>Na<sup>+</sup></b>	10.10	10,200	254.1	88.9	946.7	<0.05	-0.23	0.64
<b>SO<sub>4</sub><sup>2-</sup></b>	0.78	2,710	82.1	37.8	247.6	<0.05	-0.03	<0.05
<b>NO<sub>3</sub><sup>-</sup></b>	0.06	192	33.1	28.7	30.5	<0.05	0.42	<0.05
<b>TOC</b>	0.17	7.05	1.1	0.9	0.8	<0.05	-0.28	<0.05
<b>SAR</b>	0.35	58	3.6	2.1	6.25	<0.05	-0.20	<0.05
<b>Br<sup>-</sup></b>	27.50	71,100	1,820.6	610	6,616.72	<0.05	-0.15	<0.05
<b>F</b>	0.06	0.7	0.2	0.2	0.09	<0.05	0.55	<0.05
<b>Fe</b>	1.10	30,920	1,422.2	60.2	3,485.2	<0.05	-0.05	<0.05

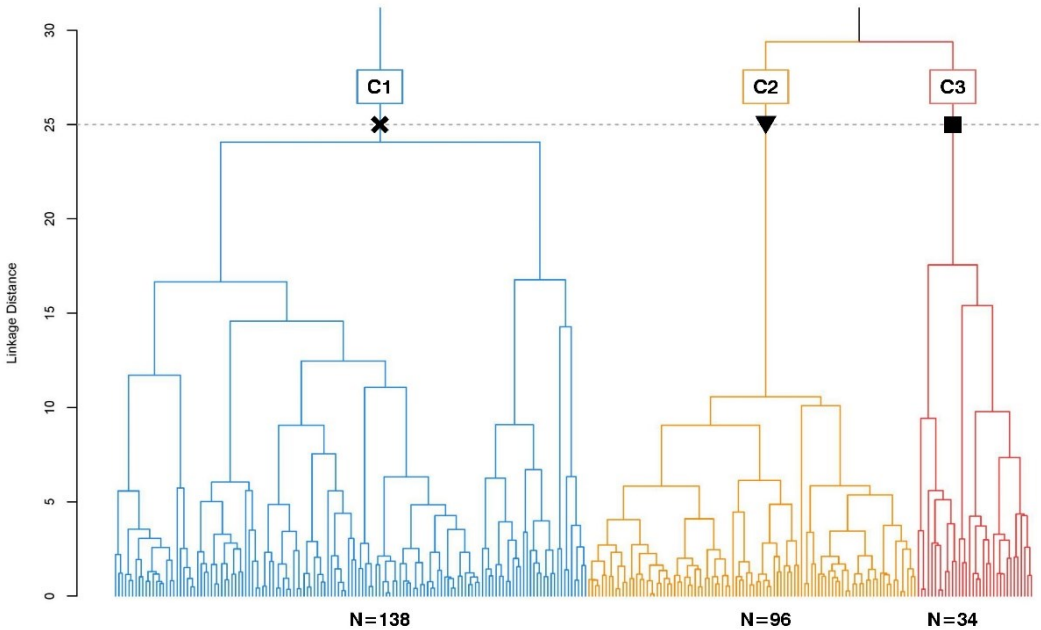


Fig. 35 - Dendrogram of HCA obtained with Tiziano Project dataset. The dashed line represents the selected level of dissimilarity (25), which identifies three clusters: C1, C2 and C3.

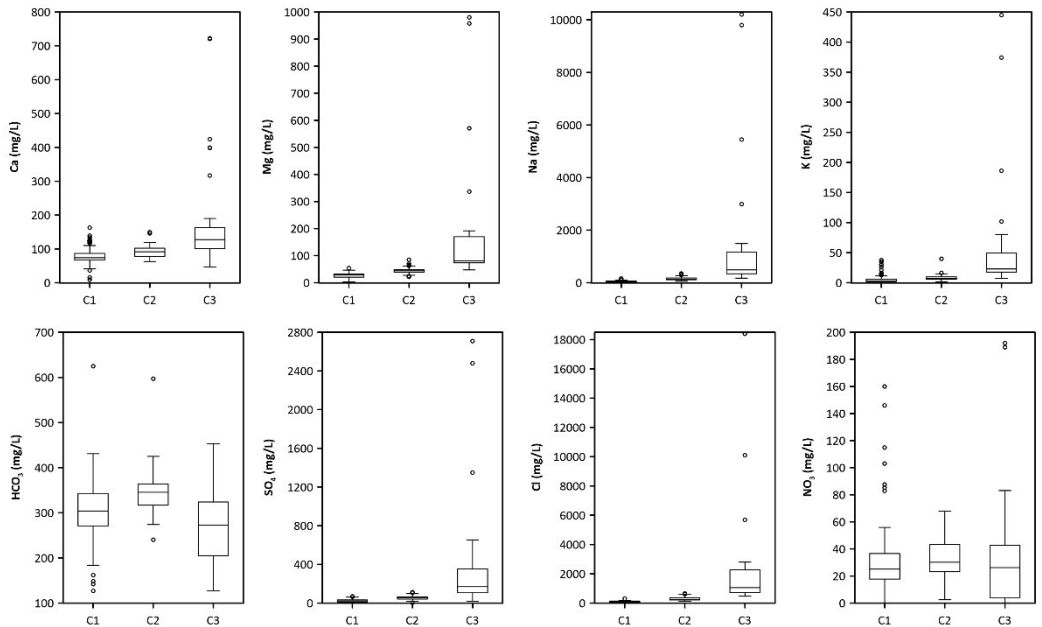


Fig. 36 - Boxplots for major ions for each cluster (C1, C2 and C3) derived by HCA.

Table 15 - Median values of physical and chemical parameters for each cluster derived by HCA.

Parameters	Unit	C1	C2	C3
HR	°f	30.7	40.4	68.35
EC	μS/cm	712.0	1,252.0	3316
HCO <sub>3</sub> <sup>-</sup>	mg/L	304.0	345.5	272.5
Ca <sup>2+</sup>	mg/L	74.0	91.7	127.5
Cl <sup>-</sup>	mg/L	80.2	246.5	1051
Mg <sup>2+</sup>	mg/L	28.6	44.3	80.9
K <sup>+</sup>	mg/L	3.1	7.6	22.75
Na <sup>+</sup>	mg/L	43.3	132.5	490
SO <sub>4</sub> <sup>-</sup>	mg/L	22.1	56.8	169
NO <sub>3</sub> <sup>-</sup>	mg/L	24.0	30.2	26.2
TOC	mg/L	1.0	0.8	1.32
SAR	-	1.1	3.0	9.49
Br <sup>-</sup>	μg/L	320.0	950.0	3,675
F	mg/L	0.2	0.2	0.235
Fe	μg/L	48.6	40.3	1,751.5

C1 cluster comprises water samples with the lowest chloride concentrations ( $\text{Cl}^- = 80.2 \text{ mg/L}$ ). In contrast, samples in clusters C2 and C3 exhibit higher chloride concentrations ( $\text{Cl}^- = 246.5 \text{ mg/L}$  and  $1,051 \text{ mg/L}$ , respectively) compared to those in C1 samples. Other parameters associated with the salinisation process, such as  $\text{Na}^+$ ,  $\text{SO}_4^{2-}$ ,  $\text{K}^+$ , also demonstrate an increasing trend from C1 to C3. Referring to the median values of each parameter as presented in Table 15 and following the classification outlined in Table 16, it can be inferred that:

- the C1 cluster represents freshwater samples,
- the C2 group refers to fresh-brackish,
- the C3 is indicative of brackish water samples.

Table 16 - Classification of groundwater based on  $\text{Cl}^-$  values.

<b>Facies Prefix</b>	<b>Range of <math>\text{Cl}^-</math> (mg/L)</b>
Oligohaline	5-30
Fresh	31-250
Fresh-brackish	251-1,000
Brackish	1,001-6,000
Brackish-salt	6,001-10,000
Saline	10,001-20,000
Hyper-saline	>20,001

The FA analyses revealed three significant factors that collectively account for 73.2 % of the total variance in the dataset. Results confirm that groundwater salinisation is the primary phenomenon affecting groundwater quality in the Salento aquifer, as summarised in Table 17.

The first factor, denoted as F1 and explaining 48.64% of the total variance, is characterised by high factor loadings associated with various parameters. These include EC (0.94),  $\text{Cl}^-$  (0.97),  $\text{Mg}^{2+}$  (0.79),  $\text{Na}^+$  (0.98), and SAR (0.98). Additionally,  $\text{Br}^-$  (0.95) exhibits a substantial factor loading, which is indicative of its relationship with groundwater salinisation. Alcalá & Custodio (2008) have previously demonstrated that the  $\text{Cl}^-/\text{Br}^-$  ratio can serve as an effective tracer for identifying the source of salinity in aquifers. The second latent factor reflects a freshening process, as high factor loadings characterise it for  $\text{Ca}^{2+}$  (0.82) and HR (0.65).

Lastly, the third factor is influenced by  $\text{NO}_3^-$  (-0.70), TOC (0.70) and Fe (0.80), albeit with a lower total variance of 11.55 %.

Table 17 - Varimax rotated factor loadings, eigenvalues, and total and cumulative variance related to the Tiziano Project dataset.

	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>	<b>Factor 4</b>
<b>HR</b>	0.70	<b>0.65</b>	0.02	0.14
<b>EC</b>	<b>0.94</b>	0.30	0.03	0.08
<b>HCO<sub>3</sub><sup>-</sup></b>	-0.12	0.45	-0.03	0.74
<b>Ca<sup>2+</sup></b>	0.41	<b>0.82</b>	-0.09	-0.06
<b>Cl<sup>-</sup></b>	<b>0.97</b>	0.14	0.05	0.00
<b>Mg<sup>2+</sup></b>	<b>0.79</b>	0.23	0.15	0.28
<b>K<sup>+</sup></b>	<b>0.83</b>	0.05	0.11	-0.16
<b>Na<sup>+</sup></b>	<b>0.98</b>	0.09	0.05	0.05
<b>SO<sub>4</sub><sup>-</sup></b>	<b>0.80</b>	0.36	-0.17	-0.16
<b>NO<sub>3</sub><sup>-</sup></b>	0.09	0.49	<b>-0.70</b>	0.02
<b>TOC</b>	0.06	0.16	<b>0.70</b>	-0.11
<b>SAR</b>	<b>0.98</b>	-0.02	0.07	0.01
<b>Br<sup>-</sup></b>	<b>0.95</b>	0.15	0.09	0.04
<b>F</b>	0.12	-0.24	-0.11	0.76
<b>Fe</b>	0.16	-0.10	<b>0.80</b>	0.01
<b>Eigenvalue</b>	7.92	2.08	1.26	1.03
<b>% Total - variance</b>	48.64	13.01	11.55	8.72
<b>% Cumulative - variance</b>	48.64	61.65	73.20	81.92

After applying MVSA, the chemical dataset was used for HFE analysis. Fig. 37 shows the HFE-Diagram with freshening/intrusion processes substages for all water samples. The Conservative Mixing Line was built with freshwater and saltwater end-members. Regarding the freshwater end-member, the HFE-D model automatically selects the chemical composition of the freshest groundwater collected during the seven sampling surveys. In contrast, the saltwater end-member refers to the average chemical analyses of saltwater samples collected in past monitoring programs (1986) in deep wells reaching saltwater beneath freshwater in the study area (refer to Table 5 in section 4.2). The samples categorised as “freshening” correspond to those with the lowest TDS and Cl<sup>-</sup> concentrations. Conversely, water samples in the “intrusion zone” are characterised

by TDS levels reaching up to 5,000 mg/L. Few samples exhibit TDS values exceeding 5,000 mg/L, accompanied by  $\text{Cl}^-$  concentrations of about 20,000 mg/L, indicating the presence of saltwater. Based on classifications derived from HCA HFE analyses, the following associations can be made:

- Waters in the C1 group are predominantly associated with the Ca- $\text{HCO}_3$  and MixCa- $\text{HCO}_3$  facies (f3, f4, and FW). This finding suggests that C1 groundwater samples primarily originate from recharge areas characterised by freshening processes.
- In contrast, the C2 group is mainly associated with the MixNa-Cl and MixNa-MixCl facies (i2 and i3). These results indicate that C2 groundwater samples are sourced from areas where there is a contribution of sodium and chloride ions, possibly reflecting the influence of seawater/saltwater intrusion.
- Finally, the C3 group is nearly exclusively linked to the Na-Cl facies (SW), strongly suggesting that C3 groundwater samples originate from zones affected by sodium-chloride-dominated water, which could be indicative of saltwater or seawater intrusion.

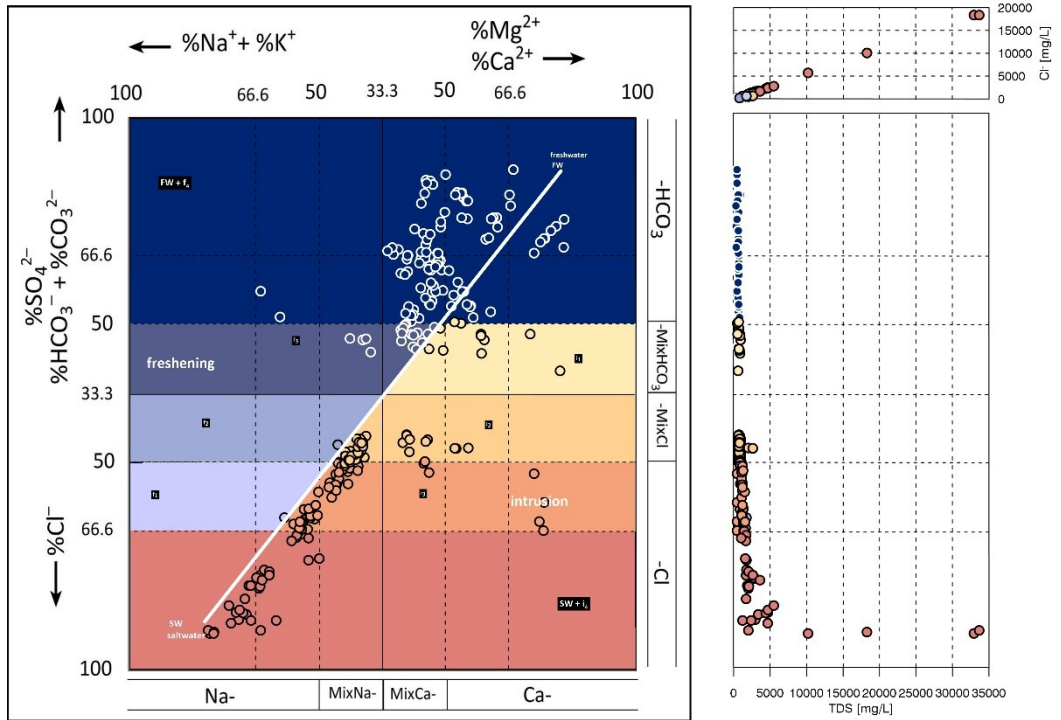


Fig. 37 - HFE-Diagram of the chemical analyses for the seven sampling surveys of the Tiziano Project. The diagrams on the right side add information about the TDS and chloride concentration related to the freshening and intrusion HFs (Giménez-Forcada, 2014, 2019).

To visually compare the HCA, FA and HFE techniques and to evaluate the temporal evolution over the seven sampling surveys from 2007 to 2011, the results were spatially plotted, specifically investigating the dynamics of groundwater salinisation. These thematic maps are prepared by interpolating the factor scores related to F1 (groundwater salinisation) derived by FA (Fig. 38, left-hand side), as well as the hydrogeochemical sub-phases derived by the HFE analysis (Fig. 38, right-hand side) and locating HCA groups with different symbols.

The spatial interpolation derived by FA distinguishes the areas more affected by F1 (groundwater salinisation) as the factor score increases and those less affected by F1 as the factor score decreases. The spatial interpolation derived by HFE-D allows distinguishing areas dominated by freshening processes and those dominated by intrusion sub-phases.



These thematic maps were determined by using the Ordinary Kriging interpolation method. The experimental variogram (Spherical) was built based on the observations for each of the seven sampling periods. Table 18 and Table 19 summarise the identified parameters, i.e., partial sill, nugget, and range for F1 factor scores and HFE substages, respectively. No cut-off was used since the spatial distribution of samples was inhomogeneous and sparse compared to the extension of the studied aquifer.

Table 18 - Ordinary Kriging parameters for F1 factor scores.

<b>Sample campaign</b>	<b>Variogram</b>	<b>Partial Sill</b>	<b>Nugget</b>	<b>Range</b>
<b>DRY_2007</b>	Sph	0.6	0.2	15,000
<b>WET_2008</b>	Sph	0.8	0.2	15,000
<b>DRY_2008</b>	Sph	0.5	0.3	15,000
<b>WET_2009</b>	Sph	0.5	0.3	15,000
<b>DRY_2009</b>	Sph	1	0.5	15,000
<b>WET_2010</b>	Sph	0.6	0.2	15,000
<b>DRY_2010</b>	Sph	1	0.5	15,000

Table 19 - Ordinary Kriging parameters for HFE substages.

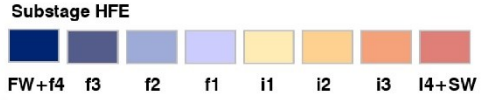
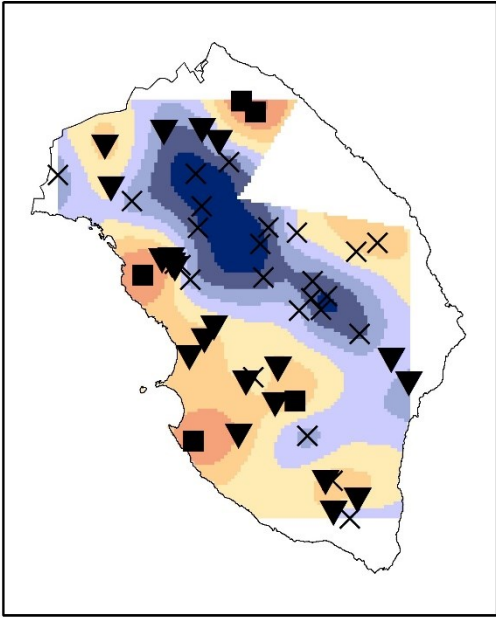
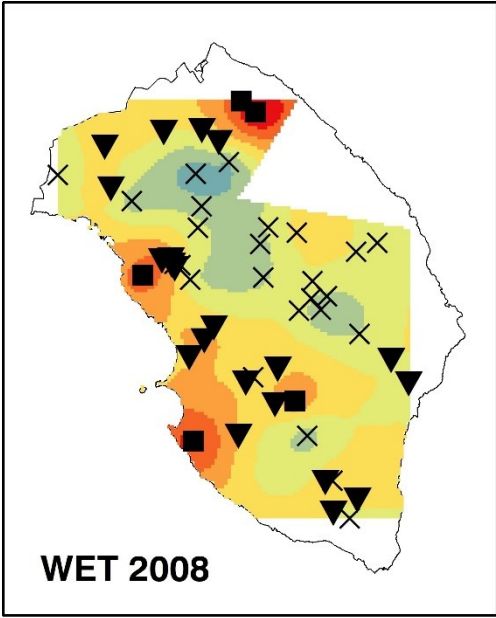
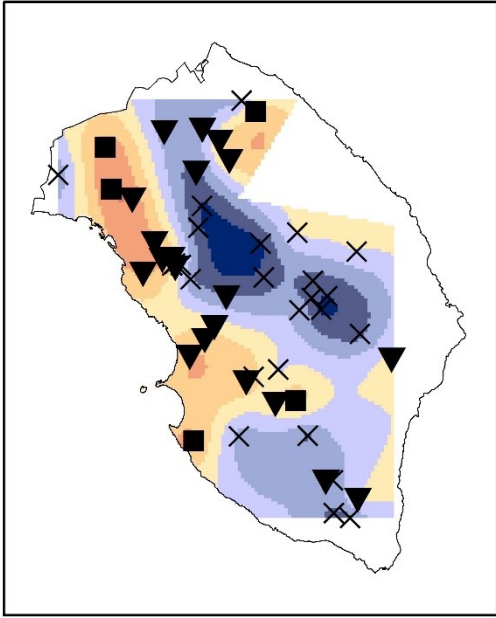
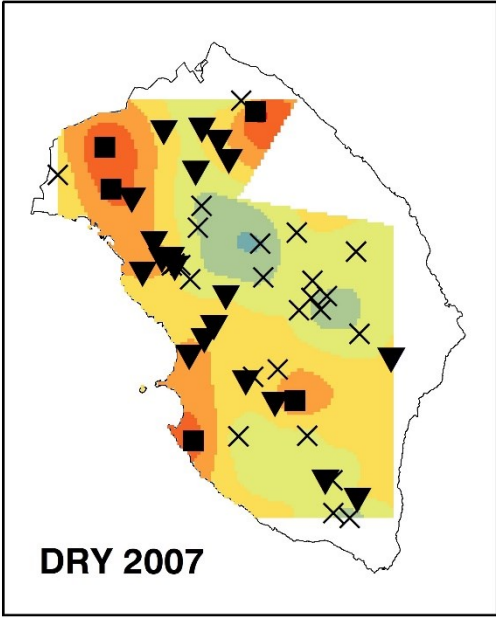
<b>Sample campaign</b>	<b>Variogram</b>	<b>Partial Sill</b>	<b>Nugget</b>	<b>Range</b>
<b>DRY_2007</b>	Sph	7	4	15,000
<b>WET_2008</b>	Sph	8	4	15,000
<b>DRY_2008</b>	Sph	8	2	15,000
<b>WET_2009</b>	Sph	8	2	15,000
<b>DRY_2009</b>	Sph	8	5	15,000
<b>WET_2010</b>	Sph	10	2	15,000
<b>DRY_2010</b>	Sph	8	5	15,000

Fig. 38 illustrates a clear correspondence between areas with negative F1 factor scores and those associated with the freshening sub-phases of the HFE classification. These areas are generally in alignment, and it is worth noting that most of the C1 samples are located within these zones. Consequently, the results obtained through statistical methods and HFE analysis are consistent in defining the recharge areas of the Salento aquifer. These recharge areas are characterised by low TDS, influenced by freshening processes, and unaffected by F1. Notably, the position and extent of these recharge areas

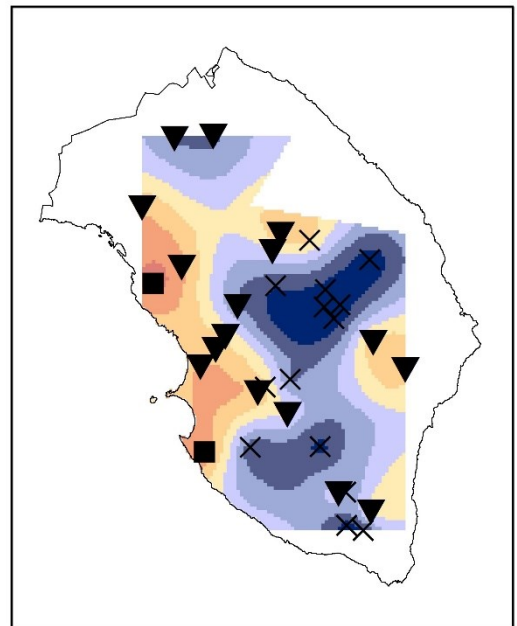
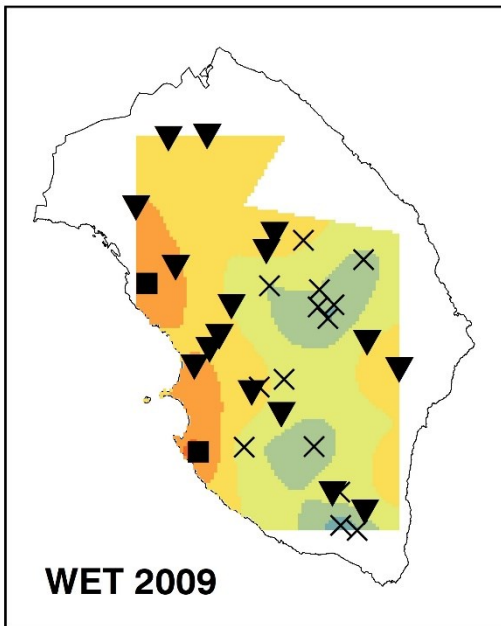
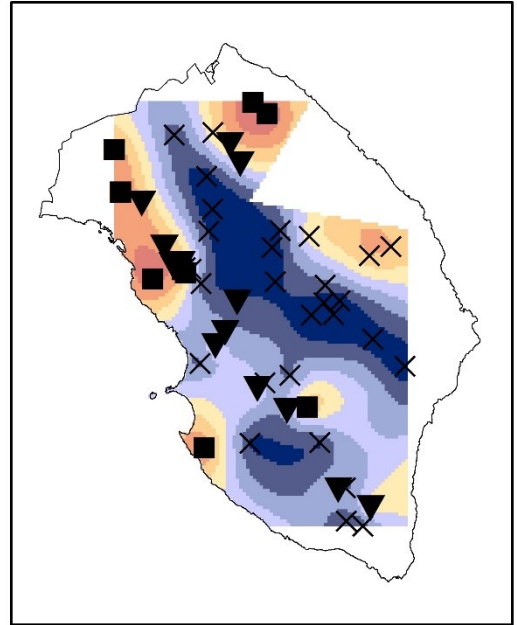
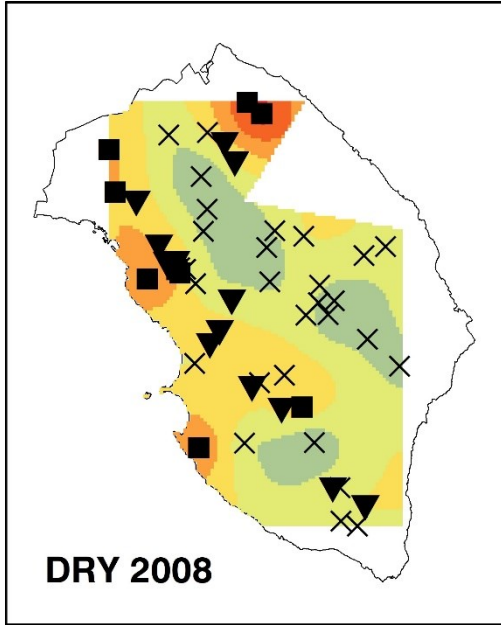
remain relatively stable over time, with only minor changes observed in the last three analysed periods, (DRY-2009/10 and WET-2010), despite limited data from sampled wells. One prominent recharge area is identifiable along the northwest border of the Salento aquifer, coinciding with the higher elevations of the adjacent Murgia aquifer. This area appears to be connected to the primary recharge zone of the Salento aquifer, situated in the central part of the Lecce Province, nearly corresponding to the endorheic part of the territory, extending into the southern portion of the Salento aquifer.

Conversely, areas with positive F1 factor scores, as depicted in Fig. 38, correspond to locations marked by C2 and C3 water samples and various salinisation sub-phases (i1, i2, i3, i4 and SW). These areas are primarily situated along the Ionian coastline and within a narrow strip extending from the Ionian to the Adriatic seas in the northern part of the aquifer. This observation corroborates previous findings from studies conducted within the Water Protection Plan of Apulia Region (2005, 2015), as well as by Polemio (2016), which identified a narrow strip affected by groundwater salinisation extending from the Ionian to the Adriatic seas between the Murgia and the Salento aquifers. No significant information can be evident on the Adriatic side coast due to a few spatially heterogeneous sampled wells.

The three statistical approaches are in good agreement and appropriate to investigate groundwater dynamics. The amount of data and the spatial distribution of the water samples influence the spatial resolution of the results. Therefore, a more homogeneous and denser spatial distribution of sampling points would be beneficial for a comprehensive understanding of the areas showing variable dynamics of salinisation.



**Cluster** X C1 ▼ C2 ■ C3



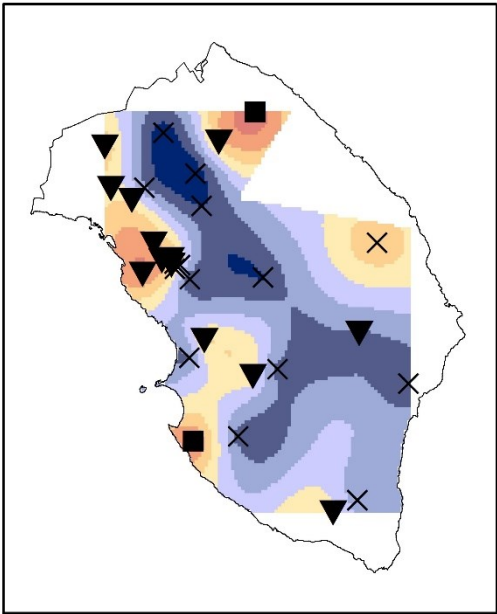
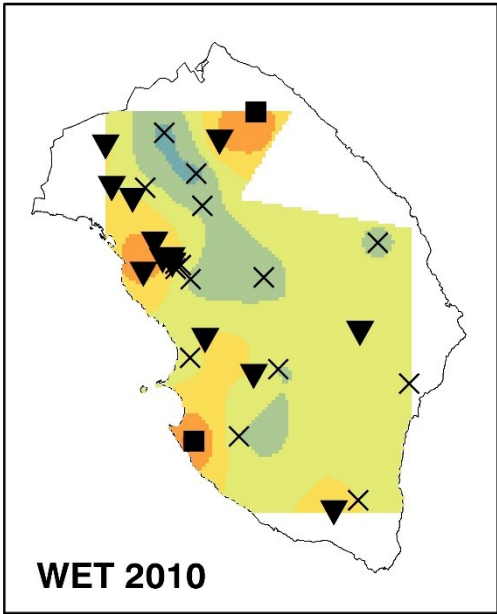
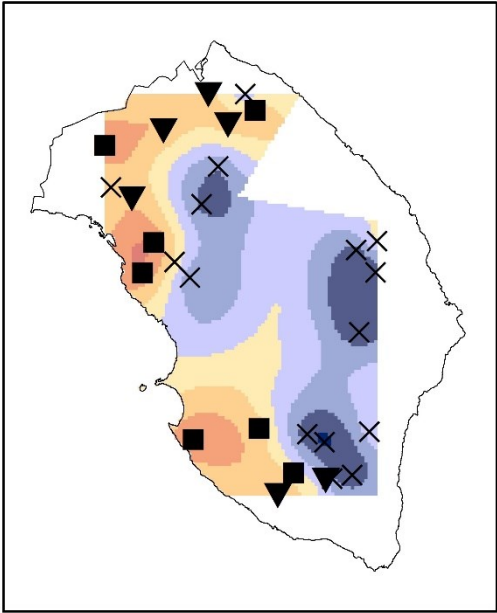
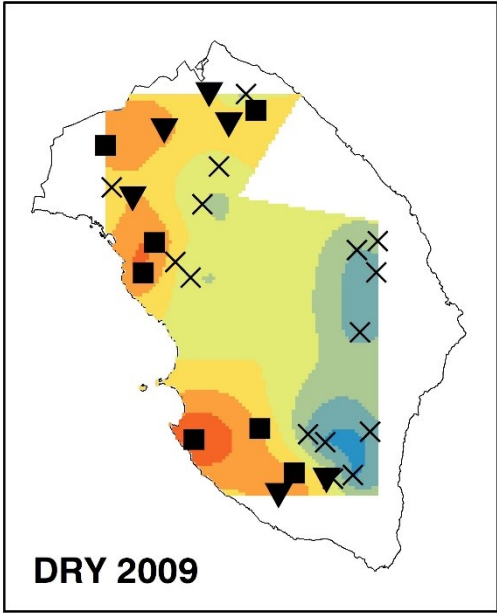
F1 – Factor Score



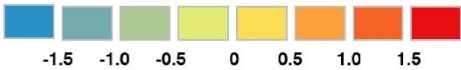
Substage HFE



Cluster X C1 ▼ C2 ■ C3



F1 - Factor Score



Substage HFE



**Cluster** × C1 ▼ C2 ■ C3

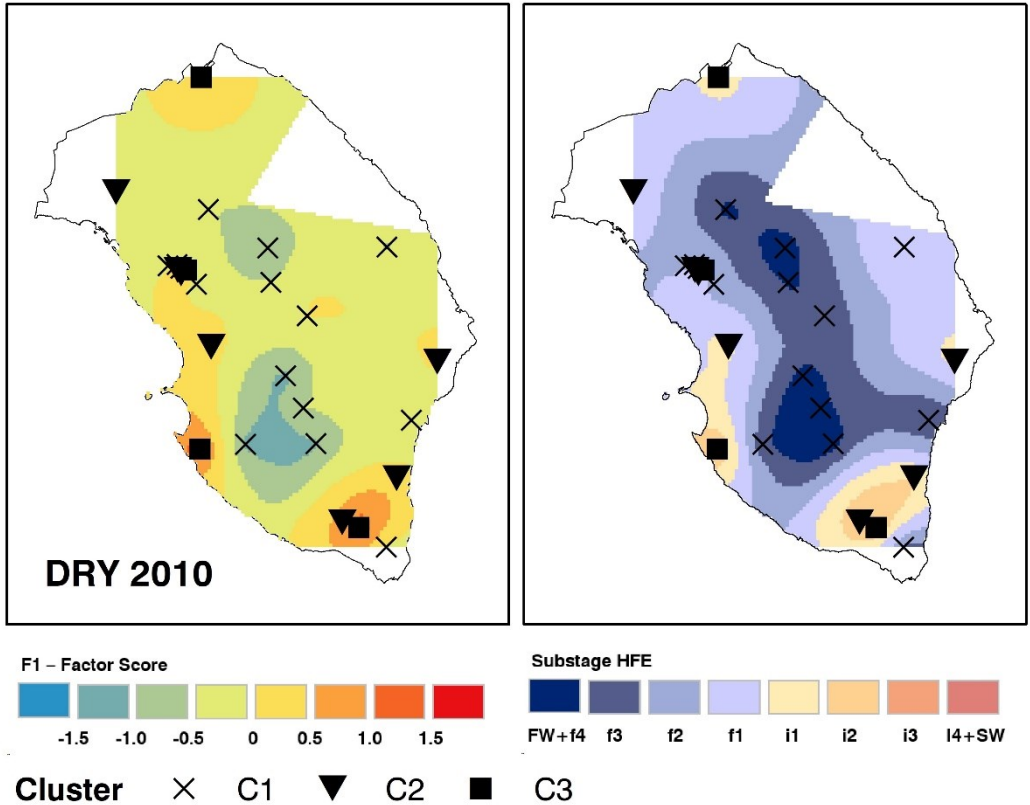


Fig. 38 - Results of FA, HFE-D and HCA are shown for each sampling period of the Tiziano Project. On the left-hand side, the thematic maps are associated with FSs of F1(groundwater salinisation) derived by FA; on the right-hand side, the thematic maps are associated with HFes derived by the HFE-D. On both sides, results are derived by Q-mode HCA.

### 6.6.2 Maggiore Project database

Similar results were achieved by Parisi et al. (2023) over the study area using the dataset originating from *Maggiore Project* (2015-2018). They organised the chemical dataset into six monitoring surveys with sampling at the end of each wet season (from April to June) and at the end of each dry season (from October to December). After a preliminary data screening (checking non-numerical/missing values, charge balance error, and normality test as discussed in 4.1.1), only wells with repeated sampling

in the six surveys were considered, obtaining a net of 31 monitoring wells (12 static and 19 pumping).

The chemical analyses concerned major cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$  and  $\text{K}^+$ ) and anions ( $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$  and  $\text{HCO}_3^-$ ), as well as some minor ions ( $\text{NH}_4^+$ ,  $\text{NO}_2^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{Br}^-$  and  $\text{F}^-$ ). The physical parameters include temperature, pH, electrical conductivity (EC) and dissolved oxygen (DO).

HCA enabled distinguishing three clusters, of which C1 concerns freshwater (TDS = 567.2 mg/L), while C2 (TDS = 933.4 mg/L) and C3 (TDS = 2,938 mg/L) refer to fresh-brackish and brackish groundwater, respectively.

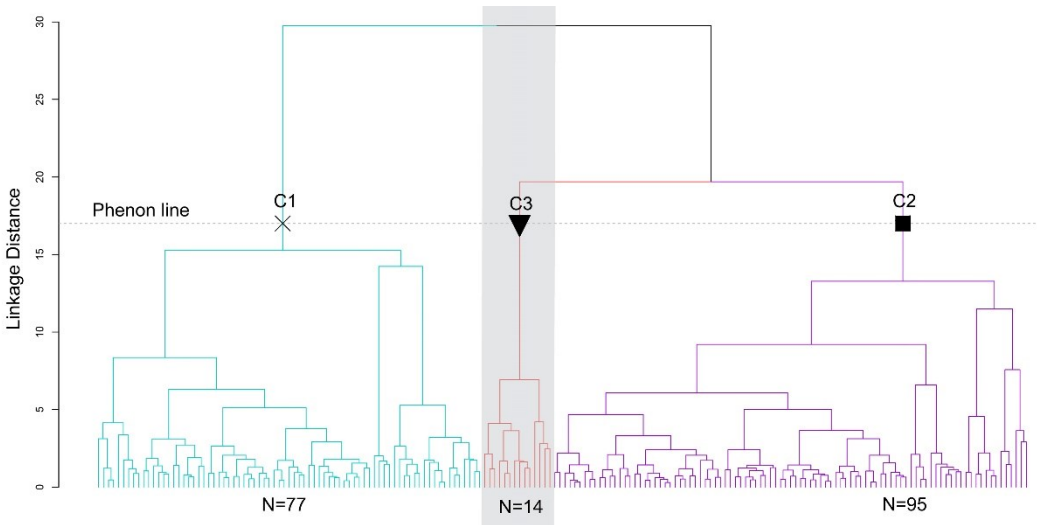


Fig. 39 - Dendrogram of HCA obtained with the Maggiore Project dataset. The dashed line represents the selected level of dissimilarity (17), which identifies three clusters: C1, C2 and C3.

According to the HFE's classification (Fig. 40), C1 mostly belongs to  $\text{Ca-HCO}_3$  and  $\text{MixCa-HCO}_3$  (f3, f4 and FW) substages, C2 is predominantly  $\text{MixNa-Cl}$  and  $\text{MixNa-MixCl}$  (i2 and i3), while C3 is nearly  $\text{Na-Cl}$  (i3, i4 and SW). Therefore, C1 groundwaters mark the areas of recharge typified by freshening processes, while C2 and C3 clusters, qualified by higher concentrations of  $\text{Cl}^-$ ,  $\text{Na}^+$ ,  $\text{K}^+$  and  $\text{SO}_4^{2-}$ , outline the zones subject to saltwater/seawater encroachment. Finally, FA results emphasised the idea that groundwater salinisation is the primary phenomenon affecting the groundwater quality

of the Salento aquifer (Table 20) since the highest factor loading associated with F1 explained the 48.26 % of the total variance, divided between  $\text{Na}^+$  (0.97),  $\text{Cl}^-$  (0.97), EC (0.93) and  $\text{Mg}^{2+}$  (0.90).

Table 20 - Varimax rotated factor loadings, eigenvalues, and total and cumulative variance related to the Maggiore Project dataset.

	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>
<b>EC</b>	<b>0.93</b>	0.21	0.10
<b>DO</b>	-0.15	-0.01	0.02
<b>pH</b>	-0.04	<b>-0.93</b>	-0.14
<b>Cl<sup>-</sup></b>	<b>0.97</b>	0.14	0.03
<b>Ca<sup>2+</sup></b>	0.51	<b>0.66</b>	0.18
<b>Mg<sup>2+</sup></b>	<b>0.90</b>	0.12	0.19
<b>Na<sup>+</sup></b>	<b>0.97</b>	0.05	0.05
<b>HCO<sub>3</sub><sup>-</sup></b>	0.15	0.21	<b>0.96</b>
<b>Eigenvalue</b>	4.37	1.38	0.91
<b>% Total - variance</b>	48.26	18.02	12.89
<b>% Cumulative - variance</b>	48.26	66.28	79.17

The spatial distribution of negative F1 factor scores, HFE substages, and the locations of the clusters identified by HCA allows for defining the recharge area of the Salento aquifer, which is associated with the lowest TDS values and dominated by freshening processes. It corresponds to the central part of the study area, coinciding with the endorheic basin and on the NW border of Salento, close to the Murgia aquifer.

The areas with positive values of F1 factor scores correspond to those affected by salinisation substages (i1, i2, i3, i4 and SW) and are indicated by C2 and C3 clusters. These areas coincide with the Ionian coastline and two strips extending from the Ionian to the Adriatic seas in the northern and southern parts of the study area.

The methodological approach used in Parisi et al. (2023) differs from that implemented with the *Tiziano Project* database in:

- i. Chemical and physical parameters used in the MVSA.
- ii. Only repeated samplings in all six monitoring surveys were considered.
- iii. The spline method applied for the spatial distribution of F1 factor scores and HFE substages.



Notwithstanding, a comparison of results demonstrated that even if based on different methodological approaches, the statistical and geochemical methods validate each other, showing comparable and complementary results in defining recharge areas and those subject to saltwater intrusion and salinisation.

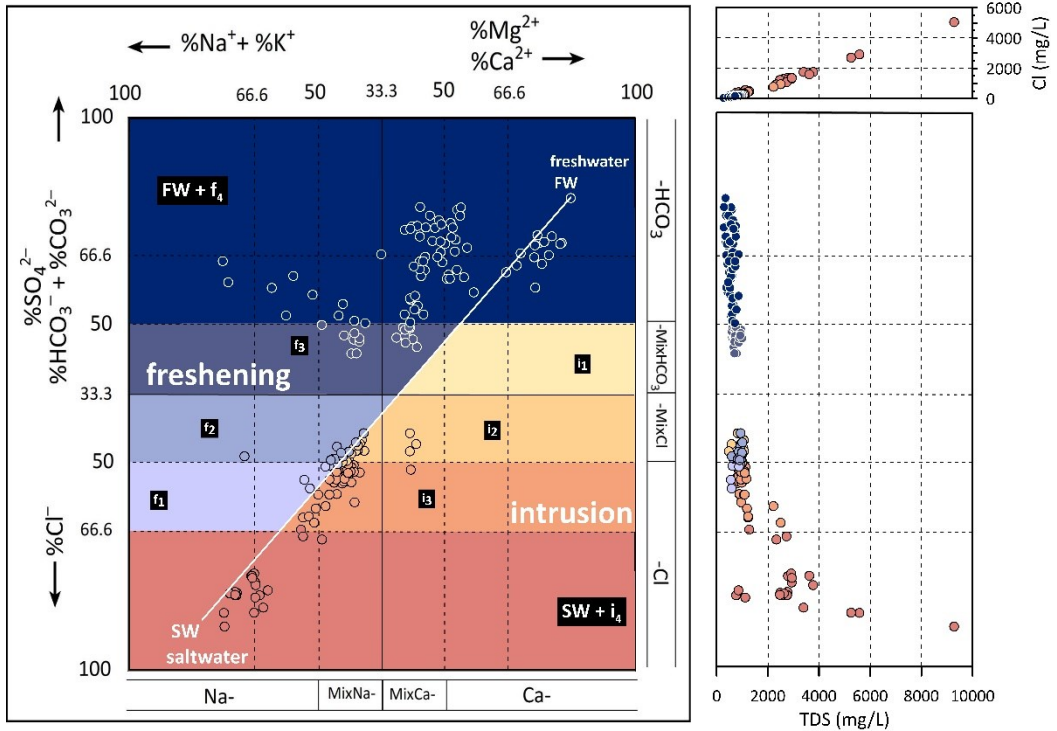
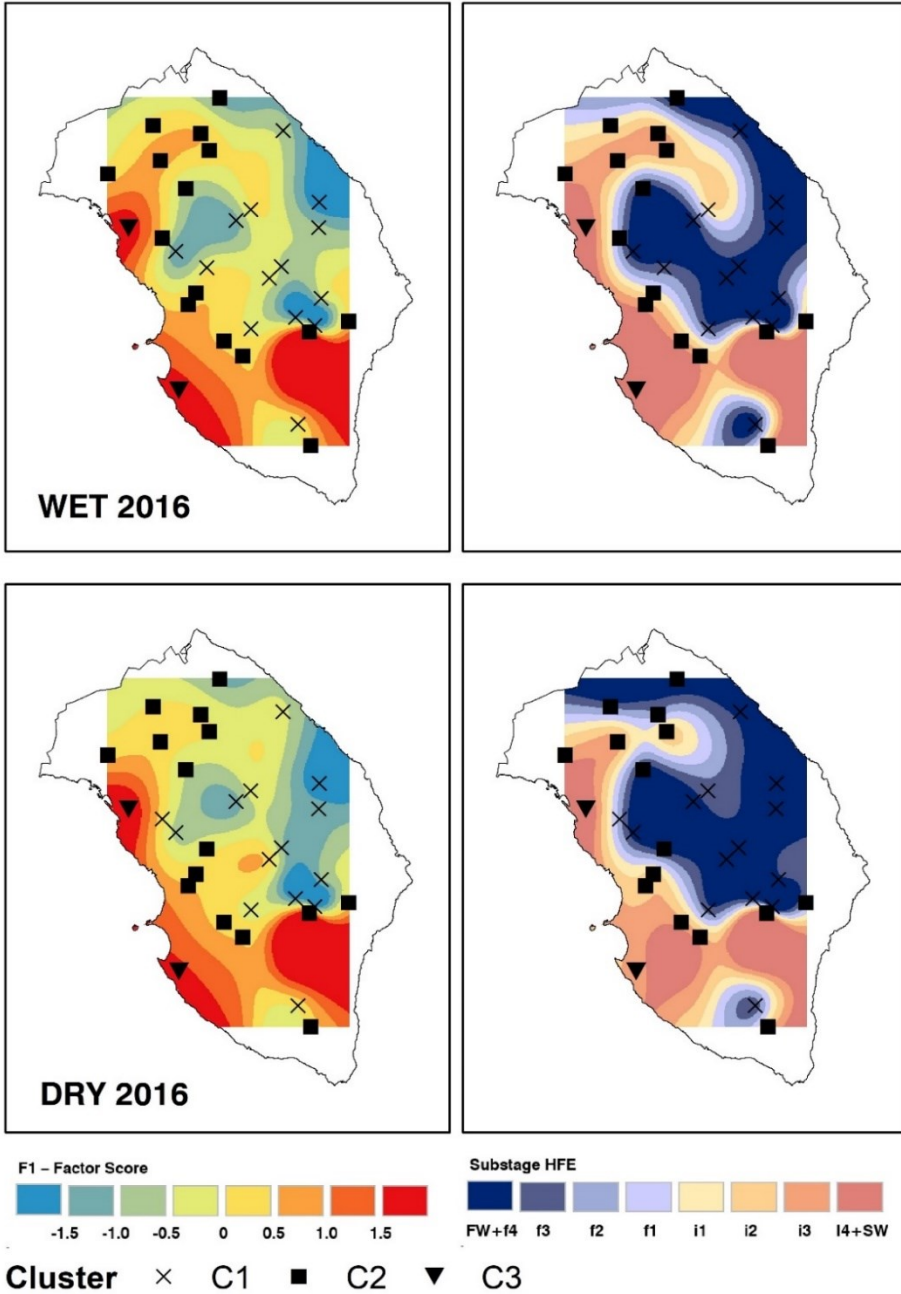
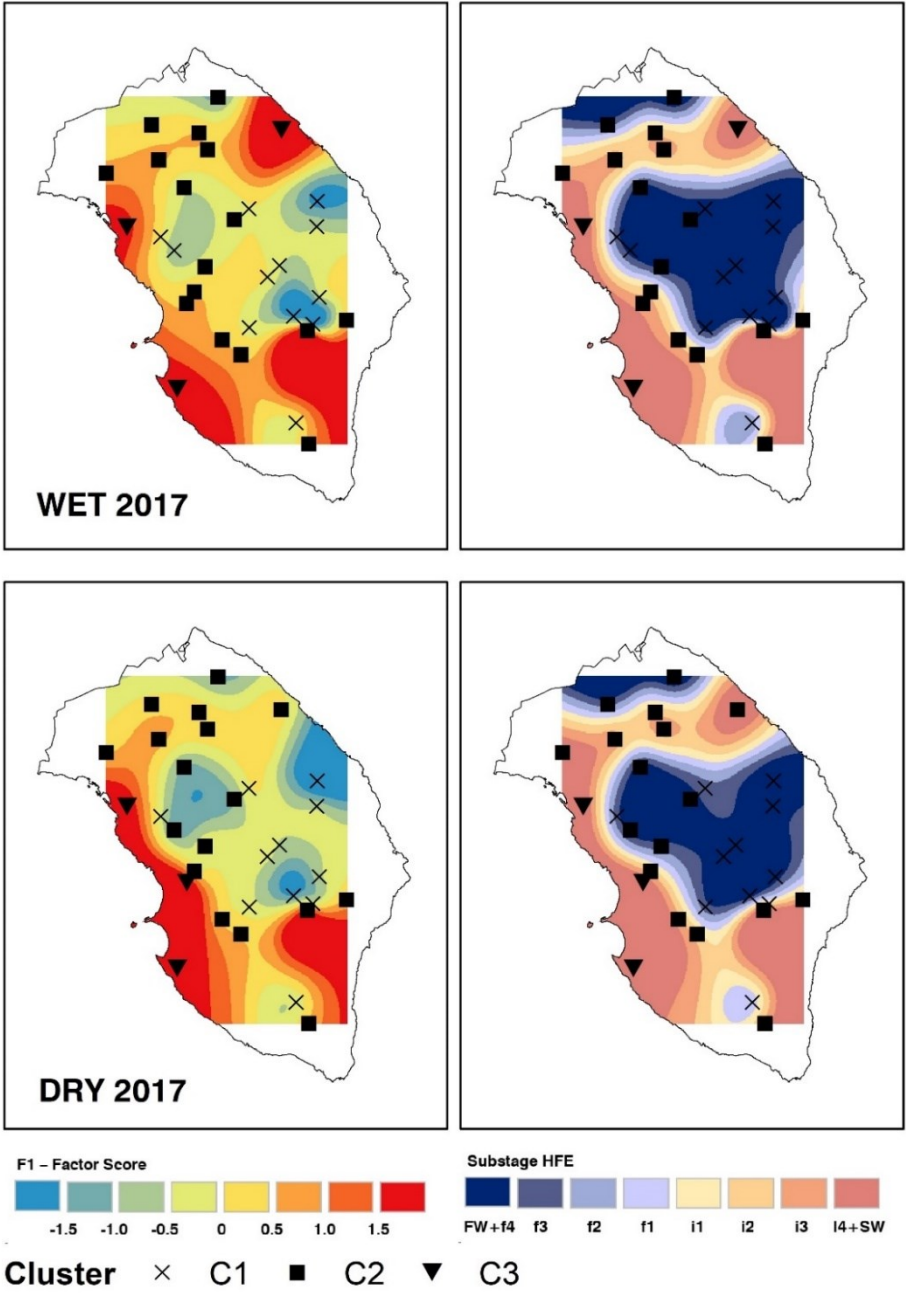


Fig. 40 - HFE-Diagram of the chemical analyses for the six sampling surveys of the Maggiore Project. The diagrams on the right side add information about the TDS and chloride concentration related to the freshening and intrusion HF (Giménez-Forcada, 2014, 2019).





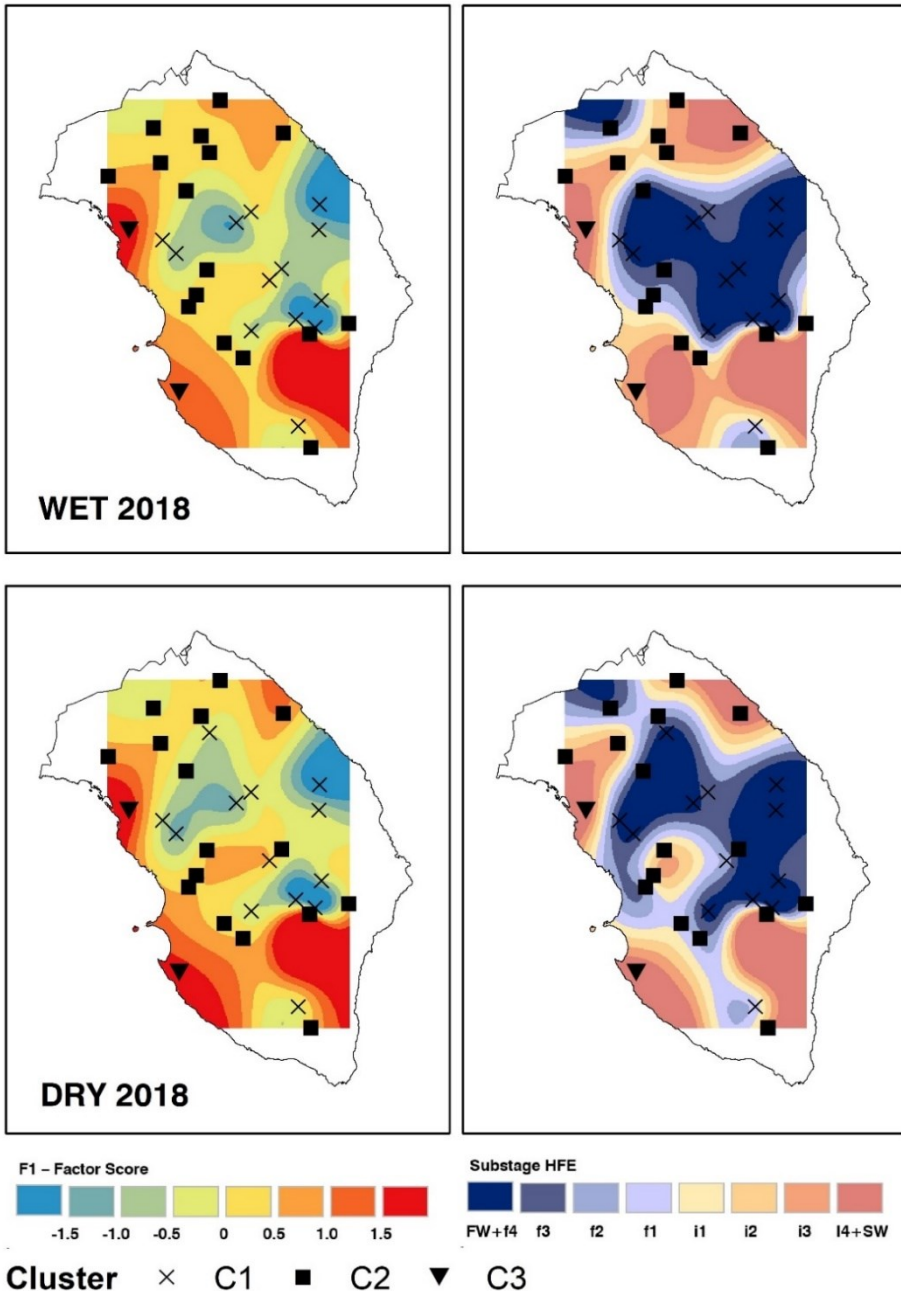


Fig. 41 - Results of FA, HFE-D and HCA are shown for each sampling period of the Maggiore Project. On the left-hand side, the thematic maps are associated with FSs of F1(groundwater salinisation) derived by FA; on the right-hand side, the thematic maps are associated with HFs derived by the HFE-D. On both sides, results are derived by HCA.

### **6.7 Indicator Kriging for the nitrate pollution assessment**

Nitrate pollution in the Salento aquifer had been documented as early as 1990, when Tulipano & Fidelibus (1993) reported concentrations of 100 mg/L in some areas. The accumulation of nitrates in groundwater is typically linked to the intensive use of fertilisers in agriculture. The excessive use and/or inappropriate application of fertilisers containing nitrates, such as ammonium nitrate and potassium nitrate, may result in a surplus of nitrates in the soil, which can subsequently infiltrate through the soil zone and reach the aquifer system, potentially affecting also the quality status and groundwater dependent ecosystems.

Different organisations and countries have imposed drinking water standards for nitrate and other chemical parameters to safeguard public health and the environment. Among these, in 1991, the European Union published the “Nitrates Directive” (91/676/CEE), designating vulnerable zones. The primary goal of this Directive is to mitigate water pollution caused by nitrates of agricultural origin and prevent further contamination. The Directive mandates drafting and implementing specific Action Plans (Aps) for European Union Member States to achieve these goals, focusing on promoting and ensuring responsible and sustainable use of fertiliser in agriculture, through compiled Codes of Good Agricultural Practices (COGAPs). The lack of continuous and long-term monitoring data poses significant challenges for conducting a thorough assessment of groundwater nitrate pollution using numerical models, limiting the spatio-temporal evaluation of pollution. Additionally, it hampers investigations into infiltration mechanisms and the time it takes for the groundwater system to respond to meteorological events.

In addition to the previous issues, the Salento aquifer shows a high degree of hydraulic conductivity anisotropy because of fracturing and karstification processes. The regional groundwater flow pattern further complicates matters. These factors collectively create conditions that promote the development of pollutant-preferential pathways, which can result in diffuse groundwater quality degradation.

For this purpose, chemical surveys from 1995 to 2021 were organised into two-time datasets (1995-2006 and 2007-2021) to focus on a general spatio-temporal overview of nitrate concentrations, aware of the differences between the analysed sampling

campaigns (concerning monitoring frequency, techniques and instruments, water sample storage, transport, and analytical methodologies followed in the laboratory) as well the complexity and the three-dimensional vulnerability of the investigated system.

A preliminary statistical analysis of nitrate concentrations has been conducted on the whole dataset and each well to identify potential trends. Fig. 42 reports the nitrate concentration patterns of six monitoring wells over the last 25 years. Data indicate a considerable increase over time in nitrate concentrations, even exceeding the recommended upper limit threshold (20 mg/L) of the EU's Nitrate Directive. In the early 1990s, it was common to find nitrate concentration values around 40 mg/L (e.g., SAL\_W35, SAL\_W79, and SAL\_W94 wells).

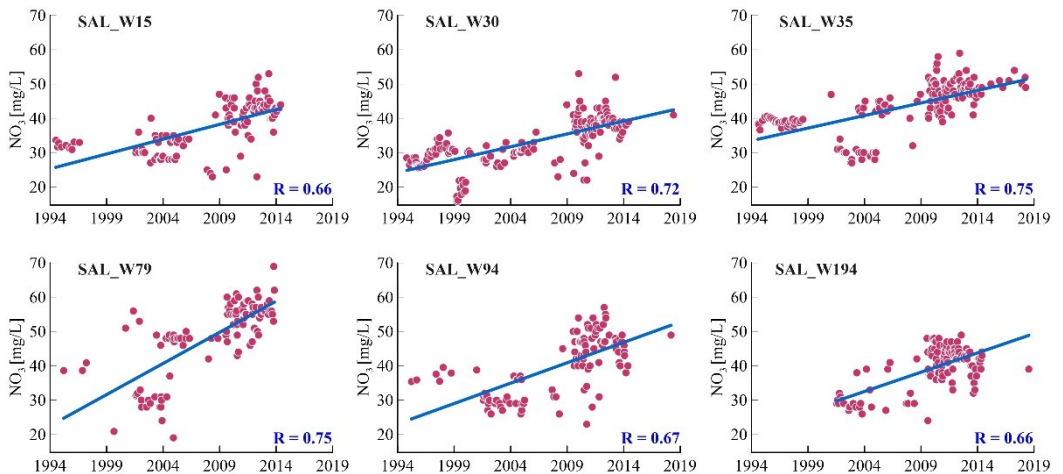


Fig. 42 - Trend of nitrate concentrations sampled from 1995 to 2021 in a few selected monitoring wells. Fig. 43 shows boxplots regarding nitrate concentrations for the same set of monitoring wells shown in Fig. 42, distinguished according to the two reference periods. In 1995-2006, the mean nitrate concentrations are slightly below the regulatory threshold of 20 mg/L, while a noticeable increase in the average nitrate values is observed from 2007 to 2021. During the second reference period, there were some outliers with nitrate concentrations exceeding 70 mg/L. (e.g., SAL\_W79) and in other monitoring wells, although they are not explicitly listed here for brevity.

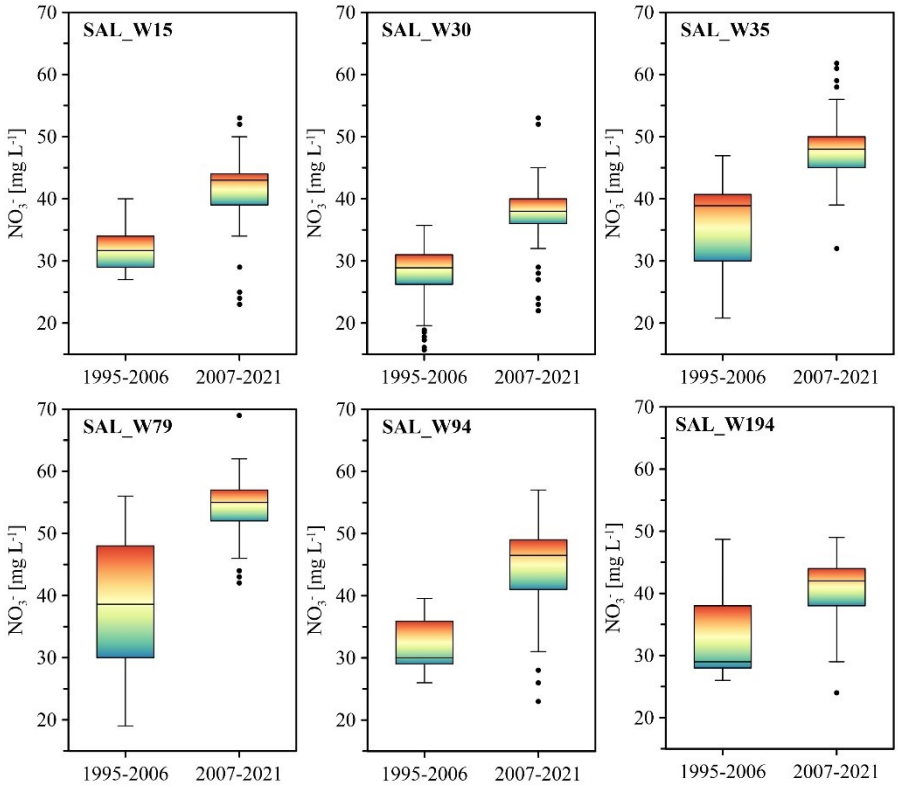


Fig. 43 - Boxplot of nitrate concentrations sampled over the two reference periods in a few selected monitoring wells.

The sampling results of nitrate concentration values monitored during the two reference periods offer valuable insights into the temporal evolution of groundwater quality in Salento. Analysing average and maximum values provides a general sense of the water quality variability in for each monitoring well, even if the occasional exceeding a certain threshold might also be attributed to instruments' errors, data outliers, or operator mistakes. Therefore, when interpreting such data, it is advisable to apply statistical techniques to identify and address potential anomalies or errors that could impact the accuracy of the results, ensuring a reliable and robust assessment. Considering nitrate concentrations in groundwater from 1995 to 2021, the frequency curves were determined for a number of wells (152) with significant samples.

Table 21 lists the frequency at which the set thresholds were exceeded, considering all available wells, and depending on the number of samples collected over time for each

one of them. Nearly all considered wells consistently exceed the 20 mg/L threshold for nitrate concentration. Specifically, 96 out of 152 wells exceed this threshold approximately 90% of the time. However, the results differ when examining the 40 mg/L and 50 mg/L levels. For 48 of the wells (about one-third of the total), the 40 mg/L threshold is exceeded more than 10% of the time. On the other hand, for the higher threshold of 50 mg/L, only about one-sixth of the wells (22) show concentration values above this level for more than 10% of the investigated period. Most worryingly, around one-sixth of those analysed exceed the 40 mg/L limit more than 50% of the time.

Table 21 - Frequency values of nitrate concentration threshold exceedances.

<b>% of Time</b>	<b>20 mg/L</b>	<b>40 mg/L</b>	<b>50 mg/L</b>
≥ 90	96	6	1
≥ 80	102	7	1
≥ 70	109	12	4
≥ 60	120	17	5
≥ 50	126	23	9
≥ 40	130	24	10
≥ 30	134	30	11
≥ 20	134	36	16
≥ 10	138	48	22

IK analyses were performed to directly estimate the local conditional probabilities of nitrate concentrations for the two reference periods using data from available wells (277 and 257 wells, respectively). The three thresholds set for the local estimation of the probability content in an unsampled location are 20, 40, and 50 mg/L. Theme maps representing the spatial distribution of these water quality limits have been compiled and reported with the corresponding standard deviation maps, i.e., a measure of errors. Average nitrate concentrations were calculated for each available well, as the exploratory analysis of the entire dataset did not reveal significant differences between wet and dry months. The probability of nitrate concentrations above 20 mg/L greatly expanded throughout the Salento aquifer compared with 1995-2006, affecting coastal and inland areas with a percentage exceedance of around 90 % (Fig. 44).



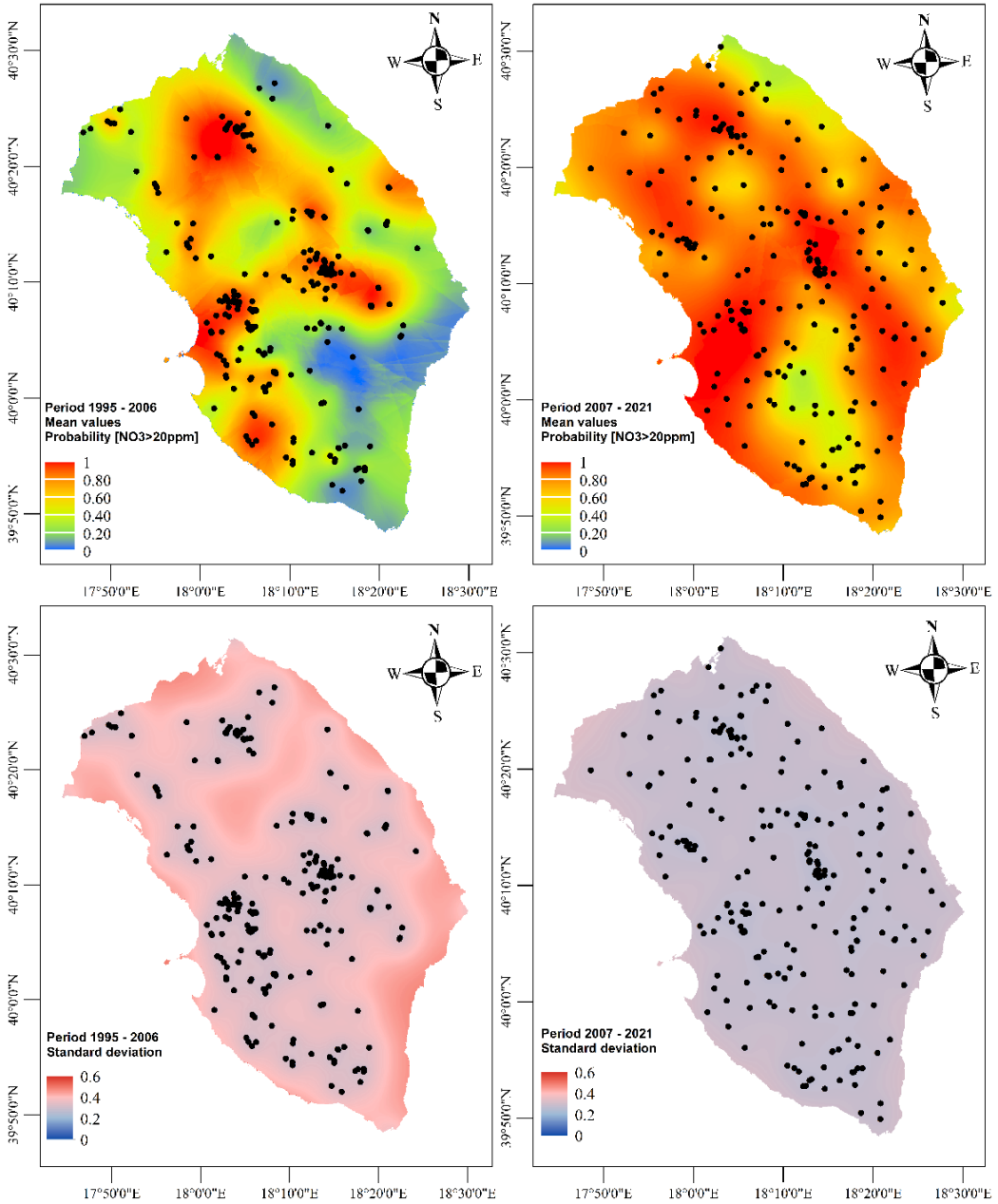


Fig. 44 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 20 mg/L threshold.

The probability of exceeding 40 mg/L evolved significantly between the first and second period from a mean percentage value of 20 % to about 60 % (Fig. 45).

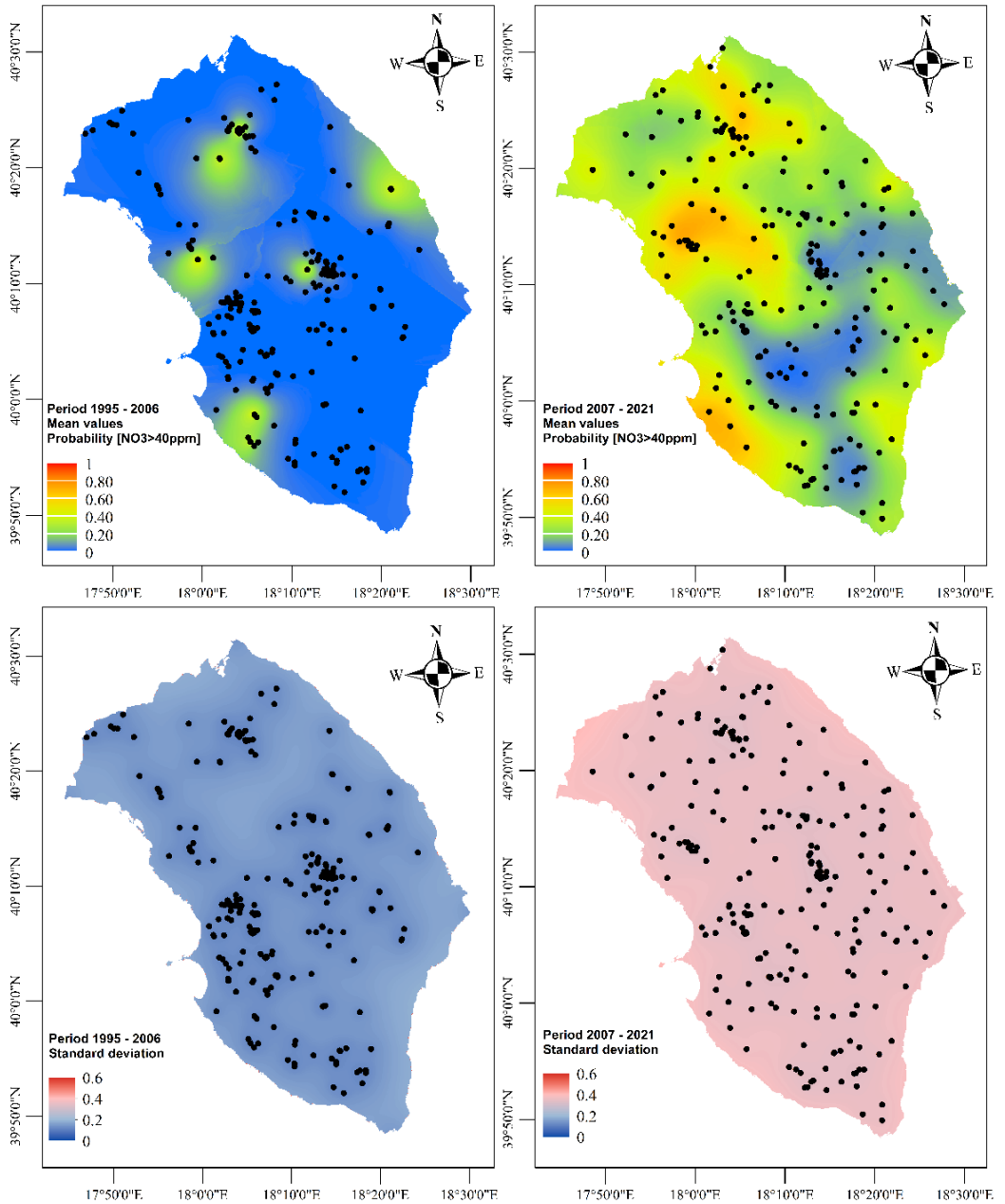


Fig. 45 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 40 mg/L threshold.

A considerable variation is also observed between the two periods, with more than 50 % exceeding the threshold of 50 mg/L for many coastal and inland areas (Fig. 46).

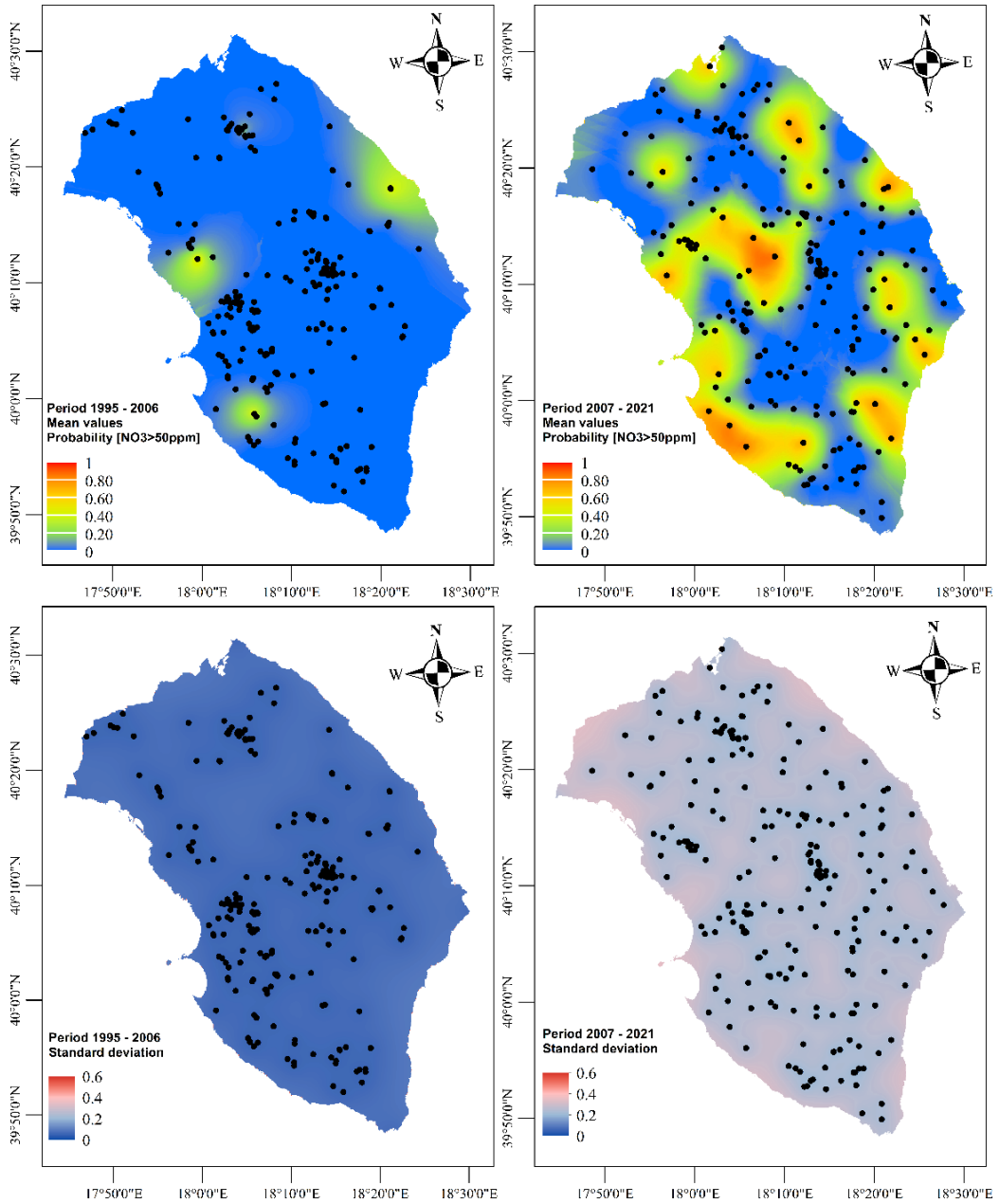


Fig. 46 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 50 mg/L threshold.

Between the first and second period, the areas with a probability of exceeding the threshold value of 20 mg/L increased by approximately 50% of the total aquifer. In the

first period, nearly the entire area of the Salento aquifer had a probability of exceedance lower than 25% for the 40 mg/L threshold. However, this percentage area became more dispersed in the second period, with higher exceedance probabilities. A similar pattern was observed for 50 mg/L, suggesting an overall increase in the possibility of exceedance and a consequent expansion of affected zones between the two periods. Furthermore, compared to the previous period, the areas with a probability of exceeding the threshold of under 25% are more expansive today and potentially more vulnerable. Table 22 summarises the above results.

Table 22 - Percentage areas corresponding to equal intervals of probability exceedance of the thresholds calculated for the two reference periods.

	20 mg/L		40 mg/L		50 mg/L	
	1995-2006	2007-2021	1995-2006	2007-2021	1995-2006	2007-2021
<b>0-25</b>	26.8	0.0	97.3	44.0	97.9	67.8
<b>25-50</b>	27.9	7.3	2.7	43.7	2.1	19.0
<b>50-75</b>	25.7	23.1	0.0	12.3	0.0	11.4
<b>75-100</b>	19.5	69.6	0.0	0.0	0.0	1.7

The results showed that the probability of finding high nitrate concentrations in groundwater changed throughout the study area and over time. Groundwater quality usually depends on anthropogenic factors and land management.

However, without a census of agricultural activities distributed in the area, it was only possible to investigate whether and how land use change over time could influence groundwater pollution. For this purpose, the nitrogen surplus areas identified in two Corine Land Cover maps (<https://land.copernicus.eu/global/>) for the years 2000 and 2018 were compared. Results were discussed in terms of percentage variations of land use areas (Fig. 47). Nitrogen surplus data at different agricultural areas were collected from the Water Protection Plan of the Apulia Region (Water Protection Plan, 2009): 66 Kg/ha for olive groves, 60 Kg/ha for vineyards, 42 Kg/ha for fruit trees, 25 Kg/ha for temporary crops associated with permanent crops and non-irrigated arable crops, and 0 Kg/ha for vegetable and herbaceous crops. The comparison includes red areas, which identify territories where the nitrogen surplus has increased over time, and the

green ones, representing the zones where the nitrogen surplus has decreased. The remaining areas are those where nitrogen surplus remained unchanged over time: (i) orange, pink, and blue areas correspond to zones with high, medium, and low nitrogen surplus, respectively; (ii) grey areas refer to zones with no nitrogen surplus, (iii) while the white ones are those where nitrogen surplus was not detected like artificial areas, wetlands, and water bodies.

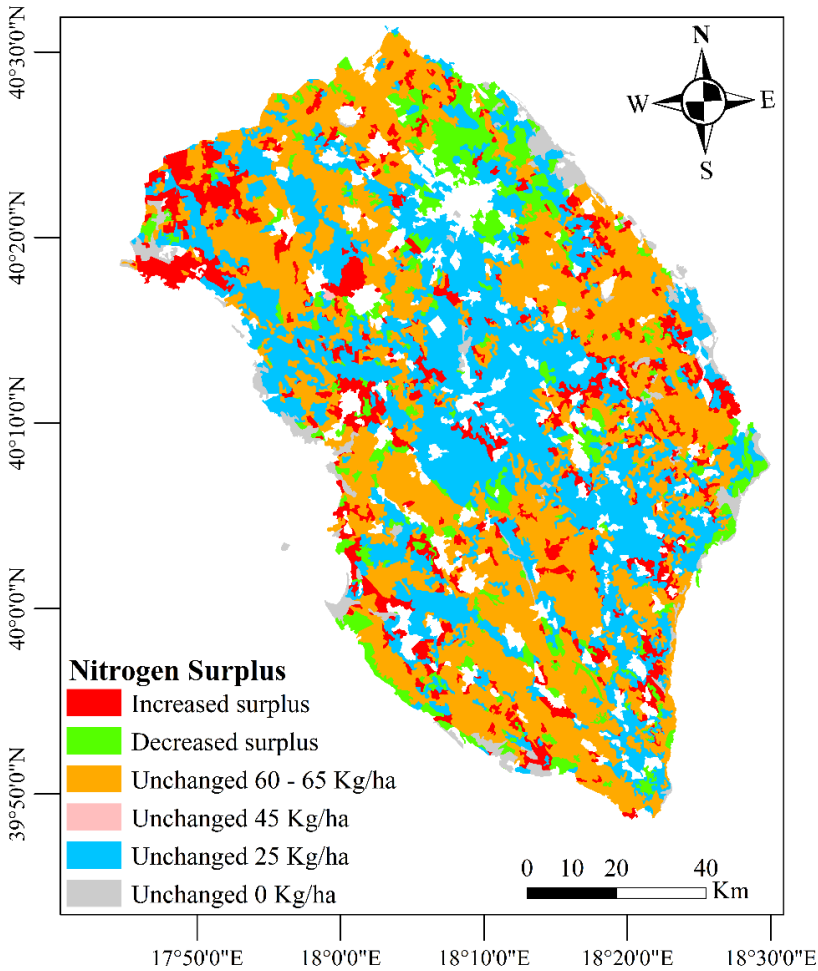


Fig. 47 - Nitrogen surplus changing in agricultural areas from 2000 to 2018

The comparison of the two Corine Land Cover maps highlights that the nitrogen surplus is distributed over the entire territory. On the contrary, IK identifies distinct zones potentially affected by nitrate pollution.

Furthermore, while national data on the nitrogen-based fertilisers used in the Apulia region (ISTAT data: <http://dati.istat.it/index.aspx?queryid=23961#>) reveal a drastic reduction from 2008 onwards (Fig. 48), the IK results suggest an expansion of the areas potentially contaminated by nitrates in 2007-2021 compared to 1995-2006.

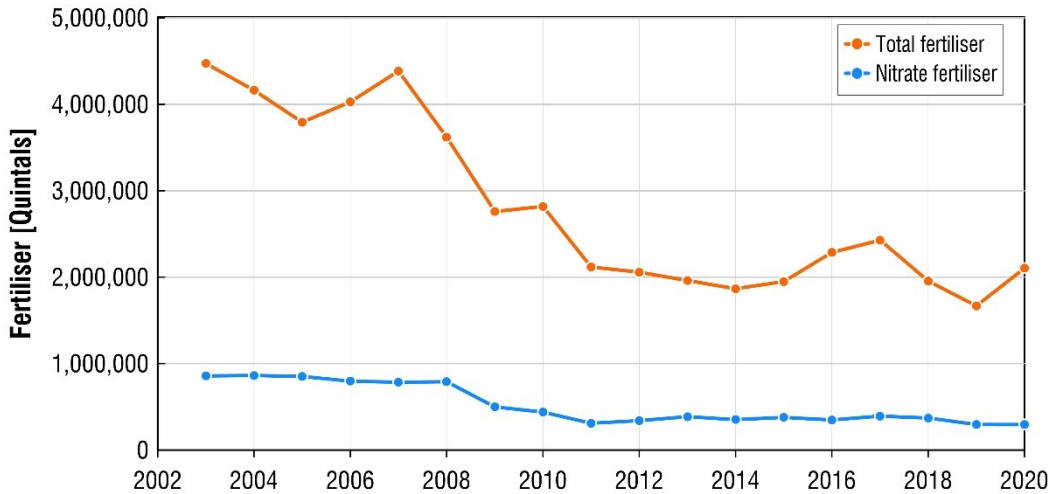


Fig. 48 - Distribution, for agricultural use, of fertilisers (fertilisers, soil conditioners and correctives) from 2003 to 2020 (Source: Istat - <http://dati.istat.it/index.aspx?queryid=23961#>)

Data on land-use change over the period under investigation and the distribution of nitrate-producing fertilisers are apparently unrelated to the nitrate concentrations observed in the second reference period from 2007 onwards. Several studies demonstrated that nitrates leaching from the soil could take decades to reach groundwater due to the transitory deposition and accumulation of nitrates in the unsaturated zone (Wang et al., 2013). The nitrate transport processes should be, in principle, more complex within karst unsaturated zones than in porous ones because in the former, the anisotropy of the hydraulic conductivity adds up to soil lithology, land use, and hydrological conditions as key control factors. Therefore, the time lag detected in many cases (Wang et al., 2013; Kaandorp et al., 2021) could justify the observed trend. If this is true, it means that we will see in the following years a recessionary limb that will bring nitrate concentration values from their peaks to values coherent with the adopted policies of reduction of fertiliser use. This result is crucial advice and would be essential to

be considered for groundwater monitoring planning as it outlines a challenging situation in the management of the resources. Groundwater quality improvement in response to mitigation measures is delayed in time, and the definition of this lag is critical from a policy and monitoring perspective (Bain et al., 2012). For this purpose, it is essential to constantly collect groundwater quality data, attempting to clarify whether what appears today is merely the result of the uncontrolled practices of the past.

## **6.8 Potential climate change in the Salento aquifer**

With the aim of evaluating the projected precipitation and temperature changes in the Salento aquifer, an ensemble of twelve Regional Climate Models (RCMs) driven by several General Circulation Models (GCMs) were collected under the Representative Concentration Pathway RCP4.5. Eight bias-correction (BC) methods were applied at daily time steps, and their results were assessed on monthly and annual time steps, using daily records from 19 and 11 precipitation and temperature (minimum and maximum) stations, respectively, for the period 1960-2005. Potential hydrological and economic impacts of the projected climate trend were finally qualitatively assessed. Two distinct paragraphs report the performance of the bias-correction methods and the future climate trend results.

### 6.8.1 Results of bias correction methods

As described in Chapter 5, an ensemble of 12 RCMs was analysed, and eight BC methods were applied to account for uncertainty due to climate projections. Since the Salento aquifer has a great storage capacity and exhibits slow response to precipitation and temperature variability (Balacco et al., 2022b), we focused on the monthly and annual assessment of the results. Furthermore, although the analyses were carried out for each station, the results were reported in terms of average values of the ensemble for characterising the climate trend on a regional dimension.

The calculated metrics for monthly precipitation and temperature data, i.e., RMSE, RHO, NSE, and BIAS, are reported in the Annex section (Table A3-Table A5). Regarding precipitation, results differ among each RCM and BC method. LS, EQM, LOCI, and PTR generally yield better scores of the assessed coefficients than the other more sophisticated techniques, especially regarding percent bias error. In contrast, temperature outcomes confirm that all methods show good performances in adjusting raw simulated data, and no significant deviations between RCMs can be detected.

Fig. 49 shows the metric results for precipitation, highlighting how all BC methods slightly improve the raw RCM datasets; there are still significant deviations in reproducing the observed characteristics depending on the analysed RCM and BC technique.



In general, the selected bias-corrected RCMs are sufficiently representative of the Salento precipitation pattern, although a significant variation exists among them compared to the observed time series. Few RCMs (i.e., AL\_CN and WF\_IP) were excluded entirely from successive precipitation analyses because their deviation was significantly inconsistent compared to the observations, even after applying the BC methods. As reported in Fig. 49, these two RCMs are indeed those with the highest variability in terms of metrics. It is worth noting that the performance of each BC method varies among all RCMs, making the selection of the best technique undetectable.

Fig. 50 illustrates the Taylor diagrams compiled by raw and bias-corrected RCM precipitation from 1971-2005 on a monthly interval. They are normalised to the standard deviation of the observed data and expressed as mm/month. These graphs simultaneously control the standard deviation and the correlation coefficient between simulated (raw or bias-corrected) and observed data. Thus, bias-corrected RCMs showing high correlation coefficients and standard deviations closer to one represent datasets more similar to the observed patterns.

Regarding raw data, except for RA\_IC, RC\_IC and RC\_MP, all the other raw RCMs data tend to underestimate monthly precipitation over 1971-2005 (standard deviation less than one). Specifically, the standard deviation varies between a minimum of 0.58 mm/month (RC\_IP) to a maximum of 1.05 mm/month (RC\_IC). Correlation coefficients range between a minimum of 0.15 (RC\_IP) to a maximum of 0.29 (CC\_CN and RC\_CN). After applying BC methods, the standard deviations increase to around one (0.58 mm/month (RC\_IP) and 1.05 mm/month (RC\_IC)), and the correlation coefficients slightly improve, especially for LS, EQM, PQM, and PTR.

DQM and QDM methods perform worse than the others, with monthly precipitation amounts higher than the observations.

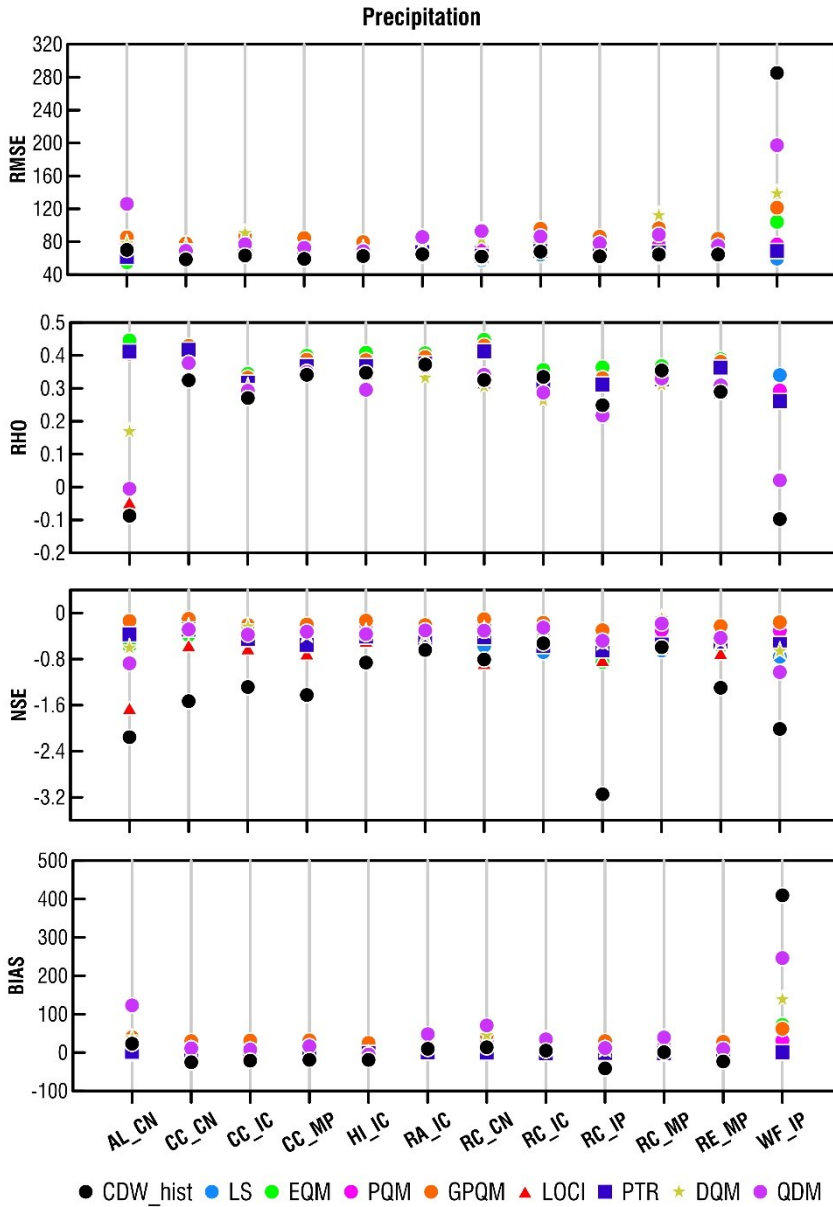


Fig. 49 - Statistical performance evaluation of the raw and bias-corrected RCM outputs against the monthly observed rainfall for the 12 RCMs in the Salento aquifer from 1971 to 2005. The RMSE, RHO, NSE and BIAS are reported. The coding of the analysed scenarios is reported in Section 5.1.1.

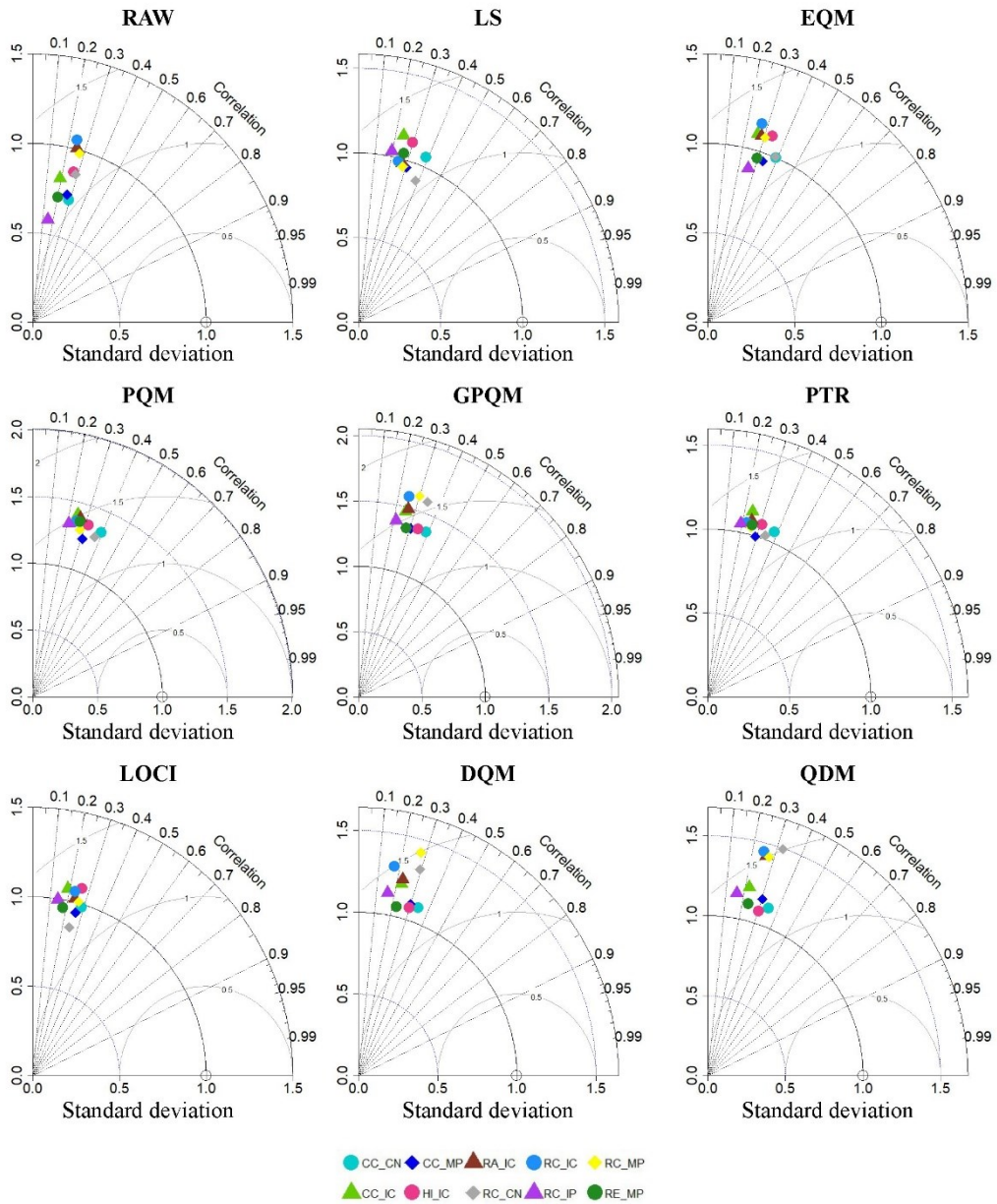


Fig. 50 - Taylor diagrams of monthly raw and bias-corrected RCMs precipitation data for the period 1971-2005.

Fig. 51 illustrates the comparison of the monthly precipitation mean values of raw and bias-corrected simulated data with the historical observations (dashed line) from 1971-2005. As detected in the Taylor diagrams, the monthly precipitation averages of raw

RCM data are generally lower than the observations, especially during autumn. The variability and uncertainty of each raw RCM compared to the observed time series are visually detectable. Although the seasonality is largely reproduced, the uncorrected monthly average precipitation of the selected RCMs significantly deviates from the observed ones, overestimating spring-summer depths, and underestimating autumn-winter ones.

After BC, the monthly bias-corrected precipitation of each ensemble member is comparable with that of the observation data, with the best fit presented for LS, EQM, and PTR techniques. GPQM, DQM, and QDM exhibit more significant variability than the observed trend, with values usually overestimating monthly precipitation, especially during the wet period.

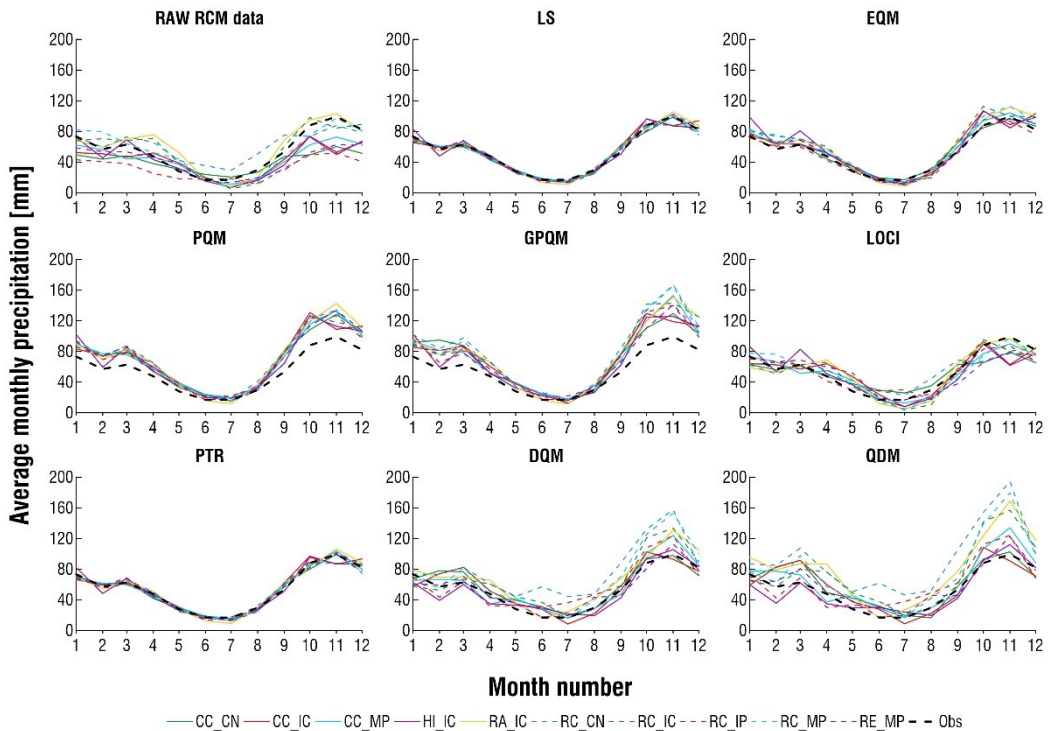


Fig. 51 - Average monthly precipitation of raw and bias-corrected RCM data for 1971-2005.

Improving simulated raw precipitation data is more complex than temperature in the bias correction context. In fact, in the case of temperature datasets, all RCMs and BC

methods perform satisfactorily, with bias-corrected results in agreement with the observation time series.

The results of the metrics (i.e., RMSE, RHO, NSE and BIAS) for monthly maximum temperature are reported in Fig. 52. The statistical performance coefficients explain a better bias correction of simulated data compared to precipitation, without significant differences among RCMs and BC techniques. RMSE coefficients are close to zero, indicating a low deviation between simulated and observed data. The already high RHO correlation coefficients for the raw simulated data improve after the application of BC methods. NSE coefficients range around zero, demonstrating the goodness of all BC methods. Finally, the BIAS percentage is very low against monthly precipitation, for which higher values and variability among each technique were obtained.

Contrarywise to precipitation, all 12 RCMs were considered for the temperature dataset. Fig. 53 shows the Taylor diagrams of maximum temperatures. Similar considerations could be discussed for minimum temperatures. All raw RCMs display high values of correlation coefficients, ranging between 0.94 and 0.96 and standard deviations varying from a minimum of 0.80 °C/month to a maximum of 1.13 °C/month. Applying BC methods improves the standard deviation more than the correlation coefficient. No significant difference in the performance between RCMs and BC methods was identified.

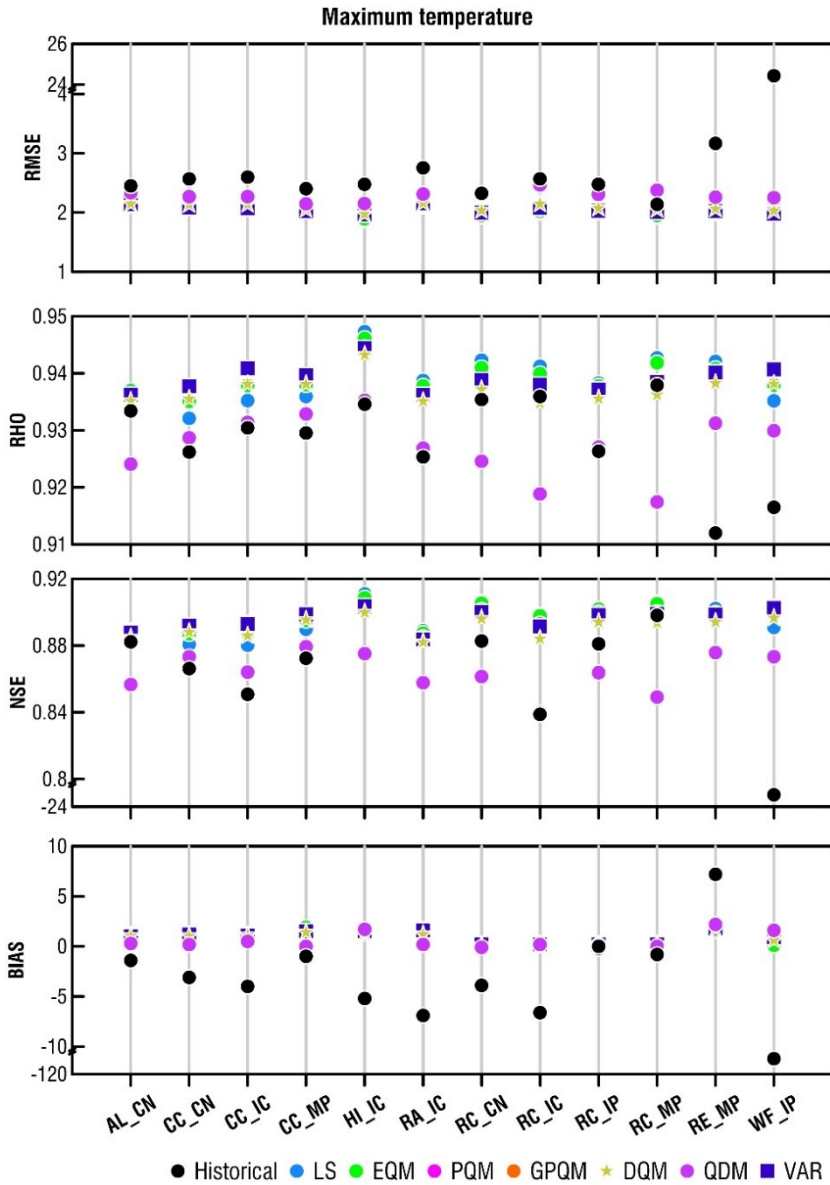


Fig. 52 - Statistical performance evaluation of the raw and bias-corrected RCM outputs against the monthly observed maximum temperature for the 12 RCMs in the Salento aquifer from 1971 to 2005. The RMSE, RHO, NSE and BIAS are reported. The coding of the analysed scenarios is reported in Section 5.1.1.

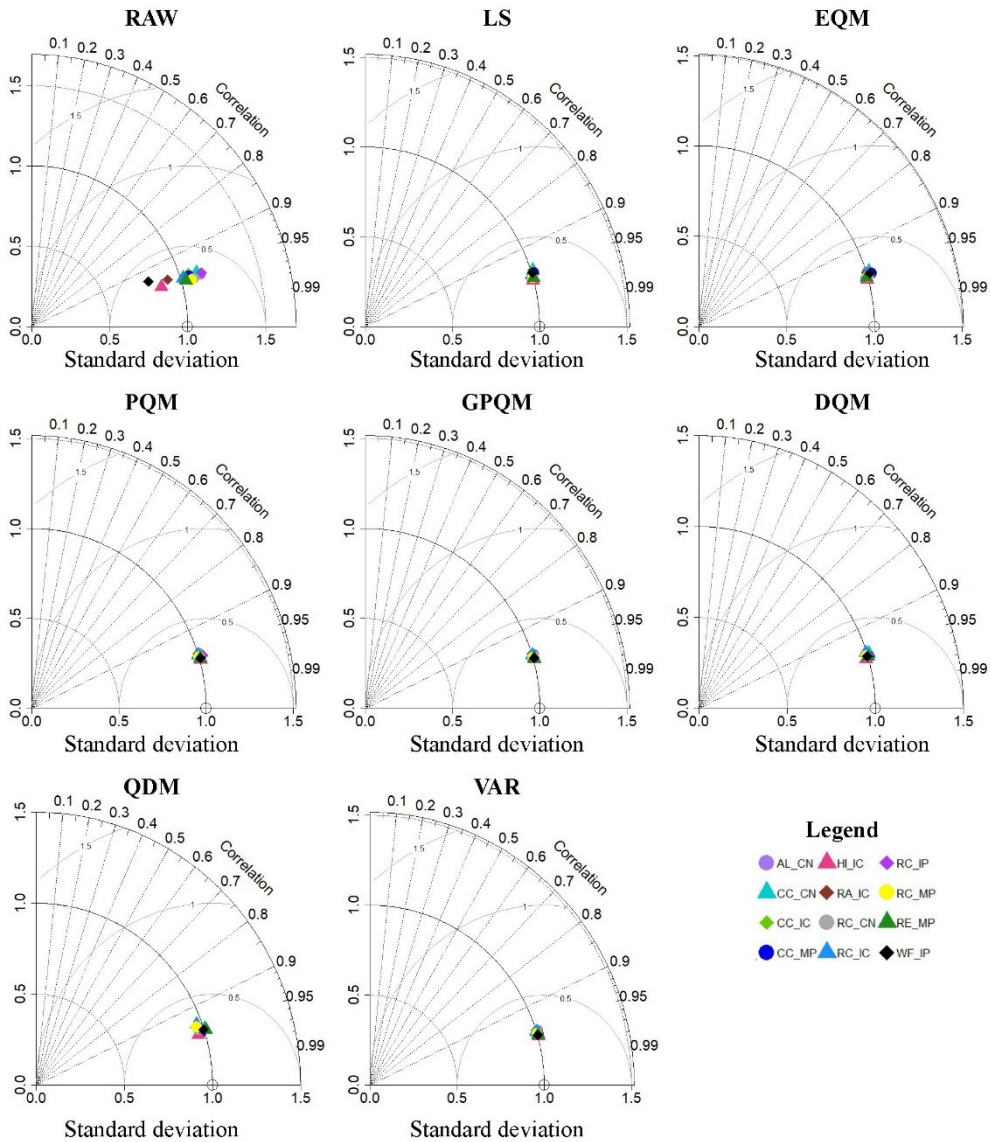


Fig. 53 - Taylor diagrams of monthly raw and bias-corrected RCMs maximum temperature data for the period 1971-2005.

Fig. 54 provides a comparison of raw/bias-corrected monthly averages of the maximum temperature of each RCM and observations. No specific difference among RCMs or BC methods is evident, as all techniques considerably reduce the bias.

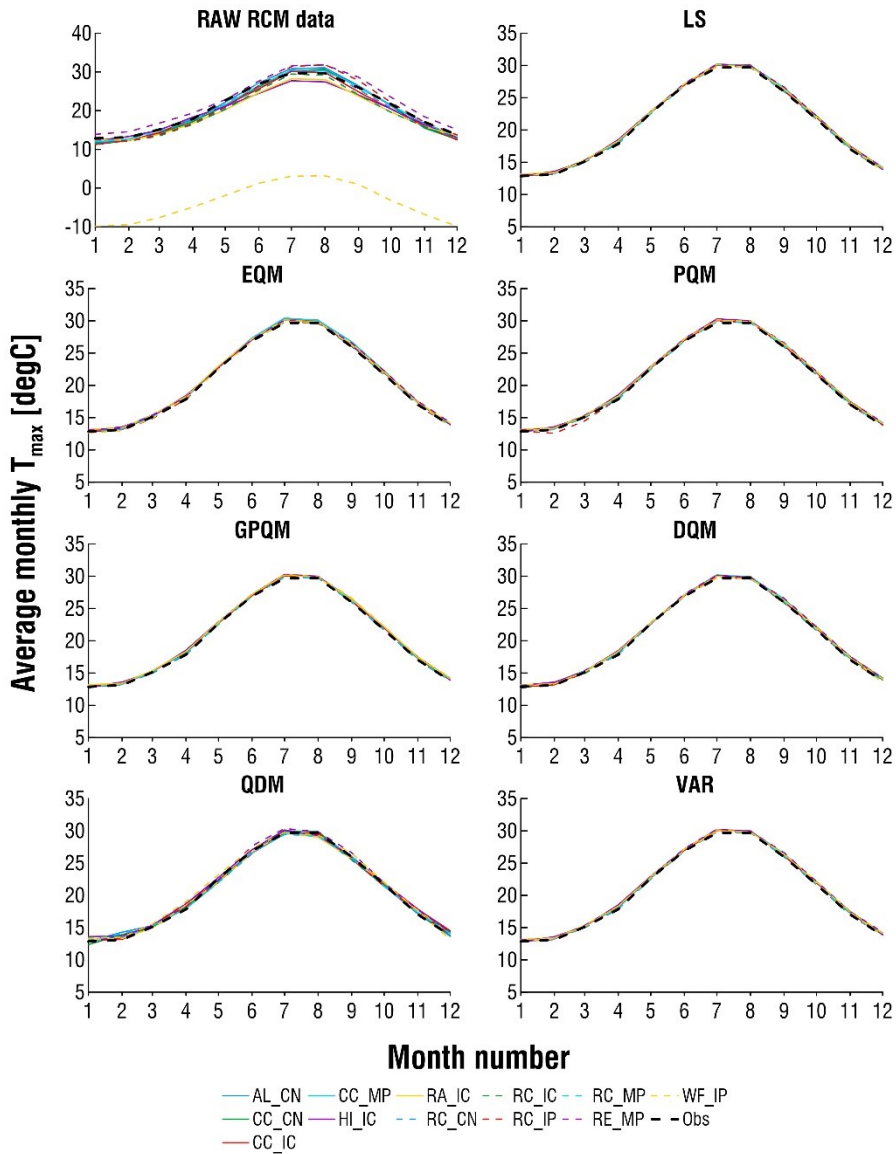


Fig. 54 - Average monthly maximum temperature of raw and bias-corrected RCM data for 1971-2005. Compromise programming (CP) results were used to select the most suitable bias-corrected data for assessing projected precipitation and temperature trends. Fig. 55 highlights the results of CP for precipitation in terms of mean values of  $L_{cp}$ . The most efficient methods are LS, PTR, and LOCI. Nevertheless, LOCI was excluded from the



future trend analysis because there are still some considerable biases in monthly average precipitation after BC, as illustrated in Fig. 51.

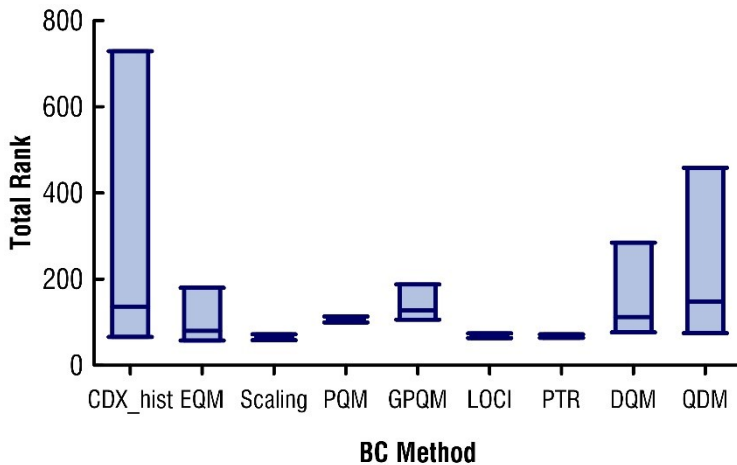


Fig. 55 - Average values of compromise programming results for precipitation.

The LS and the PTR algorithms could finally be considered the most efficient BC methodologies for precipitation time series for the case study, since quantile methods (i.e., GPQM, DQM, and QDM) tend to overestimate monthly precipitation, especially during the wet period. On the contrary, in the case of temperature datasets, all methods perform satisfactorily, and the results agree with the observation time series.

Addressing the persistent uncertainties in RCMs requires a comprehensive approach combining BC methods with multi-model ensemble techniques, as Lyra & Loukas (2023) discussed. However, many BC techniques are specifically designed for particular characteristics of catchments, primarily at the basin scale, making them less suitable for alternative contexts (Tumsa, 2022). The challenges of selecting the appropriate BC method become even more complex in arid and semi-arid regions, where precipitation patterns are characterised by scarcity, irregularity, and randomness (Mendez et al., 2020). BC methods often tend to overestimate annual mean values in these regions, with the degree of overestimation varying depending on the specific RCMs used. Furthermore, uncertainties arise regarding the ability of these methods to accurately represent climate change signals and predict future extreme weather events. In this context, it is noteworthy that climate change signals are chiefly sensitive to RCMs, given

that LS and PTR for precipitation and LS and VAR for temperature generally yield similar results. Nonetheless, there are geographical regions where the choice of BC methods significantly influences climate change patterns and extreme events assessments. For example, Tefera et al. (2023) illustrated in the Jemma sub-basin of the Upper Blue Nile Basin that the performance of Linear Scaling and Distribution Mapping techniques varies in extreme event characterisation and climate change signal representation across different RCMs and emission scenarios.

Despite advancements in this field, dealing with the compounded uncertainties introduced by the modelling process remains a challenging task. This process encompasses decisions regarding GCM and RCM simulations, evaluations of emission scenarios, BC method selection, the geographical area under investigation, and the quality of observed datasets (Noto et al., 2023b).

### 6.8.2 Future trends and potential climate change over the Salento aquifer

Precipitation and minimum and maximum temperature datasets provided by each raw and bias-corrected RCM for the entire projection period 2006-2100 were divided into two 30-year intervals (2031-2060, 2071-2100) and compared with the reference period 1971-2005 on monthly and annual intervals. The correction factors assessed in the reference period for each BC method were used to correct the bias in the future, as usually considered in climate studies.

The mean annual amount and the change signal were calculated for each RCM in the two 30-year periods from the observed reference to evaluate the effect of the selected methods on future projections. Specifically, the climate change signal refers to the percentage difference between future raw and bias-corrected RCM output for precipitation and temperature deviations from the reference period. Results reported in Table 23 and Table 24 refer to the average ensemble of all station time series.

Fig. 56 and Fig. 57 show the boxplots and climate change signals of the annual precipitation of raw and bias-corrected RCMs. It can be noted that, among the others, LS and PTR adapt simulated data of the selected RCMs to local observations regarding monthly mean precipitation during the reference period (1971-2005). However, when evaluating the future projections, the raw RCMs consistently project a reduction in the mean value

of both periods, except for RC\_CN and RA\_IC, which show higher mean precipitation (Fig. 56). In contrast, bias-corrected simulated data overestimate the mean annual amount in most cases, as exemplified by CC\_CN, CC\_IC, CC\_MP, HI\_IC, RC\_IP, RE\_MP. However, both methods appear to be suitable for preserving the climate change signal of most of the RCMs in 2031-2060 and 2071-2100 (Fig. 57). There are a few exceptions, such as CC\_CN, where LS and PTR cause a shift from a negative climate change signal in the raw data to a positive one in 2071-2100. Another exception is RA\_IC, which exhibits a changing trend from a positive signal to a negative one in the 2031-2060. Generally, it can be observed that all bias-corrected RCMs tend to damp or attenuate the climate change signals when compared to the raw data.

Data-driven methods for qualitative and quantitative characterisation of coastal aquifers

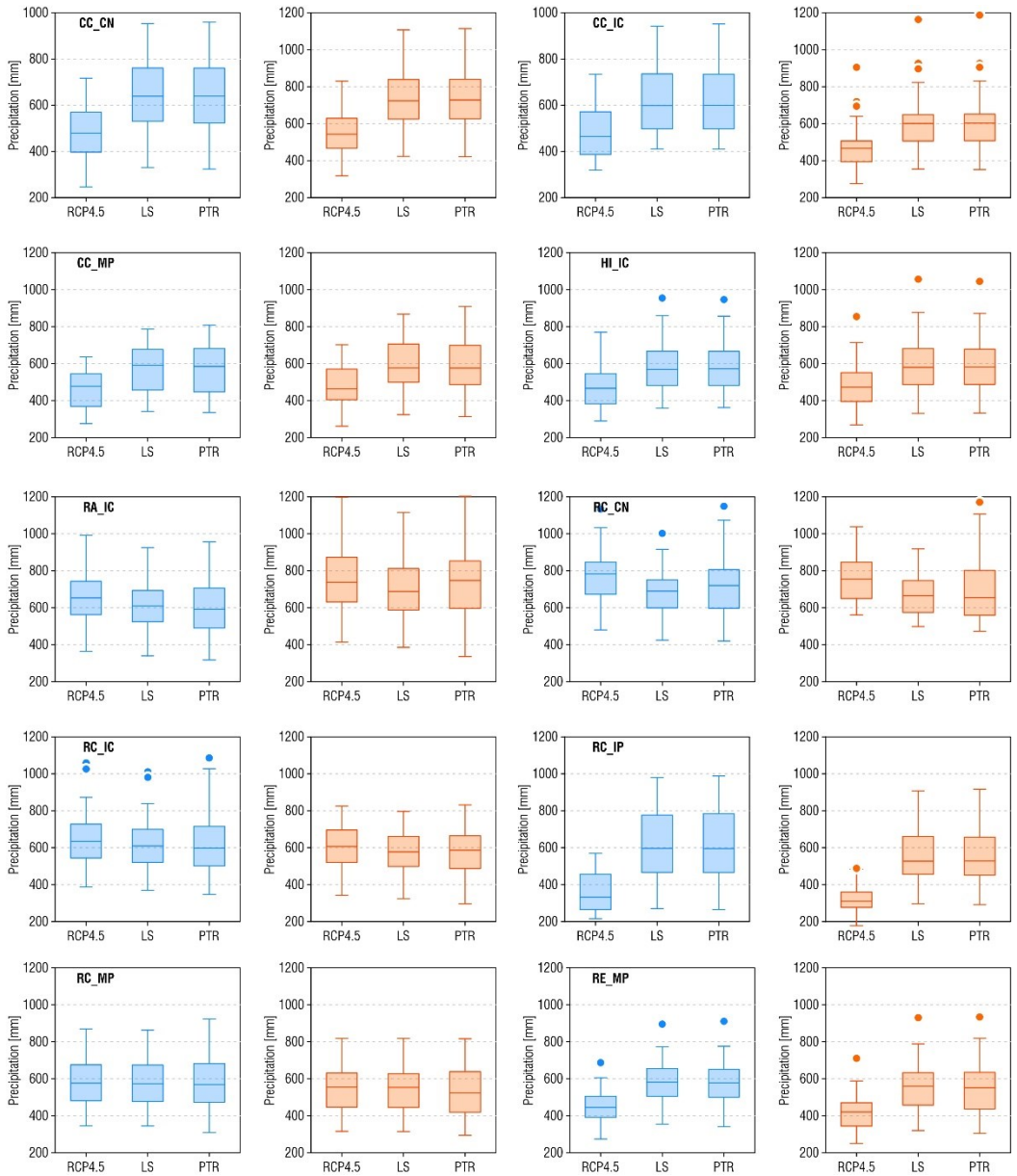


Fig. 56 - Boxplots of the annual precipitation of RCM simulations with (LS, PTR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange). In the boxplots, whiskers indicate the minimum and maximum value of precipitation; the horizontal lines refer to the 25th percentile, median, and 75th percentile from the bottom to the top of each boxplot, and the point symbols represent outliers.

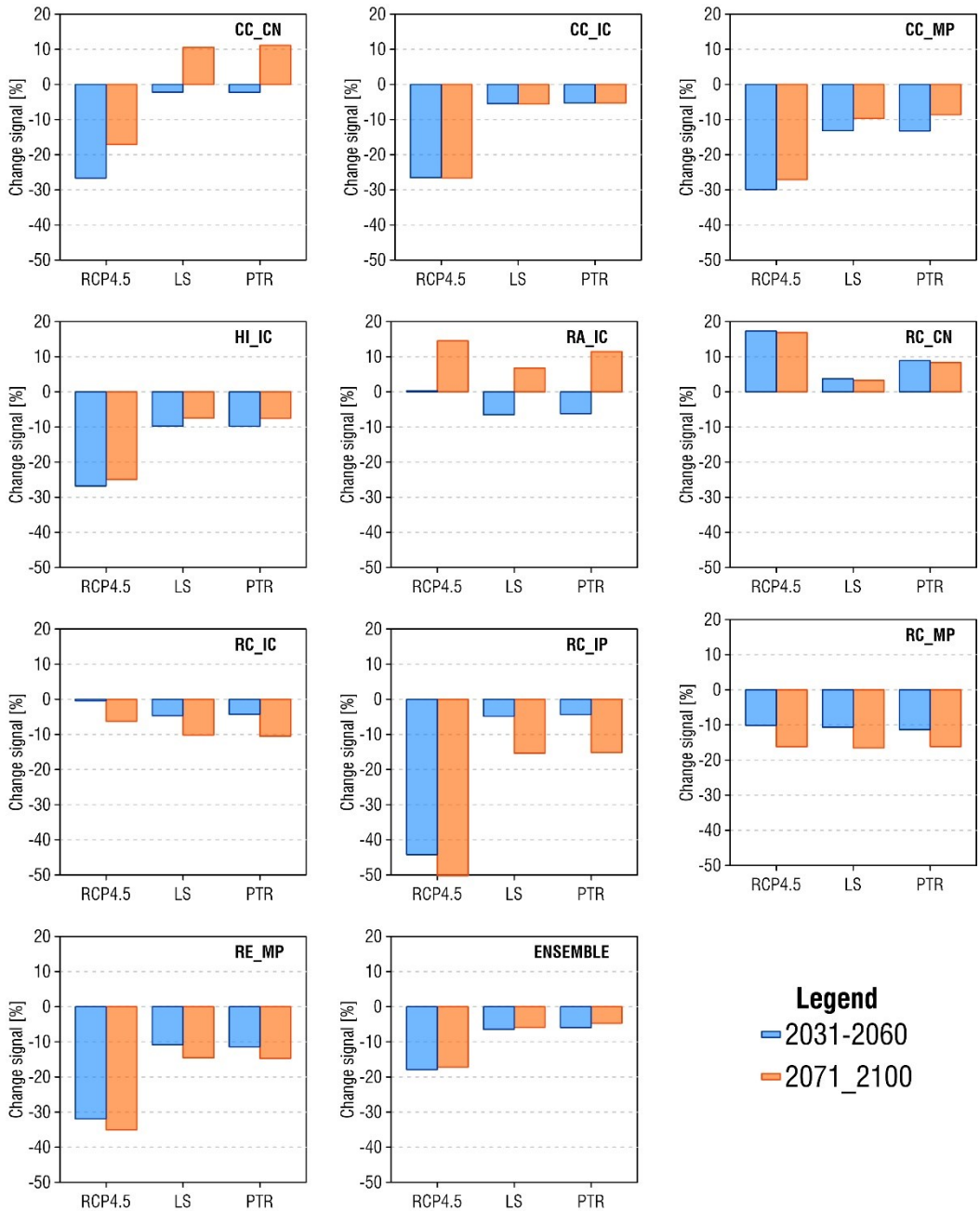


Fig. 57 - Climate change signals of annual precipitation of RCM simulations with (LS, PTR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange).

Fig. 58 shows the annual precipitation variability (blue line) resulting from the average of the 10 RCMs, bias-corrected with LS and PTR for 2031-2060 and 2071-2100. The results indicate a wide interannual variation of average precipitation. At the same time, the 95% confidence interval (light blue area) range of about 200 mm reveals the considerable uncertainty around projected annual precipitation, introduced by the different climate realisations presented from the different RCM-GCM combinations. The data reveal that the annual precipitation trend is relatively stable for the near to medium future, with no clear indication of a significant increase or decrease. This outcome was consistent across both LS and PTR methods (orange lines in Fig. 58). On the contrary, a slightly positive trend in annual precipitation became more noticeable in 2071-2100. However, bias-corrected data showed consistently lower values for both periods when compared to historical observations, pointing towards a potential decrease in precipitation over time (Table 23).

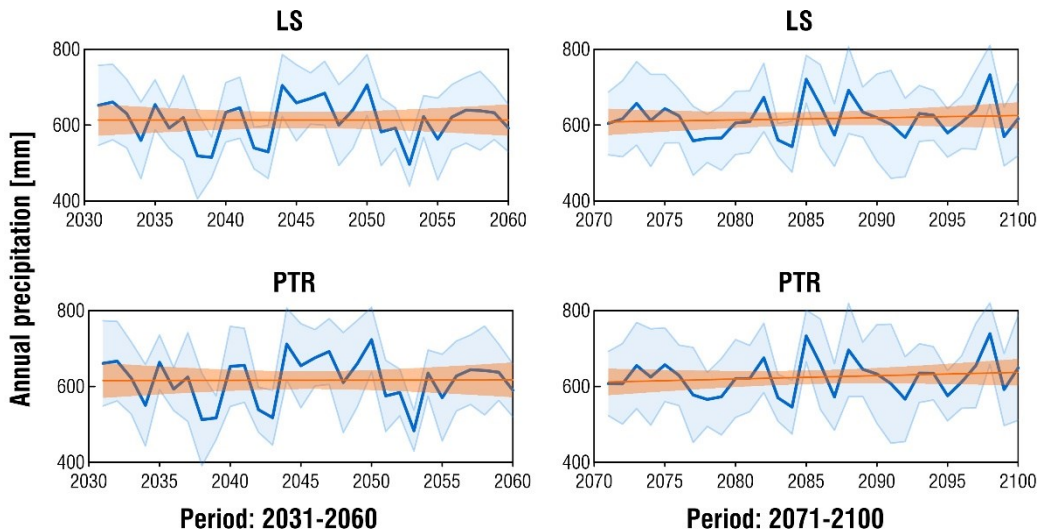


Fig. 58 - Average annual precipitation (bold blue line) and 95% confidence interval (light blue area) based on the data of the 10 RCMs, bias-corrected with LS and PTR methods for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval.

The average annual precipitation resulting from bias-corrected data with LS in 2031-2060 was 613.2 mm, while the corresponding value of PTR was 616.5 mm, thus indicating a very low deviation between the two BC methods. In 2071-2100, the average

annual precipitation resulting from bias-corrected data were 616.8 mm and 624.5 mm with LS and PTR, respectively. Thus, precipitation results are lower than the observed average annual precipitation in the historical period (1971-2005), which equals 655.3 mm. Specifically, the annual precipitation for the Salento study area shows a decrease of 6.4 % and 5.9 % for LS and 5.9 % and 4.7 % for PTR in 2031-2060 and 2071-2100, respectively, when compared to the historical period (1971-2005).

Table 23 - Summary of annual precipitation change scenarios.

BC	Period	Annual trend	Annual precipitation	Average annual precipitation [mm]	Average annual precipitation change [%]
Historical	1971-2005	+		655.3	na
LS	2031-2060	-		613.2	-6.4
PTR	2031-2060	+		616.5	-5.9
LS	2071-2100	+		616.8	-5.9
PTR	2071-2100	+		624.5	-4.7

The monthly precipitation change is mainly negative in the short and long term, with some exceptions during the late spring-early summer period (April-June), which appears to be increasing (Fig. 59).

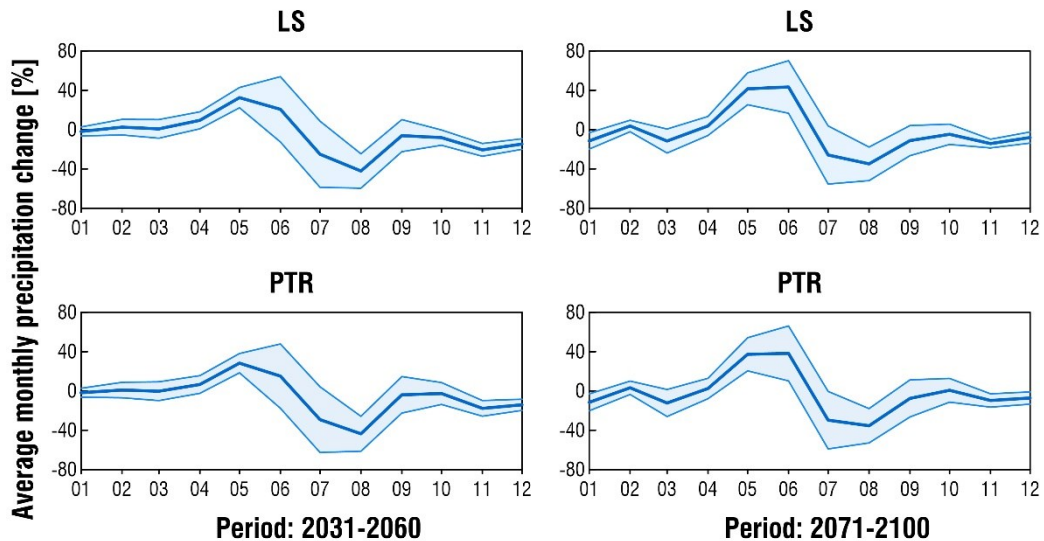


Fig. 59 - Monthly precipitation change of 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides) for LS and PTR, respectively. The blue line refers to the average monthly value, and the light blue area indicates the 95% confidence interval produced from the 10 RCMs.

For brevity, minimum and maximum temperature results are reported referring to LS and VAR methods since deviations among each method are negligible. In contrast to precipitation, different results can be detected for minimum and maximum temperature in terms of mean value, for which all raw simulations project an increasing trend compared to the reference period, except for RA\_IC, RC\_CN, and RC\_IC, which display a slight reduction in minimum temperature. However, when referring to bias-corrected data, LS and VAR often overestimate the mean annual temperature. The two BC methods preserve the positive climate change signals in both periods, except for RA\_IC, RC\_CN, RC\_IC, and RC\_MP, for which BC methods triggered a change from negative to positive climate change signals.



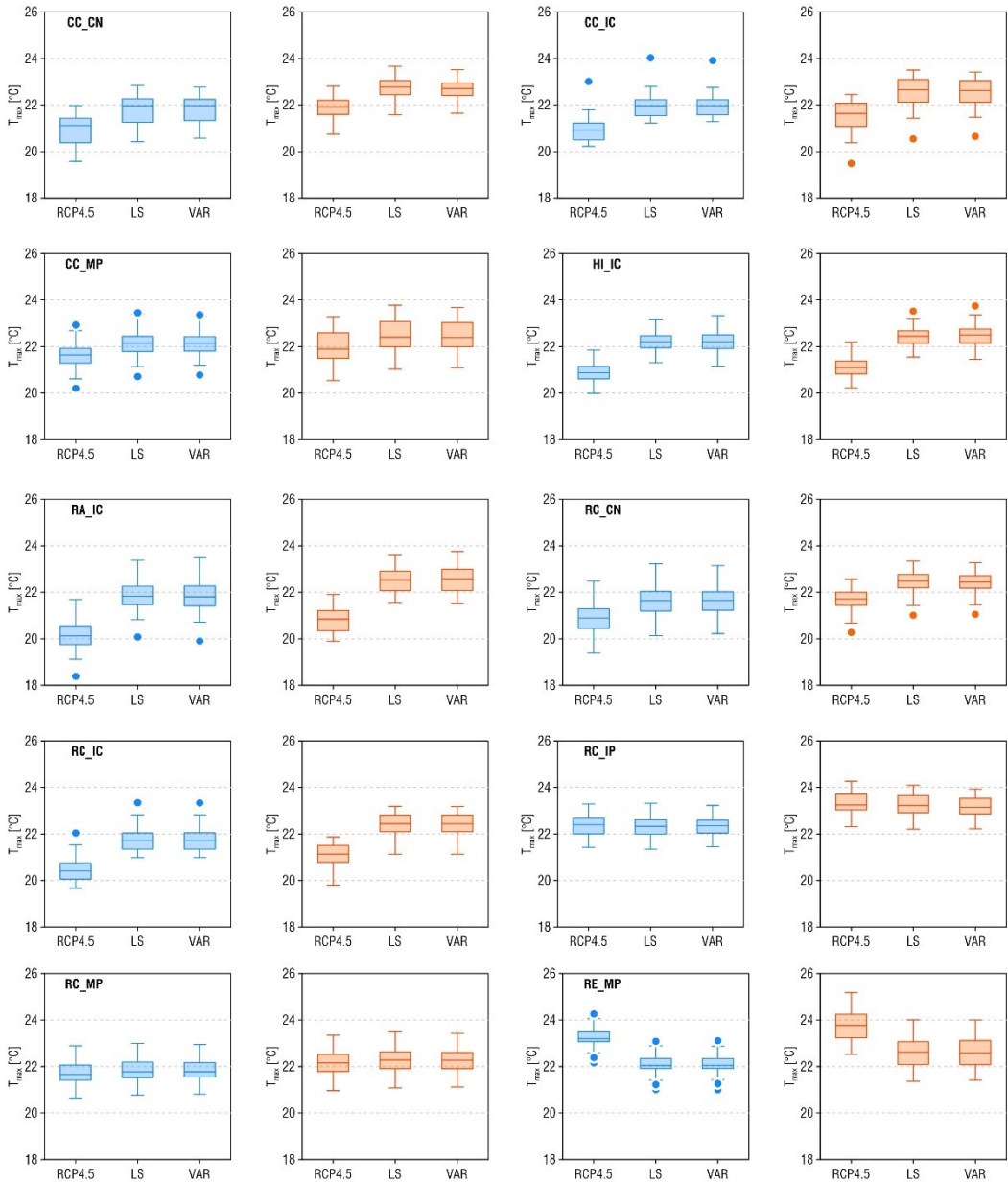


Fig. 60 - Boxplots of the maximum temperature of RCM simulations with (LS, VAR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange). In the boxplots, whiskers indicate the minimum and maximum value of  $T_{max}$ ; the horizontal lines refer to the 25th percentile, median, and 75th percentile from the bottom to the top of each boxplot, and the point symbols represent outliers.

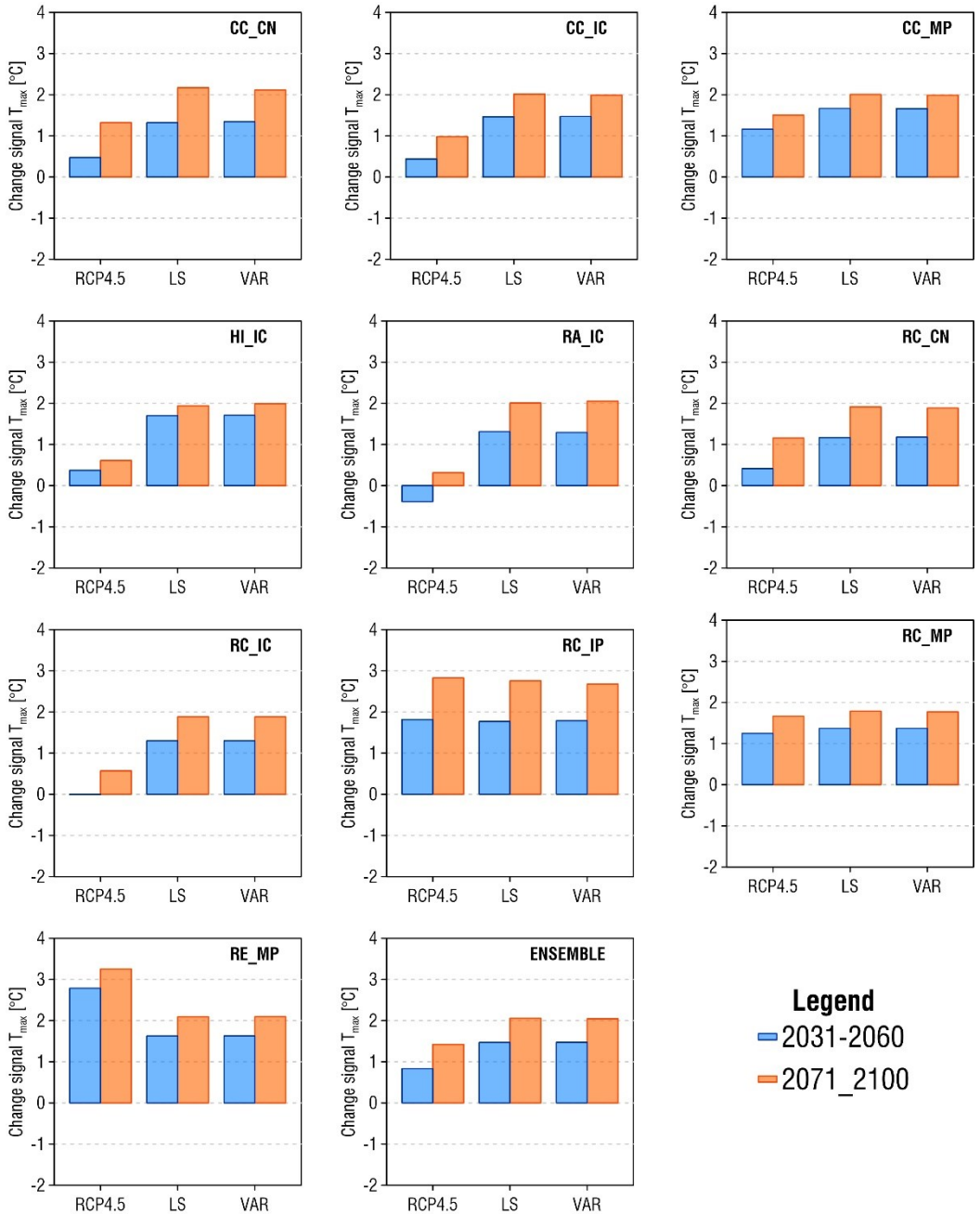


Fig. 61 - Climate change signals of maximum temperature of RCM simulations with (LS, VAR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange).

Table 24 summarises the results of all BC techniques, while Fig. 62 and Fig. 63 highlight the temporal variation in maximum and minimum temperatures (blue lines), resulting from the 12 RCMs' average bias-corrected with LS and VAR for 2031-2060 and 2071-2100 respectively. The linear trend (orange lines) outlines an increasing pattern for both periods and variables. In contrast to precipitation results, air temperature simulations show a limited interannual variation of average temperature and moderate deviations in annual temperature projections resulting from all the selected RCMs (light blue area).

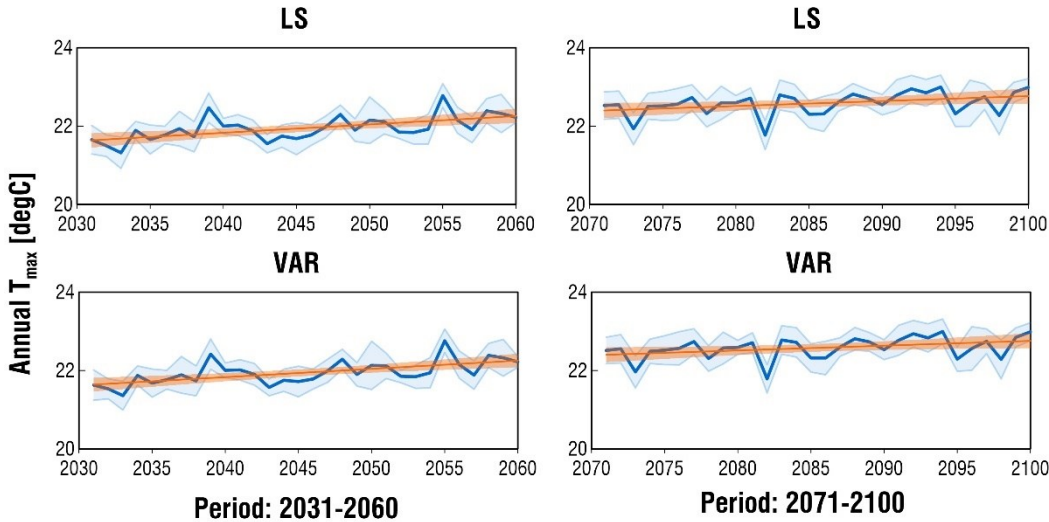


Fig. 62 - Average annual maximum temperature (blue line) and 95% confidence interval (light blue area) based on the data of the 12 RCMs, bias-corrected with LS and VAR for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval.

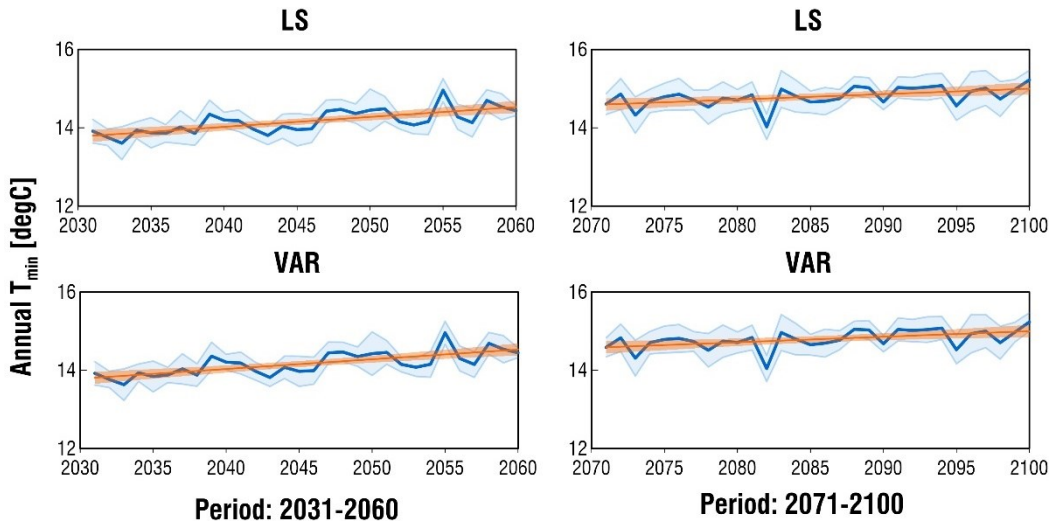


Fig. 63 - Average annual minimum temperature (blue line) and 95% confidence interval (light blue area) based on the data of the 12 RCMs, bias-corrected with LS and VAR for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval.

The average annual maximum temperature obtained from bias-corrected data was approximately 22°C in 2031-2060 and 22.6°C in 2071-2100. These values demonstrate a significant difference from the observed average maximum temperature in 1971-2005, which was 20.5°C. No significant deviation was observed between the two bias correction methods, as indicated in Table 24.

Similarly, the annual minimum temperature derived from bias-corrected data was approximately 14.1°C in 2031-2060 and 14.8°C in 2071-2100. In contrast, the observed average minimum temperature in the historical period of 1971-2005 was 12.9°C. Therefore, the bias-corrected data project higher maximum and minimum temperatures in the future. The annual maximum and minimum temperatures increase on average by 1.4°C and 1.3°C, respectively, compared to the reference period. Considerable temperature increases are observed in 2071-2100, with maximum and minimum temperatures rising by 2.1°C and 1.9°C, respectively.

Table 24 - Summary of annual maximum and minimum change scenarios.

BC	Period	Annual T <sub>max</sub> trend	Average annual T <sub>max</sub> [°C]	Average annual T <sub>max</sub> difference [°C]	Annual T <sub>min</sub> trend	Average annual T <sub>min</sub> [°C]	Average annual T <sub>min</sub> difference [°C]
Historical	1971-2005	+	20.5	na	+	12.9	na
LS	2031-2060	+	21.9	1.4	+	14.2	1.3
EQM	2031-2060	+	22	1.5	+	14.2	1.3
PQM	2031-2060	+	22	1.5	+	14	1.1
GPQM	2031-2060	+	22	1.5	+	14.2	1.3
DQM	2031-2060	+	21.9	1.4	+	14.1	1.2
QDM	2031-2060	+	21.9	1.4	+	14.2	1.3
VAR	2031-2060	+	21.9	1.4	+	14.2	1.3
LS	2071-2100	+	22.6	2.1	+	14.8	1.9
EQM	2071-2100	+	22.6	2.1	+	14.8	1.9
PQM	2071-2100	+	22.6	2.1	+	14.6	1.7
GPQM	2071-2100	+	22.6	2.1	+	14.8	1.9
DQM	2071-2100	+	22.6	2.1	+	14.8	1.9
QDM	2071-2100	+	22.6	2.1	+	14.8	1.9
VAR	2071-2100	+	22.6	2.1	+	14.8	1.9

At a monthly scale, it is evident that maximum and minimum temperatures increase for most months in the two analysed time windows, particularly during the summer. For brevity, Fig. 64 illustrates the monthly change results for maximum temperature in the two periods, considering both LS and VAR methods, as the results for minimum temperature were similar.

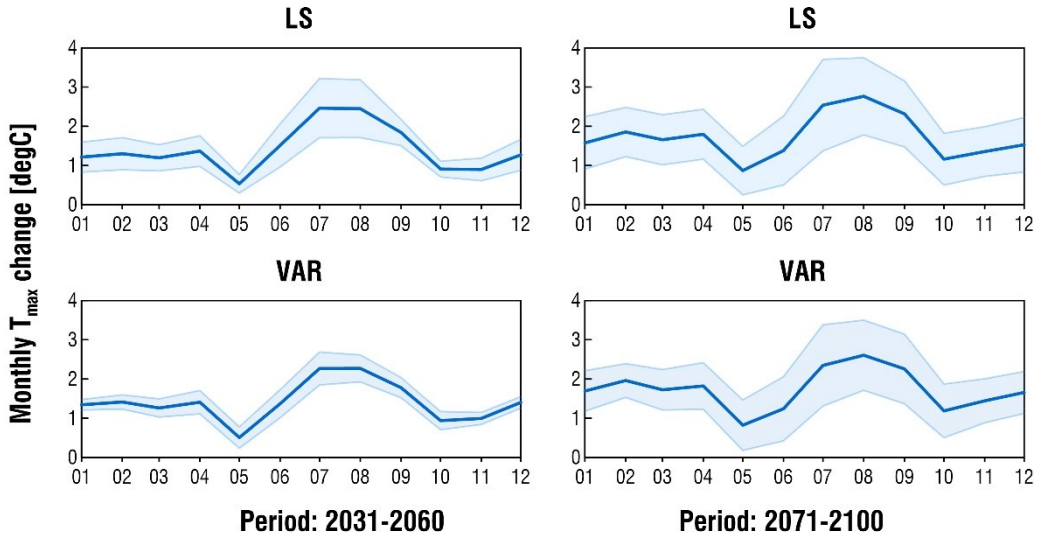


Fig. 64 - Monthly maximum temperature change of 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides) for LS and VAR methods, respectively. The blue line refers to the mean, and the light blue area indicates the 95% confidence interval produced from the 12 RCMs.

Fig. 65 shows the projection of the total annual precipitation and minimum and maximum temperature in 2015-2100 of the RCM ensemble's average values bias corrected through the LS method. The results represent a 10-year moving average to mitigate natural variability and emphasise the climate change signal. The median values of the climate models (blue line) reveal a slight decline in precipitation time series in the future, whilst the variability between the RCMs (blue shadow) indicates high uncertainty in the climate projections. In contrast, the temperature projections suggest a progressive warming in the region under investigation.

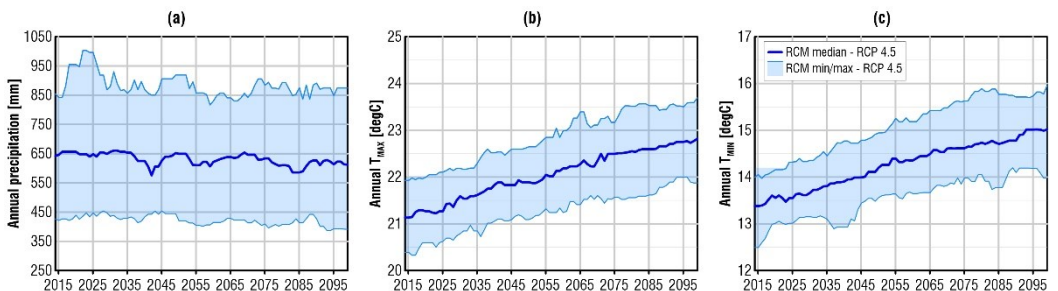


Fig. 65 - Annual precipitation (a), annual maximum temperature (b) and minimum temperature (c) in terms of a 10-year moving average in the period 2015-2100 according to the RCP4.5 scenario.

The expected slight decline in precipitation and rise in temperatures until the end of the century align with findings from other studies conducted in Southern Italy. For example, Bucchignani et al. (2016) projected a modest decrease in winter-autumn precipitation (from September to February) and a moderate to significant reduction in other months under the RCP4.5 scenario for 2071-2100. They based their analysis on high-resolution simulations using COSMO-CLM over Italy. Similarly, there was an estimated increase of over 2°C in the average temperature. D'Oria et al. (2018) also examined temperature changes, reporting an average annual minimum and maximum temperature variations of more than 2°C under the RCP4.5 scenario for the Salento Peninsula.

These trends pose significant challenges for various ecosystems, particularly in regions where water resource management has not received adequate attention. This is especially concerning in areas like Salento, where highly vulnerable aquifer systems serve as the primary source of water for the entire region. These changing climate patterns could lead to reduced aquifer recharge rates due to higher levels of evapotranspiration and, in some periods, decreased precipitation.

In numerous catchments located within the Mediterranean basin, water demand typically exhibits a rise during the summer months, primarily driven by agricultural and tourism-related needs. Consequently, the projected reduction in precipitation during these months, which coincides with periods of stronger dependency on groundwater, will worsen the already severe water shortages. This, in turn, exacerbates the problem of overexploitation in coastal aquifers, leading to groundwater quality deterioration, driven, amongst others, by saltwater intrusion.

Changes in land use can further impact aquifer recharge. The expansion of both inland and coastal urbanization, often driven by tourism development, involves converting substantial areas covered by soil into impermeable or low permeability zones. This alteration impedes infiltration and deep percolation of water, increases surface runoff, and increase flood risk substantially. Additionally, the simultaneous rise in temperatures increases water loss from such surfaces and also from the topsoil through evaporation.

Munafò et al. (2022) offered an up-to-date picture of the land cover transformation processes in Italy, revealing that the Salento region is among the territories most significantly affected by land consumption (overbuilding, construction of photovoltaic plants and abandonment). While this transformation may reduce water demand for irrigation, it also contributes to adverse changes in the infiltration and runoff regimes. On the other hand, changes in minimum and maximum temperatures could impact the agricultural sector through the subsequent increase of potential evapotranspiration, which results in higher crop water requirements. Moreover, the elevated maximum temperatures can subject crops to heat stress, which may disrupt their growth and yield, ultimately leading to substantial economic losses and deficits in food production. Groundwater availability in the Salento aquifer could be compromised due to the projected decline in precipitation and the temperature rise, as emphasized in prior research (Kapur et al., 2007; Lionello et al., 2014; D’Oria et al., 2018). This is especially so in the absence of any surface water bodies in the region, which sets all pressure on groundwater, thus exacerbating water deficit conditions and worsening its’ quality characteristics predominantly due to saltwater intrusion. Fortunately, the Apulian Water Authority is aiming to address water demand in the Apulia region by expanding the water network and optimising the use of surface water resources through the *Piano D’Ambito 2020-2045* ([https://www.autoritaidrica.puglia.it/aip/po/mostra\\_news.php?id=6](https://www.autoritaidrica.puglia.it/aip/po/mostra_news.php?id=6)). This way, the target to maintain or even reduce water withdrawals from the Salento aquifer may be reached.

Notwithstanding, when water levels decrease, the transition zone expands, reducing the thickness of the freshwater layer with a simultaneous increase in its salt content. The recovery of groundwater levels is contingent on the scale of the aquifer and can vary in duration. However, even during wet periods that may lead to water levels replenishment, groundwater quality often remains compromised, as Alfio et al. (2020) pointed out. Therefore, any future reduction in rainfall compared to the historical reference period, albeit modest (c.a. 6%), poses a definitive threat to groundwater quality and availability because the simultaneous decrease in recharge and increased groundwater extraction will most probably expedite the salinisation process.



Addressing the multifaceted challenges posed by climate change requires a holistic approach. As suggested by Lange (2019), an integrated strategy involving regional cooperation, collaborative research, and stakeholder engagement is vital. It encompasses the development of adaptation and mitigation strategies, such as improving the efficiency of resource deployment, conducting thorough assessments of technology for power generation, and increasing reliance on renewable and solar technologies. These measures are necessary to address water and energy scarcity and ensure a secure and sustainable future for Mediterranean societies.

## CONCLUSIONS

Groundwater resources worldwide are essential for satisfying water drinking demand, agriculture, and industry, particularly in semi-arid and arid regions. Climate change further exacerbates water scarcity by altering precipitation patterns and increasing temperatures, leading to higher water demand, especially during droughts. In coastal aquifers, overexploitation and limited recharge rates can result in declining groundwater levels and increased saltwater intrusion, compromising water quality and making it less suitable for various uses. Due to climate variability and human activities, many Mediterranean countries face groundwater quantity and quality challenges. Effective management practices are crucial to ensure long-term groundwater availability and quality while mitigating risks such as overexploitation and saltwater intrusion. The advent of cutting-edge sensor technologies, satellite imagery, and remote sensing tools has revolutionised aquifer data collection. These technologies enable the acquisition of real-time or near real-time and high-resolution datasets, allowing the observation of time trends from networks across large areas. Continuous monitoring data enables the simulation and prediction of the aquifer behaviour under various scenarios. This capability aids in proactive management, such as addressing groundwater depletion or contamination risks. However, in many cases, the design of the monitoring network goes beyond the preliminary understanding of the system to be investigated. Constructing the conceptual model of the hydrogeological configuration and hydrodynamic evolution of an aquifer represents the preliminary yet essential step for designing a sustainable, meaningful, trustworthy, and cost-efficient monitoring plan. In the case of coastal aquifers, this phase becomes more complex because the theory of groundwater flow in variable-density systems is considerably more intricate than under single-density conditions. Due to the absence of detailed guidelines, monitoring plans often prove unexpectedly ineffective as they depend on inadequate understanding of the natural environment. A monitoring design that fails to account for the inherent complexity of the subsurface and the associated processes inevitably leads to uncertainties in monitoring outcomes, that can be misleading.

This Thesis arises from the need to develop a comprehensive and integrated approach, which can be considered a supporting tool for decision makers, water managers and water utilities in defining strategic and proper groundwater resources management. Above all, it aims to serve as a point of reflection on the limitations still existing in our monitoring networks and, consequently, on our knowledge of the aquifer. The main challenge occurred in the hydrogeological data acquisition for the selected aquifer. Despite the continuous advancements in instrumentation, in the case of the Apulia region, the only available daily groundwater level data pertain to the Tiziano project, which covered the years 2007-2011. Additional piezometric data consist of sporadic manual measurements over time (typically every six months) conducted on a monitoring network that, while well-defined spatially, has not been consistently sampled during each sampling campaign. Often, detailed information about monitoring wells is lacking, preventing their use. All these issues rendered the application of the data-driven methodologies intricate and the interpretation of results challenging.

Furthermore, the initial idea of employing artificial intelligence, particularly machine learning, to identify future patterns in GWL data using climate models failed due to the absence of a sufficiently large dataset for training, calibration, and validation steps. While numerical models can provide a more precise understanding of aquifer behaviour, artificial intelligence enables the identification of patterns, trends, and anomalies in hydrogeological data. Numerical models are invaluable tools for understanding and managing groundwater systems, but they require careful consideration of data, model complexity, calibration, uncertainty, and validation. Overcoming these difficulties is essential to ensure the accuracy and reliability of the models, but require a significant amount of data, including geological, hydrological, and hydrogeological information, often not available. Nevertheless, these sophisticated approaches become mere numerical exercises without robust reference measurements.

To this purpose, the Thesis investigated a set of methods requiring a short-time series or discrete measurements for a preliminary assessment of the quantitative and qualitative hydrodynamic mechanism of a complex coastal karst aquifer in southern Italy. The

knowledge gained on the study area can be summarised following the key points reported in the INTRODUCTION Section.

### *1. Hydrogeological characterisation*

The conceptual model of aquifer functioning indicates that the case study does not behave, as previously believed, as a classic karst, but rather as a porous aquifer, where precipitation inputs do not immediately reach the aquifer but with a certain lag. In fact, the analyses of correlations between monthly groundwater levels and climate indices (specifically SPI and SPEI) reveal a statistically significant positive relationship, which varies locally across different time scales and lag times. The presence of a long-term correlation suggests that the Salento aquifer, despite its complexity, responds linearly to precipitation and temperature variations over an extended period, acting as a low-pass filter with significant inertia in terms of transmissivity and storage capacity. When using these correlations to estimate future groundwater levels in response to climate projections, it is advisable to employ the SPEI index. This choice enables a thorough evaluation of the combined effects resulting from variations in both precipitation and temperature, specifically in areas marked by minimal precipitation decreases and substantial temperature increases. The daily time series analyses in both time and frequency domains also confirm the slow behaviour of the case study. They indicate a general baseflow water transfer process, favoured by minor fractures and discontinuities that drain the unsaturated zone, occasionally anticipated by a rapid flow through karst conduits and large fractures during heavy precipitation. Results are locally dissimilar, with variable response times and modest correlation coefficients due to the high degree of complexity and anisotropy of the investigated system. The transfer of infiltration water occurs slowly, and the effects on groundwater levels can be observed with an average lag of 18 months, however varying from area to area.

Wavelet analysis outcomes endorse the high storage capacity of the investigated system resulting from the persistent periodicity higher than 256 days for all monitoring wells.

Groundwater level decomposition proves that the seasonal and remainder components influence less significantly than the trend. Based on the available and relatively short

recordings, all monitoring wells concur in detecting a predominantly negative trend in 2007-2011, except for the hydrological year 2009-2010, which stands out as a wet period compared to the study area's annual average. However, these trends may be considered cautiously, as they refer to a too short time frame.

The results of this study indicate that the proposed combination of methods is advantageous and can be easily applied to other regions where groundwater level and meteorological time series data are accessible. Despite the limited hydrogeological data available, this approach has enhanced the understanding of subsurface hydrodynamic processes in response to meteorological conditions in the Salento aquifer. On the other hand, studies highlighted how the accessibility to long-term, consistent, and spatially distributed measurements can provide a more accurate view of the system and support future groundwater resources planning and management.

## *2. Qualitative assessment*

From the qualitative point of view, Multivariate Statistical analysis and the Hydrogeochemical Facies Evolution Diagram provide a valuable means of the salinisation process during the two analysed monitoring periods. Both methods concur in identifying the portion of the aquifer most susceptible to saltwater intrusion, primarily along the Ionian coastline and within a specific zone extending from the Ionian to the Adriatic seas in the northern sector of the aquifer. It highlights that the size and location of recharge areas remain relatively stable over time. The northwestern boundary of the Salento aquifer, which corresponds to the higher elevations of the adjacent Murgia aquifer, is connected to the central part of the Lecce Province, closely aligning with the endorheic region of the southern portion of the study area.

Another source of pollution impacting the Salento aquifer is associated with nitrates, which may potentially originate from agricultural activities, even though national data on nitrogen-based fertilizers used in the Apulia region indicate a significant reduction from 2008 onwards. The presence of elevated nitrate concentrations in groundwater, as identified through geostatistical analysis, and the apparent disconnection from agricultural practices raise a question regarding the pathways through which these pollutants infiltrate the soil and reach the karst aquifer, as well as the time involved in this

process. This finding holds significant implications for the planning of future groundwater monitoring initiatives, indicating that improvements in groundwater quality in response to mitigation measures could be delayed over time.

### 3. *Future climatic projections*

The last part of the Thesis focuses on assessing climate change within the study area, utilising various RCMs and numerous BC methods for examining future projections. Results predict a declining trend in precipitation by approximately 6% and a rising temperature pattern of 2°C relative to the historical period (1971-2005), extending until 2100. Consequently, recharge rates could be constrained due to increased evapotranspiration levels, reduced precipitation, and augmented water demand. Without alternative water sources, anthropogenic pressures could exacerbate existing water shortages, potentially leading to groundwater quality degradation caused, amongst others, by saltwater intrusion.

Managing water resources in coastal basins represents a common yet significant challenge that increases with time due to climate change and increasing population concentration in coastal basins. It requires focused attention, targeted solutions, and sharing experiences among similar contexts. Leveraging advanced technologies for data collection, analysis, and modelling empowers stakeholders to make informed decisions. However, this phase must be preceded by a proper definition of the hydrogeological conceptual model and hydrodynamic processes to make the monitoring plan effective and cost-efficient. Scientific and technological progress, collaboration among stakeholders from diverse disciplines and the sharing of results from different regions of the globe can help us address the complex challenges facing aquifer management. Temporal and spatial continuity of measurements is crucial in this context.

## BIBLIOGRAPHY

- Abdullah, A.Y. M., Bhuiyan, M. H., Kiselev, G., Dewan, A., Hassan, Q. K., Rafiuddin, M. (2022). Extreme temperature and rainfall events in Bangladesh: A comparison between coastal and inland areas. *International Journal of Climatology*, 42(6), 3253–3273. <https://doi.org/10.1002/joc.6911>
- Agresti, A. (2012). *Analysis of Ordinal Categorical Data*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA.
- Aguilera, H., Guardiola-Albert, C., Serrano-Hidalgo, C. (2020). Estimating extremely large amounts of missing precipitation data. *Journal of Hydroinformatics*, 22 (3): 578–592. <https://doi.org/10.2166/hydro.2020.127>
- Aieb, A., Madani, K., Scarpa, M., Bonaccorso, B., Lefsih, K. (2019). A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria, *Heliyon*. 5(2). <https://doi.org/10.1016/j.heliyon.2019.e01247>
- Alcalá, F.J. & Custodio, E. (2008). Using the Cl/Br ratio as a tracer to identify the origin of salinity in aquifers in Spain and Portugal. *Journal of Hydrology*, 359(1–2), 189–207. <https://doi.org/10.1016/j.jhydrol.2008.06.028>
- Alfio, M.R., Balacco, G., Parisi, A., Totaro, V., Fidelibus, M.D. (2020). Drought Index as Indicator of Salinization of the Salento Aquifer (Southern Italy). *Water*, 12, 1927. <https://doi.org/10.3390/w12071927>
- Alfio, M.R., Pisinaras, V., Panagopoulos, A., Balacco, G. (2023). A comprehensive assessment of RCP4.5 projections and bias-correction techniques in a complex coastal karstic aquifer in the Mediterranean. *Front. Earth Sci.*, Volume 11. <https://doi.org/10.3389/feart.2023.1231296>
- Alfio, M.R., Pisinaras, V., Panagopoulos, A., Balacco, G. (2023). Groundwater level response to precipitation at the Hydrological Observatory of Pinios (central Greece). *Groundwater for Sustainable Development* (Under review).
- Algieri, C., Pugliese, V., Coppola, G., Stefano Curcio, S., Calabro, V., Chakraborty, S., (2022). Arsenic removal from groundwater by membrane technology: advantages, disadvantages, and effect on human health. *Groundwater for Sustainable Development* 19, 100815. <https://doi.org/10.1016/j.gsd.2022.100815>
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998). *Crop Evapotranspiration: Guidelines for Computing Crop Requirements*. *Irrig. Drain. Pap.*; United Nations FAO: Rome, Italy; Volume 56.
- Alsumaiei, A.A. (2020). Monitoring Hydrometeorological Droughts Using a Simplified Precipitation Index. *Climate*, 8, 19.
- Alther, G.A. (1979). A simplified statistical sequence applied to routine water quality analysis: a case history. *Groundwater* 17 (6), 556–561. <https://doi.org/10.1111/j.1745-584.1979.tb03356.x>

- Amraoui, F., Razack, M. & Bouchaou, L. (2003). Turbidity dynamics in karstic systems. Example of Ribaa and Bittit springs in the Middle Atlas Morocco. *Hydrological Sciences Journal*, 48 (6), 971–984. <https://doi.org/10.1623/hysj.48.6.971.51418>
- An, L., Ren, X., Hao, Y., Yeh, T.C.J. (2019). Utilizing precipitation and spring discharge data to identify groundwater quick flow belts in a karst spring catchment. *Journal of Hydrometeorology* 20 (10), 2057–2068. <https://doi.org/10.1175/JHM-D-18-0261.1>
- Angelini, P. (1997). Correlation and spectral analysis of two hydrogeological systems in Central Italy. *Hydrological Sciences Journal* 42 (3), 425–438. <https://doi.org/10.1080/02626669709492038>
- Arampatzis, G., Panagopoulos, A., Pisinaras, V., Tziritis, E., Wendland, F. (2018). Identifying potential effects of climate change on the development of water resources in Pinios River Basin, Central Greece. *Appl Water Sci* 8, 51. <https://doi.org/10.1007/s13201-018-0690-1>
- Armanuos, A.M., Al-Ansari, N., Yaseen, Z.M. (2020). Cross Assessment of Twenty-One Different Methods for Missing Precipitation Data Estimation. *Atmosphere*, 11(4), 389. <https://doi.org/10.3390/atmos11040389>
- Arumi, J.L., Oyarzún, R., Sandoval, M., (2009). Natural protection against groundwater pollution by nitrates in the central valley of Chile. *Hydrol. Sci. J.* <https://doi.org/10.1623/hysj.50.2.331.60654>
- Arumugam, T., Kinattinkara, S., Kannithottathil, S., et al. (2023). I. Comparative assessment of groundwater quality indices of Kannur District, Kerala, India using multivariate statistical approaches and GIS. *Environ Monit Assess* 195, 29. <https://doi.org/10.1007/s10661-022-10538-2>
- Aureli, A., Ganoulis, J., Margat, J. (2008). Groundwater Resources in the Mediterranean Region: Importance, Uses and Sharing, *Water in the Mediterranean*. pp 96-105
- Aydin-Kandemir, F. & Erlat E. (2023). Assessment of the relationship of the salt-covered area and the groundwater storage/drought indicators in the disappearing Lake Tuz in Turkey (1985–2021). *Environmental Monitoring and Assessment*, 195, 333. [doi:10.1007/s10661-022-10793-3](https://doi.org/10.1007/s10661-022-10793-3)
- Babre, A., Kalvans, A., Avotniece, Z., Retiķe, I., et al. (2022). The use of predefined drought indices for the assessment of groundwater drought episodes in the Baltic States over the period 1989–2018. *J. Hydrol. Reg. Stud.*, 40, 101049. [doi: 10.1016/j.ejrh.2022.101049](https://doi.org/10.1016/j.ejrh.2022.101049).
- Baena-Ruiz, L., et al. (2020). Summarizing the impacts of future potential global change scenarios on seawater intrusion at the aquifer scale. *Environ. Earth Sci.*, 79, 99. [doi:10.1007/s12665-020-8847-2](https://doi.org/10.1007/s12665-020-8847-2)
- Bagheri-Gavkosh, M., Hosseini, S.M., Ataie-Ashtiani, B., Sohani, Y., et al. (2021). Land subsidence: A global challenge. *Science of The Total Environment*, Volume 778, 146193, <https://doi.org/10.1016/j.scitotenv.2021.146193>
- Bahrami, M., Khaksar, E., Khaksar, E. (2020). Spatial variation assessment of groundwater quality using multivariate statistical analysis (Case study: Fasa plain, Iran). *J. Groundw. Sci. Eng.* 8 (3), 230–243. <https://doi.org/10.19637/j.cnki.2305-7068.2020.03.004>



- Bailly-Comte, V., Jourde, H., Roesch, A., Pistre, S., Batiot-Guilhe, C. (2008). Time-series analyses for karst/river interactions assessment: case of the Coulazou river southern France. *Journal of Hydrology* 349 (1–2), 98–114. <https://doi.org/10.1016/j.jhydrol.2007.10.028>
- Bailly-Comte, V., Martin, J.B., Screamon, E.J. (2011). Time variant cross correlation to assess residence time of water and implication for hydraulics of a sink-rise karst system. *Water Resources Research* 47 (5), (16 pages). <https://doi.org/10.1029/2010WR009613>
- Bain, D.J., Green, M.B., Campbell, J.L., Chamblee, J.F., et al. (2012). Legacy Effects in Material Flux: Structural Catchment Changes Predate Long-Term Studies. *BioScience*, Volume 62, Issue 6, Pages 575–584, [.org/10.1525/bio.2012.62.6.8](https://doi.org/10.1525/bio.2012.62.6.8)
- Balacco, G., Alfio, M.R., Fidelibus, M.D. (2022a). Groundwater Drought Analysis under Data Scarcity: The Case of the Salento Aquifer (Italy). *Sustainability*. 2022; 14(2):707. <https://doi.org/10.3390/su14020707>
- Balacco, G., Alfio, M.R., Parisi, A., Panagopoulos, A., Fidelibus, M.D. (2022b). Application of short time series analysis for the hydrodynamic characterization of a coastal karst aquifer: the Salento aquifer (Southern Italy). *Journal of Hydroinformatics*; 24 (2): 420–443. <https://doi.org/10.2166/hydro.2022.135>
- Balacco, G., Fiorese, G.D., Alfio, M.R. (2023). Assessment of groundwater nitrate pollution using the Indicator Kriging approach. *Groundwater for Sustainable Development*, 2023, p. 100920. <https://doi.org/10.1016/j.gsd.2023.100920>
- Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods. Theory and Applications*. John Wiley & Sons, New York.
- Beaulieu, C., Ouarda, T.B.M.J., Seidou, O. (2007). Synthèse des techniques d’homogénéisation des séries climatiques et analyse d’applicabilité aux séries de précipitations. *Hydrological Sciences Journal*, 52 (1), 18-37. <https://doi.org/10.1623/hysj.52.1.18>
- Beguera, S., Vicente-Serrano, S.M., Reig, F., Latorre, B. (2013). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.*, 34, 3001–3023.
- ben Aissia, M.A., Chebana, F. and Ouarda, T.B.M.J. (2017). Multivariate missing data in hydrology – review and applications. *Advances in Water Resources*, 110, 299–309. <https://doi.org/10.1016/j.advwatres.2017.10.002>
- Benavente, J. & Pulido-Bosch, A.P. (1985). Application of correlation and spectral procedures to the study of discharge in a karstic system (Eastern Spain). In *Karst Water Resources. Proceedings of the Ankara - Antalya Symposium, July 1985*. IAHS Publ. no. 161.
- Bloomfield, J.P. & Marchant, B.P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrology and Earth System Sciences*, 17, 4769–4787. doi:10.5194/hess-17-4769-2013
- Box, G.E.P., Jenkins, G. M., Reinsel, G.C. (2013). *Time series analysis: forecasting and control*. 4th edn. <https://doi.org/10.1002/9781118619193>
- Box, G.E. & Cox, D.R. (1964). An analysis of transformations. *J. R. Stat. Soc. Ser. B* 26, 211–252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>

- Box, G.E. & Cox, D.R. (1982). An analysis of transformations revisited, rebutted. *J. Am. Stat. Assoc.* 77 (377), 209–210. <https://doi.org/10.1080/01621459.1982.10477788>
- Bruno, G. & Zezza, F. (1992). Typologies of instability and methods of stability verification for the coastal cliffs of Salento (Apulia, Italy). *Bollettino di Oceanologia Teorica ed Applicata*, Vol. X, nn. 2-3-4, April-July-October
- Bucchignani, E., Montesarchio, M., Zollo, A.L., Mercogliano, P. (2016), High-resolution climate simulations with COSMO-CLM over Italy: performance evaluation and climate projections for the 21st century. *Int. J. Climatol.*, 36: 735-756. <https://doi.org/10.1002/joc.4379>
- Cai, Z. & Offerdinger, U. (2016). Analysis of groundwater-level response to rainfall and estimation of annual recharge in fractured hard rock aquifers, NW Ireland. *Journal of Hydrology*, 535, 71–84. <https://doi.org/10.1016/j.jhydrol.2016.01.066>
- Cannon, A. J., Sobie, S.R., Murdock, T.Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, 28(17), 6938–6959. <http://www.jstor.org/stable/26196079>
- Caretta, M.A., et al., (2022). Water. In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 551–712, doi:10.1017/9781009325844.006.
- Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S., Stenseth, N.C. (2008). Wavelet analysis of ecological time series. In *Oecologia*, 156, 287-304. <https://doi.org/10.1007/s00442-008-0993-2>
- Chang, K.Y., Xu, L. Starr, G., Paw U, K.T. (2018). A drought indicator reflecting ecosystem responses to water availability: The Normalized Ecosystem Drought Index. *Agric. For. Meteorol.* 2018, 250–251, 102–117.
- Chatfield, C. & Xing, H. (2013). The analysis of time series: an introduction with R. In *Journal of Chemical Information and Modeling*.
- Chen, J., Brissette, F.P., Leconte, R. (2011a). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J. Hydrol.*, 401 (3–4), 190-202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>
- Chen, J., Brissette, F.P., Poulin, A., Leconte, R. (2011b). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resour. Res.*, 47 (12), W12509. <https://doi.org/10.1029/2011WR010602>
- Chen, J., Brissette, F.P., Lucas-Picher P. (2015), Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *J. Geophys. Res. Atmos.*, 120, 1123–1136. doi:10.1002/2014JD022635
- Chiaudani, A., Di Curzio, D., Palmucci, W., Pasculli, A., Polemio, M., Rusi, S. (2017). Statistical and fractal approaches on long time-series to surface-water/groundwater relationship assessment: a central Italy alluvial plain case study. *Water* 9 (11), 850. <https://doi.org/10.3390/w9110850>

- Chica-Olmo, M., Luque-Espinar, J.A., Rodriguez-Galiano, V., Pardo-Igúzquiza, E., Chica-Rivas, L. (2014). Categorical Indicator Kriging for assessing the risk of groundwater nitrate pollution: the case of Vega de Granada aquifer (SE Spain). *Sci. Total Environ.*, Volumes 470–471, 229–239. <https://doi.org/10.1016/j.scitotenv.2013.09.077>.
- Chon, N.,S. (2010). Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data. Master's Thesis, University of Pittsburgh. (available on line: <http://d-scholarship-dev.library.pitt.edu/8056/>) (accessed 24 September 2021)
- Chow, V.T. (1969). Stochastic analysis of hydrologic systems. Research Report No. 26, Water Resources Center, University of Illinois at Urbana-Champaign.
- Christensen, O.B., Christensen, J.H., Machenhauer, B., Botzet, M. (1998) Very high-resolution regional climate simulations over Scandinavia-Present climate. *J Climate* 11:3204–3229. [https://doi.org/10.1175/1520-0442\(1998\)011<3204:VHRRCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3204:VHRRCS>2.0.CO;2)
- Ciaranfi, N., Pieri, P., Ricchetti, G. (1988). Note alla carta geologica delle Murge e del Salento (Puglia centro -meridionale). *Mem. Della Soc. Geol. Ital.*, 41, 449–460.
- Cleveland, R. B., Cleveland, W.S., McRae, J.E., Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion). *Journal of Official Statistics*.
- Colin, J., De ´ que ´ , M., Radu, R., Somot, S. (2010). Sensitivity study of heavy precipitations in Limited Area Model climate simulation: influence of the size of the domain and the use of the spectral nudging technique. *Tellus A* 62:591–604. <https://doi.org/10.1111/j.1600-0870.2010.00467.x>
- Collados-Lara, A.J., Pulido-Velazquez, D., Pardo-Igúzquiza, E. (2018). An Integrated Statistical Method to Generate Potential Future Climate Scenarios to Analyse Droughts. *Water*, 10, 1224.
- Colyer, A., Butler, A., Peach, D., Hughes, A. (2022). How groundwater time series and aquifer property data explain heterogeneity in the Permo-Triassic sandstone aquifers of the Eden Valley, Cumbria, UK. *Hydrogeol J*, 30, 445-462. doi:10.1007/s10040-021-02437-6
- Cotecchia, V. (1977). Studi e ricerche sulle acque sotterranee e sull'intrusione marina in Puglia (Penisola Salentina), Quaderno n. 20 dell'Istituto di Ricerca sulle Acque del C.N.R., Roma. (in Italian)
- Cotecchia, V. (2017). Area Idrogeologica del Salento. In *Memorie Descrittive Della Carta Geologica d'Italia ISPRA—Istituto Superiore per la Protezione e la Ricerca Ambientale: Rome, Italy*; 92, pp. 312–442. (in Italian)
- D'Oria, M., Tanda, M.G., Todaro, V. (2018). Assessment of local climate change: Historical trends and RCM multi-model projections over the Salento Area (Italy), *Water*, 10(8), 978. <https://doi.org/10.3390/w10080978>
- D'Oria, M., Balacco, G., Todaro, V., Alfio, M.R., Tanda, M.G. (2023). Effects of climate change on a coastal karst aquifer in a semi-arid region. *Hydrological Sciences Journal (Under review)*.
- Dalton, M.G. & Upchurch, S.B. (1978). Interpretation of hydrochemical facies by factor analysis. *Groundwater* 16, 228–233. <https://doi.org/10.1111/j.1745-6584.1978.tb03229.x>.

- Davis, J.C. (1986). *Statistics and Data Analysis in Geology*. 2nd ed. Wiley, New York.
- De Filippis, G., Margiotta, S., Branca, C., Negri, S.L. (2019). A modelling approach for assessing the hydrogeological equilibrium of the karst, coastal aquifer of the Salento peninsula Southeastern Italy: evaluating the effects of a MAR facility for wastewater reuse. *Geofluids* 19 pages. <https://doi.org/10.1155/2019/5714535>.
- Delbart, C., Valdés, D., Barbecot, F., Tognelli, A., Couchoux, L. (2016). Spatial organization of the impulse response in a karst aquifer. *Journal of Hydrology* 537, 18–26. <https://doi.org/10.1016/j.jhydrol.2016.03.029>
- Delbart, C., Valdes, D., Barbecot, F., Tognelli, A., Richon, P., Couchoux, L. (2014). Temporal variability of karst aquifer response time established by the sliding-windows cross-correlation method. *Journal of Hydrology*, 511, 580-588. <https://doi.org/10.1016/j.jhydrol.2014.02.008>
- Dieng, D., Cannon, A. J., Laux, P., Hald, C., et al. (2022). Multivariate bias-correction of high-resolution regional climate change simulations for West Africa: Performance and climate change implications. *Journal of Geophysical Research: Atmospheres*, 127, e2021JD034836. <https://doi.org/10.1029/2021JD034836>
- Diggle, P.J. (1993). Time series: a biostatistical introduction. *Biometrics* 49 (4), 1286. <https://doi.org/10.2307/2532287>
- Douville, H., Raghavan, K., Renwick, J., Allan, R. P., et al. (2021). Water Cycle Changes. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press, Cambridge.
- Droogers, P., Allen, R.G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrig.Drain. Syst.* 2002, 16, 33–45.
- Duarte, L.V., Formiga, K.T.M., Costa, V.A.F. (2022). Comparison of Methods for Filling Daily and Monthly Rainfall Missing Data: Statistical Models or Imputation of Satellite Retrievals? *Water* 2022, 14, 3144. <https://doi.org/10.3390/w14193144>
- Dufresne, J.L., Foujols, M.A., Denvil, S., Caubel, A., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dynam.* 40, 2123–2165. <https://doi.org/10.1007/s00382-012-1636-1>
- Duvert, C., Jourde, H., Raiber, M. & Cox, M. E. (2015). Correlation and spectral analyses to assess the response of a shallow aquifer to low and high frequency rainfall fluctuations. *Journal of Hydrology* 527, 894–907. <https://doi.org/10.1016/j.jhydrol.2015.05.054>
- Duy, N.L., Nguyen, T.V.K., Nguyen, D.V., Tran, A.T., et al. (2021). Groundwater dynamics in the Vietnamese Mekong Delta: Trends, memory effects, and response times. *Journal of Hydrology: Regional Studies*, Vol 33, 100746 <https://doi.org/10.1016/j.ejrh.2020.100746>

- Fan, X., Min, T., Dai, X. (2023). The Spatio-Temporal Dynamic Patterns of Shallow Groundwater Level and Salinity: The Yellow River Delta, China. *Water* 2023, 15, 1426. <https://doi.org/10.3390/w15071426>
- Fantini, A., Raffaele, F., Torma, C., Bacer, S., et al. (2018). Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. *Clim Dyn* 51, 877–900. <https://doi.org/10.1007/s00382-016-3453-4>
- Farge M. (1992). Wavelet transforms and their applications to turbulence. *Annu Rev Fluid Mech.* doi:10.1146/annurev.fluid.24.1.395
- Farnham, I.M., Singh, A.K., Stetzenbach, K.J., Johannesson, K.H. (2002). Treatment of nondelects in multivariate analysis of groundwater geochemistry data. *Chemom. Intell. Lab. Syst.* 60, 265–281. [https://doi.org/10.1016/S0169-7439\(01\)00201-](https://doi.org/10.1016/S0169-7439(01)00201-)
- Fidelibus, M.D., Calò, G., Tinelli, R., Tulipano, L. (2011). Salt ground waters in the Salento karstic coastal aquifer (Apulia, southern Italy). In: Lambrakis, N., Stournaras, G., Katsanou, K. (Eds.), *Advances in the Research of Aquatic Environment*. Environ. Earth Sci. Springer, Berlin, Heidelberg, pp. 407–415 [https://doi.org/10.1007/978-3-642-19902-8\\_48](https://doi.org/10.1007/978-3-642-19902-8_48).
- Fidelibus, M.D. & Pulido-Bosch, A. (2019). Groundwater temperature as an indicator of the vulnerability of karst coastal aquifers. *Geosciences* 2019 9, 23. <https://doi.org/10.3390/geosciences9010023>.
- Fiorillo, F. & Doglioni, A. (2010). The relation between karst spring discharge and rainfall by cross-correlation analysis Campania, Southern Italy. *Hydrogeology Journal* 18, 1881–1895. <https://doi.org/10.1007/s10040-010-0666-1>
- Gado, T.A., Mohameden, M.B., Rashwan, I.M.H. (2022). Bias correction of regional climate model simulations for the impact assessment of the climate change in Egypt. *Environ Sci Pollut Res* 29, 20200–20220. <https://doi.org/10.1007/s11356-021-17189-9>
- Gambini, R. & Tozzi, M. (1996). Tertiary Geodynamic Evolution of the Southern Adria Microplate. *Terra Nov.*, 8, 593–602, doi:10.1111/j.1365-3121.1996.tb00789.x.
- Gampe, D., Schmid, J., Ludwing, R. (2019). Impact of Reference Dataset Selection on RCM Evaluation, Bias Correction, and Resulting Climate Change Signals of Precipitation. *J. Hydrometeorolo.* 20 (9), 1813–1828. <https://doi.org/10.1175/JHM-D-18-0108.1>
- Ghesquière, O., Walter, J., Chesnaux, R., Rouleau, A. (2015). Scenarios of groundwater chemical evolution in a region of the Canadian shield based on multivariate statistical analysis. *J. Hydrol. Reg. Stud.* 4, 246–266. <https://doi.org/10.1016/j.ejrh.2015.06.004>.
- Giménez-Forcada, E. (2010). Dynamic of seawater interface using hydrochemical facies evolution diagram. *Groundwater* 48 (2), 212–216. <https://doi.org/10.1111/j.1745-6584.2009.00649.x>.
- Giménez-Forcada, E. (2014). Space/time development of seawater intrusion: a study case in vinaroz coastal plain (Eastern Spain) using HFE-diagram, and spatial distribution of hydrochemical facies. *J. Hydrol.* 517, 617–627. <https://doi.org/10.1016/j.jhydrol.2014.05.056>.

- Giménez-Forcada, E., Sánchez San Román, F.J. (2015). An Excel Macro to plot the HFE-diagram to identify sea water intrusion phases. *Groundwater* 53 (5), 819–824. <https://doi.org/10.1111/gwat.12280>.
- Giménez-Forcada, E. (2019). Use of the hydrochemical facies diagram (HFE-D) for the evaluation of salinization by seawater intrusion in the coastal Oropesa plain: comparative analysis with the coastal vinaroz plain, Spain. *HydroResearch* 2, 76–84. <https://doi.org/10.1016/j.hydres.2019.11.007>.
- Giorgetta, M.A., Jungclaus, J., Reick, C.H., Legutke, S., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Sy.* 5, 572–597. <https://doi.org/10.1002/jame.20038>
- Giorgi, F., Lionello, P. (2008). Climate change projections for the Mediterranean region, *Global and Planetary Change*, 63 (2–3), 90–104. <https://doi.org/10.1016/j.gloplacha.2007.09.005>
- Glacken, I.M. & Blackney, P.C.J. (1998). A practitioners implementation of indicator kriging. In: *Proceedings of a One Day Symposium: Beyond Ordinary Kriging*, Geostatistical Association of Australasia, Perth, Australia, pp. 27–39.
- Green, T. R., Taniguchi, M., Kooi, H., Gurdak, J. J., et al. (2011). Beneath the surface of global change: Impacts of climate change on groundwater. In *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2011.05.002>
- Grinsted, A., Moore, J.C., Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlin. Processes Geophys.*, 11, 561–566, <https://doi.org/10.5194/npg-11-561-2004>
- Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol. J.* 10, 455–474. <https://doi.org/10.1007/s10040-002-0196-6>.
- Gutjahr, O. & Heinemann, G. (2013). Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theoretical and Applied Climatology*, 114(4), 511–529. <https://doi.org/10.1007/s00704-013-0834-z>
- Hargreaves, G.H. (1994). Defining and using reference evapotranspiration. *J. Irrig. Drain. Eng.*, 120, 1132–1139.
- Hausfather, Z. & Peters, G.P. (2020). Emissions—the ‘business as usual’ story is misleading. *Nature*, 577(7792), 618–620.
- Hazeleger, W., Severijns, C., Semmler, T., Stefanescu, S., et al. (2010). EC-Earth: A Seamless Earth-System Prediction Approach in Action. *B. Am. Meteorol. Soc.* 91, 1357–1364. <https://doi.org/10.1175/2010BAMS2877.1>
- Herrmann, M., Somot, S., Calmanti, S., Dubois, C., Sevault, F. (2011). Representation of daily wind speed spatial and temporal variability and intense wind events over the Mediterranean Sea using dynamical downscaling: impact of the regional climate model configuration. *Nat Hazards Earth Syst Sci* 11:1983–2001. <https://doi.org/10.5194/nhess-11-1983-2011>

- Holman, I.P., Rivas-Casado, M., Bloomfield, J.P., Gurdak, J.J. (2011). Identifying non-stationary groundwater level response to North Atlantic ocean-atmosphere teleconnection patterns using wavelet coherence. *Hydrogeol J.*, 19, 1269-1278. doi:10.1007/s10040-011-0755-9
- Hyndman, R.J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia.
- Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N., Nishimori, M. (2017). Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes, *J. Geophys. Res. Atmos.*, 122, 7800– 7819, doi:10.1002/2017JD026613.
- Imagawa, C., Takeuchi, J., Kawachi, T., Chono, S., Ishida, K. (2013). Statistical analyses and modeling approaches to hydrodynamic characteristics in alluvial aquifer. *Hydrol. Process.*, 27, 4017–4027. <https://doi.org/10.1002/hyp.9538>.
- IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.
- Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., et al. (2019). The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software*, 111, 42-54. <https://doi.org/10.1016/j.envsoft.2018.09.009>
- Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., et al. (2012). Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions. *Atmosphere* 3:181–199. <https://doi.org/10.3390/atmos3010181>
- Jacob, D., Petersen, J., Eggert, B., et al. (2014). EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Change* 14 (2), 563–578. <https://doi.org/10.1007/s10113-013-0499-2>
- Jiao, J. & Post, V. (2019). *Coastal Hydrogeology*. In *Coastal Hydrogeology* (p. I). Cambridge: Cambridge University Press.
- Jiménez Cisneros, B.E., T. Oki, N.W. Arnell, G. Benito, et al. (2014). Freshwater Resources. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*. [C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P. R. Mastrandrea and L. L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229–269.
- Jing, X., Luo, J., Wang, J., Ganggang, Z., Na, W. (2022). A Multi-imputation Method to Deal with Hydro-Meteorological Missing Values by Integrating Chain Equations and Random Forest. *Water Resour Manage* 36, 1159–1173. <https://doi.org/10.1007/s11269-021-03037-5>
- Johnson, R.A. & Wichern, D.W. (1992). *Applied multivariate statistical analysis*. 3rd ed. Prentice-Hall International, Englewood Cliffs, New Jersey.

- Journel, A.G. (1983). Nonparametric estimation of spatial distributions. *J. Int. Assoc. Math. Geol.* <https://doi.org/10.1007/BF01031292>.
- Kaandorp, V. P., Broers, H. P., van der Velde, Y., Rozemeijer, J., de Louw, P. G. B. (2021). Time lags of nitrate, chloride, and tritium in streams assessed by dynamic groundwater flow tracking in a lowland landscape, *Hydrol. Earth Syst. Sci.*, 25, 3691–3711, <https://doi.org/10.5194/hess-25-3691-2021>
- Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika* 39, 31–36. <https://doi.org/10.1007/BF02291575>.
- Kaiser, H.F. (1981). A revised measure of sampling adequacy for factor-analytic data matrices. *Educ. Psychol. Meas.* 41 (2), 379–381. <https://doi.org/10.1177/001316448104100216>.
- Kapur, B., Steduto, P., Todorovic, M. (2007). Prediction of Climatic Change for the Next 100 Years in the Apulia Region, Southern Italy, *Ital. J. Agron. / Riv. Agron.*, 2007, 4:365-371. <https://doi.org/10.4081/ija.2007.365>
- Karangoda, R.C. & Nanayakkara, K.G.N. (2023). Use of the water quality index and multivariate analysis to assess groundwater quality for drinking purpose in Ratnapura district, Sri Lanka, *Groundwater for Sustainable Development*, Volume 21, 100910. <https://doi.org/10.1016/j.gsd.2023.100910>.
- Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81–93.
- Khan, S., Gabriel, H. F., Rana, T. (2008). Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas. *Irrigation and Drainage Systems.* <https://doi.org/10.1007/s10795-008-9049-3>
- Kim, K.B., Kwon, H.H., Han, D. (2015). Bias correction methods for regional climate model simulations considering the distributional parametric uncertainty underlying the observations. *J. Hydrol.* 530, 568–579. <https://doi.org/10.1016/j.jhydrol.2015.10.015>
- Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4, 83–91.
- Kong, F., Song, J., Crosbie, R. S., Barron, O., Schafer, D., Pigois, J.P. (2021). Groundwater Hydrograph Decomposition With the HydroSight Model. *Frontiers in Environmental Science*, 9. <https://doi.org/10.3389/fenvs.2021.736400>
- Krige, D.G. (1951). A statistical approach to some basic mine problems on the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afr.* 52: 119–139
- Kumar, R., Musuuza, J.L., Van Loon, A.F., Teuling, A.J., et al. (2016). Multiscale evaluation of the Standardised Precipitation Index as a groundwater drought indicator. *Hydrology and Earth System Sciences*, 20, 1117-1131. doi:10.5194/hess-20-1117-2016
- Kupiainen, M., Samuelsson, P., Jones, C., Jansson, C., et al. (2011). Rossby Centre regional atmospheric model, RCA4. *Rosby Centre Newsletter*, June.
- Lafare, A. E. A., Peach, D. W., Hughes, A. G. (2016). Use of seasonal trend decomposition to understand groundwater behaviour in the Permo-Triassic Sandstone aquifer, Eden Valley, UK. *Hydrogeology Journal*, 24, 141-158. <https://doi.org/10.1007/s10040-015-1309-3>



- Lange, M.A. (2019). Impacts of Climate Change on the Eastern Mediterranean and the Middle East and North Africa Region and the Water–Energy Nexus. *Atmosphere*, 10, 455. <https://doi.org/10.3390/atmos10080455>
- Larocque, M., Mangin, A., Razack, M., Banton, O. (1998). Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *Journal of Hydrology*, 205 (3–4), 217–231. [https://doi.org/10.1016/S0022-1694\(97\)00155-8](https://doi.org/10.1016/S0022-1694(97)00155-8)
- Leander, R., Buishand, T.A. (2007). Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrol.* 332 (3–4), 487–496. <https://doi.org/10.1016/j.jhydrol.2006.08.006>
- Lee, L.J.E., Lawrence, D.S.L., Price, M. (2006). Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England. *Journal of Hydrology* 330 (3–4), 604–620. <https://doi.org/10.1016/j.jhydrol.2006.04.025>
- Li, J., Ruhe, G., Al-Emran, A., Richter, M.M. (2007). A flexible method for software effort estimation by analogy. *Empir Software Eng* 12, 65–106. <https://doi.org/10.1007/s10664-006-7552-4>
- Lionello, P., Congedi, L., Reale, M. Scarascia, L., Tanzarella, A. (2014). Sensitivity of typical Mediterranean crops to past and future evolution of seasonal temperature and precipitation in Apulia. *Reg Environ Change* 14, 2025–2038. <https://doi.org/10.1007/s10113-013-0482-y>
- Lionello, P. & Scarascia, L. (2018). The relation between climate change in the Mediterranean region and global warming. *Reg Environ Change* 18, 1481–1493. <https://doi.org/10.1007/s10113-018-1290-1>
- Little R.J.A. & Rubin D.B.(1987). *Statistical analysis with missing data*. New York: John Wiley & Sons, Inc.
- Liu, Z., Zhang, X., Fang, R. (2018). Multi-scale linkages of winter drought variability to ENSO and the Arctic Oscillation: A case study in Shaanxi, North China. *Atmospheric Research*, 200, 117-125. <https://doi.org/10.1016/j.atmosres.2017.10.012>
- Lotsi, A., Asiedou, L., Katsepor, J. (2017). Comparison of Imputation Methods for Missing Values in Longitudinal Data Under Missing Completely at Random (MCAR) Mechanism. *African Journal of Applied Statistics*, 4 (1), 241-258. DOI: 10.16929/ajas/241.213
- Loukas, A. Vasiliades, L. Dalezios, N.R. (2003). Intercomparison of Meteorological Drought Indices for Drought Assessment and Monitoring in Greece. In *Proceedings of the International Conference on Environmental Science and Technology*, Lemnos Island, Greece, 8–10 September 2003; Volume B, pp. 484–491.
- Lyra, A. & Loukas, A. (2023). Simulation and Evaluation of Water Resources Management Scenarios Under Climate Change for Adaptive Management of Coastal Agricultural Watersheds. *Water Resour Manage* 37, 2625–2642. <https://doi.org/10.1007/s11269-022-03392-x>
- Machiwal, D., Cloutier, V., Güler, C., Kazakis, N. (2018). A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environ. Earth Sci.* 77 (19). <https://doi.org/10.1007/s12665-018-7872-x> Article Number: 681.

- Mangin, A. (1984). Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *Journal of Hydrology* 67 (1–4), 25–43. <https://doi.org/10.1016/0022-16948490230-0>
- Maraun, D. (2016). Bias correcting climate change simulations-A critical review. *Current Climate Change Reports*, 2(3), 211–220. <https://doi.org/10.1007/s40641-016-0050-x>
- Marsh, C.J., Barwell, L.J., Gavish, Y., Kunin, W.E. (2018). “downscale: An R Package for Downscaling Species Occupancy from Coarse-Grain Data to Predict Occupancy at Fine-Grain Sizes.” *Journal of Statistical Software*, 86(Code Snippet 3), 1–20. <https://www.jstatsoft.org/v86/c03/>
- Matheron, G. (1963). Principles of geostatistics. *Econ. Geol.* 58, 1246–1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>.
- McKee, T.B., Doesken, N.J., Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, Anaheim, CL, 17–22 January, 179–183.
- Mendez, M., Maathuis, B., Hein-Griggs, D., Alvarado-Gamboa, L.F. (2020). Performance evaluation of bias correction methods for climate change monthly precipitation projections over Costa Rica. *Water (Switzerland)* 12 (2). doi:10.3390/w12020482.
- Meng, Q., Xing, L., Li Liu, Xing, X., Zhao, Z., Zhang, F., Li C. (2021). Time-lag characteristics of the response of karst springs to precipitation in the northern China. *Environmental Earth Sciences*. <https://doi.org/10.1007/s12665-021-09640-4>
- Meyers, S.D., Kelly, B.G., O’Brien, J.J. (1993). An introduction of wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon Weather Rev.*, 121(10). doi:10.1175/1520-0493(1993)121<2858:AITWAI>2.0.CO;2
- Molénat, J., Davy, P., Gascuel-Oudou, C., Durand, P. Study of three subsurface hydrologic systems based on spectral and cross-spectral analysis of time series. *J. Hydrol.* 1999, 222, 152–164. [https://doi.org/10.1016/s0022-1694\(99\)00107-9](https://doi.org/10.1016/s0022-1694(99)00107-9).
- Morris, B.L., Lawrence, A.R.L., Chilton, P.J.C., Adams, B., Calow, R.C., Klinck, B.A. (2003). Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management. United Nations Environment Programme, 126pp.
- Munafò, M. (2022). Consumo di suolo, dinamiche territoriali e servizi ecosistemici. Edizione 2022, Report SNPA (in Italian).
- Myers, J.L., Well, A.D., Lorch, R.F. *Research Design and Statistical Analysis*, 3rd ed.; Routledge: London, UK, 2013.
- Nash, J.E. & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — a discussion of principles. *Journal of Hydrology*, 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- National Geographic Society, *Freshwater Resources*, February 28, 2023. <https://education.nationalgeographic.org/resource/freshwater-resources/>

- Ndehedehe, C.E., Ferreira, V.G., Adeyeri, O.E., Correa, F.M et al. (2023). Global assessment of drought characteristics in the Anthropocene. *Resources, Environment and Sustainability*, 12, 100105. doi:10.1016/j.resenv.2022.100105
- Nikolaidis, N.P., Bidoglio, G., Bouraoui, F., Cardoso, A.C. (2014). *Water Quality of the Mediterranean, Comprehensive Water Quality and Purification*, Elsevier, Pages 230-250, <https://doi.org/10.1016/B978-0-12-382182-9.00076-1>
- Niranjannaik, M., Kumar, A., Beg, Z., Singh, A., Swarnkar, S., Gaurav, K. (2022). Groundwater Variability in a Semi-Arid River Basin, Central India. *Hydrology* 2022, 9, 222. <https://doi.org/10.3390/hydrology9120222>
- Noto, L.V., Cipolla, G., Francipane, A. et al. (2023a). Climate Change in the Mediterranean Basin (Part I): Induced Alterations on Climate Forcings and Hydrological Processes. *Water Resour Manage* 37, 2287–2305. <https://doi.org/10.1007/s11269-022-03400-0>
- Noto, L.V., Cipolla, G., Pumo, D. et al. (2023b). Climate Change in the Mediterranean Basin (Part II): A Review of Challenges and Uncertainties in Climate Change Modeling and Impact Analyses. *Water Resour Manage* 37, 2307–2323. <https://doi.org/10.1007/s11269-023-03444-w>
- Pacheco Castro, R., Pacheco Ávila, J., Ye, M., Cabrera Sansores, A. (2018). Groundwater quality: analysis of its temporal and spatial variability in a karst aquifer. *Groundwater* 56 (1), 62–72. <https://doi.org/10.1111/gwat.12546>.
- Padilla, A. & Pulido-Bosch, A. (1995). Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis. *Journal of Hydrology*. [https://doi.org/10.1016/0022-1694\(94\)02648-U](https://doi.org/10.1016/0022-1694(94)02648-U)
- Panagopoulos, G. & Lambrakis, N. (2006). The contribution of time series analysis to the study of the hydrodynamic characteristics of the karst systems: application on two typical karst aquifers of Greece Trifilia, Almyros Crete. *Journal of Hydrology* 329 (3–4), 368–376. <https://doi.org/10.1016/j.jhydrol.2006.02.023>
- Papathodorou, P.G., Lambrakis, N., Panagopoulos, G. (2007). Application of multivariate statistical procedures to the hydrochemical study of a coastal aquifer: an example from Crete-Greece. *Hydrol. Process.* 21 (11), 1482–1495. <https://doi.org/10.1002/hyp.6322>.
- Pardo-Igúzquiza, E., Collados-Lara, A. J., Pulido-Velazquez, D. (2019). Potential future impact of climate change on recharge in the Sierra de las Nieves (southern Spain) high-relief karst aquifer using regional climate models and statistical corrections. *Environmental Earth Sciences*, 78, 598. doi:10.1007/s12665-019-8594-4
- Parisi, A., Monno, V., Fidelibus, M.D. (2018). Cascading vulnerability scenarios in the management of groundwater depletion and salinization in semi-arid areas. *Int. J. Disaster Risk Reduct.*, 30, 292–305.
- Parisi, A., Alfio, M.R., Balacco, G., Güler, C., Fidelibus, M.D. (2023). Analyzing spatial and temporal evolution of groundwater salinization through Multivariate Statistical Analysis and Hydrogeochemical Facies Evolution-Diagram. *Sci Total Environ.* 1;862:160697. doi: 10.1016/j.scitotenv.2022.160697. Epub 2022 Dec 7. PMID: 36493817.

- Pearson, K. (1896). Mathematical contributions to the theory of evolution III. Regression, heredity and panmixia. *Philos. Trans. R Soc.Lond.*, 187, 253–318.
- Pei, Z., Fang, S., Wang, L., Yang, W. (2020). Comparative Analysis of Drought Indicated by the SPI and SPEI at Various Timescales in Inner Mongolia, China. *Water*, 12, 1925. <https://doi.org/10.3390/w12071925>.
- Percival, D.B. & Walden, A.T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge
- Pfeifer, S., Bülow, K., Gobiet, A., Hänsler, A., et al. (2015). Robustness of Ensemble Climate Projections Analysed with Climate Signal Maps: Seasonal and Extreme Precipitation for Germany. *Atmosphere* 2015, 6, 677-698. <https://doi.org/10.3390/atmos6050677>
- Piani, C., Haerter J. O., Coppola E. (2009). Statistical bias correction for daily precipitation in regional climate models over Europe, *Theoretical and Applied Climatology*, 99, 187-192. <https://doi.org/10.1007/s00704-009-0134-9>
- Piccini, C., Marchetti, A., Farina, R., Francaviglia, R. (2012). Application of indicator kriging to evaluate the probability of exceeding nitrate contamination thresholds. *Int. J. Environ. Res.* 6 (4), 853–862. <https://doi.org/10.22059/IJER.2012.556>.
- Polemio, M. (2016). Monitoring and management of karstic coastal groundwater in a changing environment (Southern Italy): a review of a regional experience. *Water* 8, 148. <https://doi.org/10.3390/w8040148>.
- Portoghese, I., Bruno, E., Dumas, P., Guyennon, N., et al. (2013). Impacts of Climate Change on Freshwater Bodies: Quantitative Aspects. In *Advances in Global Change Research*. [https://doi.org/10.1007/978-94-007-5781-3\\_9](https://doi.org/10.1007/978-94-007-5781-3_9)
- Potop, V., Boroneant, C., Možný, M., Štěpánek, P., Skalák, P. (2014). Observed spatiotemporal characteristics of drought on various time scales over the Czech Republic. *Theoretical and Applied Climatology*. <https://doi.org/10.1007/s00704-013-0908-y>
- Priestley, M.B. (1981). *Spectral Analysis and Time Series*, 1, Academic Press, New York 890 pp.
- Prusty, P. & Farooq, S.H. (2020). Seawater intrusion in the coastal aquifers of India-a review. *HydroResearch* 3, 61–74. <https://doi.org/10.1016/j.hydres.2020.06.001>.
- Pulido-Bosch, A. (2021). *Principles of Karst Hydrogeology: Conceptual Models, Time Series Analysis, Hydrogeochemistry and Groundwater Exploitation*. Springer, Cham. doi:10.1007/978-3-030-55370-8.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (Accessed 7 December 2022).
- Randall, D., Wood, R., Bony, S., Colman, R., et al. (2007). Climate models and their evaluation. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, et al. (Eds.), *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge & New York, NY, Cambridge University Press.

- Ravenscroft, P. & Lytton, L. (2022). *Seeing the Invisible: A Strategic Report on Groundwater Quality*. Washington, DC: World Bank. <http://hdl.handle.net/10986/37197>
- Reimann, C. & Filzmoser, P. (2000). Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environ. Geol.* 39, 1001–1014. <https://doi.org/10.1007/s002549900081>
- Rockel, B., Will, A., Hence, A. (2008). Special issue regional climate modelling with COSMO-CLM (CCLM). *Meteorol Z.* 17:347– 348. doi:10.1127/0941-2948/2008/0309
- Romesburg, H.C. (1984). *Cluster Analysis for Researchers*. Lifetime Learning Publications, Belmont, California.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.
- Sağır, Ç., Kurtuluş, B., Razack, M. (2020). Hydrodynamic characterization of Mugla karst aquifer using correlation and spectral analyses on the rainfall and springs water-level time series. *Water Switzerland* 12, 85. <https://doi.org/10.3390/w12010085>
- Samuelsson, P., Jones, C., Wille 'n, U., Ullerstig, A., et al. (2011). The Rossby Centre Regional Climate Model RCA3: model description and performance. *Tellus A* 63:4–23. <https://doi.org/10.1111/j.1600-0870.2010.00478.x>
- Sanford, R.F., Pierson, C.T., Crovelli, R.A. (1993). An objective replacement method for censored geochemical data. *Math. Geol.* 25, 59–80. <https://doi.org/10.1007/BF00890676>
- Sattari, M.T., Rezazadeh-Joudi, A., Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*; 48 (4): 1032–1044. <https://doi.org/10.2166/nh.2016.364>
- Sattari, M.T., Falsafian, K., Irvem, A., Shahab, S., Qasem, S.N. (2020). Potential of kernel and tree-based machine-learning models for estimating missing data of rainfall, *Engineering Applications of Computational Fluid Mechanics*, 14(1), 1078-1094. <https://doi.org/10.1080/19942060.2020.1803971>
- Schmidli, J., Frei, C., Vidale, P.L. (2006). Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *Int. J. Climatol.* 26 (5), 679–689. <https://doi.org/10.1002/joc.1287>
- Secci, D., Tanda, M.G., D'Oria, M., Todaro, V., Fagandini, C. (2021). Impacts of climate change on groundwater droughts by means of standardised indices and regional climate models. *Journal of Hydrology*, 603(D). doi:10.1016/j.jhydrol.2021.127154
- Senatore, A., Fuoco, D., Maiolo, M., Mendicino, G., Smiatek, G., Kunstmann, H. (2022). Evaluating the uncertainty of climate model structure and bias correction on the hydrological impact of projected climate change in a Mediterranean catchment, *Journal of Hydrology: Regional Studies*, 42, 101120. <https://doi.org/10.1016/j.ejrh.2022.101120>
- Shapoori, V., Peterson, T. J., Western, A. W., Costelloe, J. F. (2015). Top-down groundwater hydrograph time-series modeling for climate-pumping decomposition. *Hydrogeology Journal.*, 23, 819-836. <https://doi.org/10.1007/s10040-014-1223-0>

- Simolo, C., Brunetti, M., Maugeri, M., Nanni, T. (2009). Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *Int. J. Clim.*, 30, 1564–1576.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., et al. (2008). A description of the advanced research WRF version 3. (No. NCAR/TN-475+STR). University Corporation for Atmospheric Research. <http://dx.doi.org/10.5065/D68S4MVH>
- Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Ann.Math. Stat.* 19 (2), 279–281. <https://doi.org/10.1214/aoms/1177730256>.
- Smitha, P.S., Narasimhan, B., Sudheer, K.P., Annamalai, H. (2018). An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment, *Journal of Hydrology*, 556, Pages 100-118. <https://doi.org/10.1016/j.jhydrol.2017.11.010>
- Solano S, C., Vargas-Azofeifa, I., Castillo-Muñoz, R., Huapaya R.P, S. (2023). Multivariate data analysis applied to groundwater geochemical characterization, Central Pacific, Costa Rica, *Applied Geochemistry*, Volume 151, 105599. <https://doi.org/10.1016/j.apgeochem.2023.105599>.
- Song, Q., Shepperd, M., Chen, X., Liu, J. (2008). Can k-NN imputation improve the performance of C4.5 with small software project data sets? A comparative evaluation, *Journal of Systems and Software*, 81(12), 2361-2370. <https://doi.org/10.1016/j.jss.2008.05.008>
- Sordo-Ward, A., Granados, A., Iglesias, A., Garrote, L., Bejarano, M.D. (2019). Adaptation Effort and Performance of Water Management Strategies to Face Climate Change Impacts in Six Representative Basins of Southern Europe. *Water*, 11, 1078. <https://doi.org/10.3390/w11051078>
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *Am. J. Psychol.*, 15, 72–101.
- Sperna Weiland, F.C., Visser, R.D., Greve, P., Bisselink, B., Brunner, L., Weerts, A.H. (2021). Estimating Regionalised Hydrological Impacts of Climate Change Over Europe by Performance-Based Weighting of CORDEX Projections. *Front. Water* 3:713537. doi: 10.3389/frwa.2021.713537
- Tadolini, T. & Tulipano, L. (1979). The evolution of Fresh Water-salt Equilibrium in Connection with Draft from the Coastal Carbonate and Karstic Acquirer of the Salentin Peninsula (South Italy). 4° Salt Water Intrusion Meeting, Hannover.
- Tadolini, T., Tulipano, L., Fidelibus, M.D. (1982). Significativi aspetti del chimismo delle acque sotterranee circolanti nell’acquifero carsico della penisola salentina (Italia Meridionale). *Geologia Applicata e Idrogeologia* 17, 411–426. (in Italian).
- Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106(D7):7183–7192
- Taylor, K.E., Stouffer, R.J., Meehl, G.A. (2012). An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* 93, 485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>

- Tefera, G.W., Dile, Y.T., Ray, R.L. (2023). Evaluating the Impact of Statistical Bias Correction on Climate Change Signal and Extreme Indices in the Jemma Sub-Basin of Blue Nile Basin. *Sustainability*, 15, 10513. <https://doi.org/10.3390/su151310513>
- Teutschbein, C. & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456, 12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- Themeßl, M. J., Gobiet, A., Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, *Climatic Change*, 112, 449–468.
- Thomson, A.M., Calvin, K.V., Smith, S.J. et al. (2011). RCP4.5: a pathway for stabilisation of radiative forcing by 2100. *Climatic Change* 109, 77. <https://doi.org/10.1007/s10584-011-0151-4>
- Thornthwaite, C.W. (1948). An Approach toward a Rational Classification of Climate. *Geogr. Rev.*, 38, 54–94.
- Torrence, C. & Compo, G.P. (1998). A Practical Guide to Wavelet Analysis. *Bull Am Meteorol Soc.*, 79(1). doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2
- Tulipano, L. & Fidelibus, M.D. (1993). Metodologie per la valutazione degli effetti di rilascio da reflui urbani sulla distribuzione dei nitrati nelle acque sotterranee delle unità idrogeologiche Murgia e Salento (Italia Meridionale). *Atti del 2° Convegno Internazionale di Geoidrologia*. Firenze (in Italian).
- Tulipano, L. & Fidelibus, M.D. (2002). Mechanism of groundwater salinisation in a coastal karstic aquifer subject to over-exploitation. *Proc. 17th SWIM, Delft, The Netherlands*.
- Tumsa, B.C. (2022). Performance assessment of six bias correction methods using observed and RCM data at upper Awash basin, Oromia, Ethiopia. *Journal of Water and Climate Change*, Vol 13 No 2, 664 doi: 10.2166/wcc.2021.181
- Turrado, C.C., López Mdel, C., Lasheras, F.S., Gómez, B.A., Rollé, J.L., Juez, F.J. (2014). Missing data imputation of solar radiation data under different atmospheric conditions. *Sensors (Basel)*. Oct 29;14(11):20382-99. doi: 10.3390/s141120382.
- Uddameri, V., Singaraju, S., Hernandez, E.A. (2019). Is Standardised Precipitation Index (SPI) a Useful Indicator to Forecast Groundwater Droughts? — Insights from a Karst Aquifer. *Journal of the American Water Resources Association*, 55 (1), 70-88. doi:10.1111/1752-1688.12698
- UNESCO (2022). Groundwater: Making the invisible visible. *Water Security for Peace and Development*, 22-27 March. 2023. 9th World Water Forum in Dakar, Senegal.
- van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Van Loon, A.F. (2015). Hydrological drought explained. *Wiley Interdiscip. Rev. Water*, 2, 359–392.

- Van Loon, A.F. Van Huijgevoort, M.H.J. Van Lanen, H.A.J. (2012). Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. *Hydrol. Earth Syst. Sci.* 2012, 16, 4057–4078.
- van Meijgaard, E., van Uft, B., van de Berg, W. J., Bosveld, F. C., van den Hurk, B., Lenderink, G., and Siebesma, A. P. (2008). The KNMI regional atmospheric climate model RACMO version 2.1 (KNMI TR-302), Technical Report TR-302.
- van Vuuren, D.P., Edmonds, J., Kainuma, M. et al. (2011). The representative concentration pathways: an overview. *Climatic Change* 109, 5. <https://doi.org/10.1007/s10584-011-0148-z>
- Varis, O., Kajander, T., Lemmelä, R. (2004). Climate and water: from climate models to water resources management and vice versa. *Clim. Change* 66 (3), 321–344. <https://doi.org/10.1023/B:CLIM.0000044622.42657.d4>
- Vicente-Serrano, S.M., Beguería S., López-Moreno, J.I. (2010). A multi-scalar drought index sensitive to global warming: The Standardised Precipitation Evapotranspiration Index – SPEI. *J. Clim.* 23: 1696–1718.
- Voltaire, A., Sanchez-Gomez, E., Salas y M´elia, D., et al. (2013). CNRM-CM5.1 global climate model: description and basic evaluation. *Clim Dyn* 40, 2091–2121. <https://doi.org/10.1007/s00382-011-1259-y>.
- Walter, I.A., Allen, R.G., Elliot, R., Jensen, M.E., Itenfisu, D., Mecham, B., et al. (2000). ASCE's standardized reference evapotranspiration equation. In *Proceedings National Irrigation Symposium*; Evans R.G., Benham B.L., Trooien T.P., Eds.; ASAE: Phoenix, AZ, USA, 2000; pp. 209–215.
- Wanders, N. van Lanen, H.A.J. van Loon, A.F. (2010). Indicators for drought characterization on a global scale; *Watch Tech. Rep. 24*; Wageningen Universiteit: Wageningen, The Netherlands, 2010.
- Wang, D., Li, P., He, X., He, S. (2023). Exploring the response of shallow groundwater to precipitation in the northern piedmont of the Qinling Mountains, China, *Urban Climate*, 47, 101379, <https://doi.org/10.1016/j.uclim.2022.101379>
- Wang, L., Butcher, A.S., Stuart, M.E. et al. (2013). The nitrate time bomb: a numerical way to investigate nitrate storage and lag time in the unsaturated zone. *Environ Geochem Health* 35, 667–681 (2013). <https://doi.org/10.1007/s10653-013-9550-y>
- Wang, L., Stuart, M.E., Lewis, M.A., Ward, R.S., et al. (2016). The changing trend in nitrate concentrations in major aquifers due to historical nitrate loading from agricultural land across England and Wales from 1925 to 2150. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2015.10.127>
- Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* 58 (301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>



- Water Protection Plan (2009). Piano di Tutela delle acque (PTA). Relazione generale. (in Italian). [http://www.sit.puglia.it/portal/portale\\_pianificazione\\_regionale/Piano%20di%20Tutella%20delle%20Acque/Documenti/PTA/PTADocumentiDownloadWindow?title=Piano%20di%20Tutela%20delle%20Acque%20-%20PTA%20Approvazione%202009-015&piano=PTA\\_2009&entity=fascicolo&action=2](http://www.sit.puglia.it/portal/portale_pianificazione_regionale/Piano%20di%20Tutella%20delle%20Acque/Documenti/PTA/PTADocumentiDownloadWindow?title=Piano%20di%20Tutela%20delle%20Acque%20-%20PTA%20Approvazione%202009-015&piano=PTA_2009&entity=fascicolo&action=2)
- Water Protection Plan (2015). Piano di Tutela delle acque – Il Ciclo (Water protection plan – Il Cycle). (in Italian). Available at. [http://www.sit.puglia.it/portal/portale\\_pianificazione\\_regionale/Piano%20di%20Tutella%20delle%20Acque](http://www.sit.puglia.it/portal/portale_pianificazione_regionale/Piano%20di%20Tutella%20delle%20Acque). (Accessed 7 December 2022).
- Werner, A.D., Alcoe, D.W., Ordens, C.M. et al. (2011). Current Practice and Future Challenges in Coastal Aquifer Management: Flux-Based and Trigger-Level Approaches with Application to an Australian Case Study. *Water Resour Manage* 25, 1831–1853. <https://doi.org/10.1007/s11269-011-9777-2>
- Wesonga R. (2015). On multivariate imputation and forecasting of decadal wind speed missing data. *Springerplus*. Jan 13;4:12. doi: 10.1186/s40064-014-0774-9.
- Wilcke, R. A. I., Mendlik, T., Gobiet, A. (2013). Multi-variable error correction of regional climate models, *Climatic Change*, 120, 871–887.
- Willmott, C.J. (1981). On the validation of models. *Physical Geography*, 2, 184–194. <https://doi.org/10.1080/02723646.1981.10642213>.
- World Health Organization (2012). World Health Organization Guidelines for drinking-water quality
- Xia, J., Wu, X., Zhan, C., Qiao, Y., Hong, S., Yang, P., Zou, L. (2019). Evaluating the dynamics of groundwater depletion for an arid land in the Tarim Basin, China. *Water (Switzerland)*, 11(2), 186. <https://doi.org/10.3390/w11020186>
- Zeleny, M. (1973). Compromise programming. In: *Multiple Criteria Decision Making*, eds: J. Cochrane L. & M. Zeleny, pp. 262–301, University of South Carolina Press, Columbia, SC.
- Zhang, Z., Chen, X., Chen, X., Shi, P. (2013). Quantifying time lag of epikarst-spring hydrograph response to rainfall using correlation and spectral analyses. *Hydrogeology Journal*. <https://doi.org/10.1007/s10040-013-1041-9>
- Zhang, Z. (2016). Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of Translational Medicine*, 4, 30. doi: 10.3978/j.issn.2305-5839.2015.12.63.

## List of figures

Fig. 1 - Location of the Salento Peninsula and delimitation of the study area.....	8
Fig. 2 - Reference Geological Map of Salento aquifer with traces of geological sections. The legend details are in Table 1.....	11
Fig. 3 - Schematic stratigraphic column and representative geological cross-sections (traces in Fig. 2). .....	13
Fig. 4 - Geomorphological features of the Salento study area. ....	15
Fig. 5 - Water level map for the Salento aquifer (Water Protection Plan, 2009).....	17
Fig. 6 - Location of meteorological gauge stations. ....	18
Fig. 7 - Tiziano Project's monitoring net over the Salento territory with the indication of the selected wells. ....	20
Fig. 8 - Daily hydrograph of the selected eleven monitoring wells. ....	21
Fig. 9 - Groundwater quality monitoring network of the Tiziano Project. The numerical tag refers to the code of each sampling point. ....	22
Fig. 10 - Chemical groundwater monitoring network of Maggiore Project. ....	23
Fig. 11 - Location in the study area of sampling points with nitrate analyses related to 1995-2006 (left-hand side) and 2007-2011 (right-hand side) periods.....	24
Fig. 12 - Filling-in missing values procedure. ....	29
Fig. 13 - Hydrogeochemical Facies Evolution-Diagram (HFE-D) proposed by Giménez-Forcada (2010, 2014). ....	54
Fig. 14 - Experimental variogram and its main parameters. ....	56
Fig. 15 - Example of theoretical variograms. ....	56
Fig. 16 - Meteorological station selected over the study area.....	65
Fig. 17 - Monthly precipitation (left-hand side column), maximum (middle side column) and minimum (right-hand side column) temperature comparison between observed and estimated values obtained with the five selected infilling methods. The reference period for precipitation is 1971-1976, while for temperature is 2000-2002.....	74
Fig. 18 - Number of iteration boxplots and final R, NSE, and SI coefficients for each filling method referring to Pr (top side), Tmax (middle side) and Tmin (bottom side), respectively.....	75
Fig. 19 - Temporal evolution of SPI and SPEI, respectively at the representative time scales of (a, b) 6-months, (c, d) 12-months, (e, f) 24-months for the Lecce rain gauge station from 1951 to 2020. The year marks are centred on the 1st of July of each year.....	79
Fig. 20 - Correlation coefficients between GWLs and the corresponding SPI from 6- to 48-months calculated with Pearson's (green lines), Kendall's (blue lines), and Spearman's (purple lines) methods.....	80
Fig. 21 - Correlation coefficients between GWLs and the corresponding SPEI from 6- to 48-months calculated with Pearson's (green lines), Kendall's (blue lines), and Spearman's (purple lines) methods.....	81

Fig. 22 - Scatterplots and R-squared between GWLs and the corresponding SPI for the more correlated time scale (18 months for all monitoring wells except 9-months for LE-NC4). ..... 82

Fig. 23 - Heat maps of the Pearson correlation coefficients between GWLs and SPIs (left) and GWLs and SPEIs (right) for all wells and different time windows and lags using the subset of data collected during late winter-early spring. .... 84

Fig. 24 - Groundwater level anomalies, relative to 1986-2095, in terms of 10-year moving average for the eight wells according to the SPI-GWL relationships and the RCP4.5 and RCP8.5 scenarios. The GWLs are averaged in February-April and represent the late winter-early spring period. .... 86

Fig. 25 - Groundwater level anomalies, relative to 1986-2095, in terms of 10-year moving average for the eight wells according to the SPEI-GWL relationships and the RCP4.5 and RCP8.5 scenarios. The GWLs are averaged in February-April and, therefore, represent the late winter-early spring period. .... 87

Fig. 26 - Autocorrelation function of Ruffano precipitation time series for each considered hydrological year. .... 89

Fig. 27 - Autocorrelation functions of the short-term analyses conducted for each well. .... 92

Fig. 28 - Cross-correlation functions of the short-term analyses conducted for each well. .... 93

Fig. 29 - a) Autocorrelation and (b) Cross-correlation functions of the long-term analyses conducted for each well on the entire available time series. .... 94

Fig. 30 - Cross Amplitude and precipitation spectral density functions for Ruffano station and LE\_P1TAU of (a) 2008-09 and (b) 2010-11 hydrological cycle. .... 96

Fig. 31 - GWL-STL decomposition classified into Group 1 (a, b, c), Group 2 (d, e, f) and Group 3 (g, h, i). The decomposition reports the Seasonal component as the first graph for every category, the trend in the middle and the remainder as the last. Ticks are centred in October. .... 100

Fig. 32 - Continuous wavelet transform (CWT) of daily rainfall series for (a) Ruffano and (b) Lecce stations. The area covered by a white screen represents the cone of influence (COI). The thick black contour represents the significant region at a 95 % confidence level. .... 101

Fig. 33 - Continuous wavelet transform (CWT) of daily GWL series for the Salento wells. The area covered by a white screen represents the cone of influence (COI). The thick contour represents the significant region at a 95 % confidence level. .... 104

Fig. 34 - Wavelet coherence between GWLs time series and precipitation stations. .... 106

Fig. 35 - Dendrogram of HCA obtained with Tiziano Project dataset. The dashed line represents the selected level of dissimilarity (25), which identifies three clusters: C1, C2 and C3. .... 108

Fig. 36 - Boxplots for major ions for each cluster (C1, C2 and C3) derived by HCA. .... 109

Fig. 37 - HFE-Diagram of the chemical analyses for the seven sampling surveys of the Tiziano Project. The diagrams on the right side add information about the TDS and chloride concentration related to the freshening and intrusion HFs (Giménez-Forcada, 2014, 2019). .... 113

Fig. 38 - Results of FA, HFE-D and HCA are shown for each sampling period of the Tiziano Project. On the left-hand side, the thematic maps are associated with FSs of

F1 (groundwater salinisation) derived by FA; on the right-hand side, the thematic maps are associated with HF<sub>s</sub> derived by the HFE-D. On both sides, results are derived by Q-mode HCA. .... 119

Fig. 39 - Dendrogram of HCA obtained with the Maggiore Project dataset. The dashed line represents the selected level of dissimilarity (17), which identifies three clusters: C1, C2 and C3. .... 120

Fig. 40 - HFE-Diagram of the chemical analyses for the six sampling surveys of the Maggiore Project. The diagrams on the right side add information about the TDS and chloride concentration related to the freshening and intrusion HF<sub>s</sub> (Giménez-Forcada, 2014, 2019). .... 122

Fig. 41 - Results of FA, HFE-D and HCA are shown for each sampling period of the Maggiore Project. On the left-hand side, the thematic maps are associated with FS<sub>s</sub> of F1 (groundwater salinisation) derived by FA; on the right-hand side, the thematic maps are associated with HF<sub>s</sub> derived by the HFE-D. On both sides, results are derived by HCA. .... 125

Fig. 42 - Trend of nitrate concentrations sampled from 1995 to 2021 in a few selected monitoring wells. .... 127

Fig. 43 - Boxplot of nitrate concentrations sampled over the two reference periods in a few selected monitoring wells. .... 128

Fig. 44 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 20 mg/L threshold. .... 130

Fig. 45 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 40 mg/L threshold. .... 131

Fig. 46 - Nitrate probability maps (up) and associated standard deviation (down) for the two reference periods with a 50 mg/L threshold. .... 132

Fig. 47 - Nitrogen surplus changing in agricultural areas from 2000 to 2018 ..... 134

Fig. 48 - Distribution, for agricultural use, of fertilisers (fertilisers, soil conditioners and correctives) from 2003 to 2020 (Source: Istat - <http://dati.istat.it/index.aspx?queryid=23961#>) ..... 135

Fig. 49 - Statistical performance evaluation of the raw and bias-corrected RCM outputs against the monthly observed rainfall for the 12 RCMs in the Salento aquifer from 1971 to 2005. The RMSE, RHO, NSE and BIAS are reported. The coding of the analysed scenarios is reported in Section 5.1.1. .... 139

Fig. 50 - Taylor diagrams of monthly raw and bias-corrected RCMs precipitation data for the period 1971-2005. .... 140

Fig. 51 - Average monthly precipitation of raw and bias-corrected RCM data for 1971-2005. .... 141

Fig. 52 - Statistical performance evaluation of the raw and bias-corrected RCM outputs against the monthly observed maximum temperature for the 12 RCMs in the Salento aquifer from 1971 to 2005. The RMSE, RHO, NSE and BIAS are reported. The coding of the analysed scenarios is reported in Section 5.1.1. .... 143

Fig. 53 - Taylor diagrams of monthly raw and bias-corrected RCMs maximum temperature data for the period 1971-2005. .... 144

Fig. 54 - Average monthly maximum temperature of raw and bias-corrected RCM data for 1971-2005..... 145

Fig. 55 - Average values of compromise programming results for precipitation..... 146

Fig. 56 - Boxplots of the annual precipitation of RCM simulations with (LS, PTR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange). In the boxplots, whiskers indicate the minimum and maximum value of precipitation; the horizontal lines refer to the 25th percentile, median, and 75th percentile from the bottom to the top of each boxplot, and the point symbols represent outliers. .... 149

Fig. 57 - Climate change signals of annual precipitation of RCM simulations with (LS, PTR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (orange). ..... 150

Fig. 58 - Average annual precipitation (bold blue line) and 95% confidence interval (light blue area) based on the data of the 10 RCMs, bias-corrected with LS and PTR methods for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval. .... 151

Fig. 59 - Monthly precipitation change of 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides) for LS and PTR, respectively. The blue line refers to the average monthly value, and the light blue area indicates the 95% confidence interval produced from the 10 RCMs. .... 152

Fig. 60 - Boxplots of the maximum temperature of RCM simulations with (LS, VAR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (in orange). In the boxplots, whiskers indicate the minimum and maximum value of Tmax; the horizontal lines refer to the 25th percentile, median, and 75th percentile from the bottom to the top of each boxplot, and the point symbols represent outliers. .... 154

Fig. 61 - Climate change signals of maximum temperature of RCM simulations with (LS, VAR) and without (RCP4.5) bias correction for 2031-2060 (in blue) and 2071-2100 (orange). ..... 155

Fig. 62 - Average annual maximum temperature (blue line) and 95% confidence interval (light blue area) based on the data of the 12 RCMs, bias-corrected with LS and VAR for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval..... 156

Fig. 63 - Average annual minimum temperature (blue line) and 95% confidence interval (light blue area) based on the data of the 12 RCMs, bias-corrected with LS and VAR for the period 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides), respectively. Orange line indicates the trend, and the light orange area refers to its 95% confidence interval..... 157

Fig. 64 - Monthly maximum temperature change of 2031-2060 (left-hand top and bottom sides) and 2071-2100 (right-hand top and bottom sides) for LS and VAR methods, respectively. The blue line refers to the mean, and the light blue area indicates the 95% confidence interval produced from the 12 RCMs. .... 159

Fig. 65 - Annual precipitation (a), annual maximum temperature (b) and minimum temperature (c) in terms of a 10-year moving average in the period 2015-2100 according to the RCP4.5 scenario. .... 159

## List of tables

Table 1 - Geology of the Salento aquifer (source: MEDSAL Project, report D.2.3) .....	10
Table 2 - Main statistics of NO <sub>3</sub> <sup>-</sup> concentrations. Data are reported in mg/L. ....	24
Table 3 - Missing values percentages over the historical period 1960-2005 (1971-2005 for SAL-MS6 and SAL-MS14).....	27
Table 4 - SPI's drought category according to Mckee et al. (1993). ....	37
Table 5 - Saltwater end-member chemical parameters set for CML definition expressed in mg/L. ....	54
Table 6 - Theoretical variogram functions. ....	57
Table 7 - List and acronyms of used GCM-RCM combinations. ....	64
Table 8 - Lists and acronyms of the implemented BC methods. ....	66
Table 9 - Wells and associated weather stations.....	78
Table 10 - Wells, associated meteorological gauging stations, maximum correlation coefficients between February-April GWLs and SPIs and their accumulation periods and lags in months. Slopes and intercepts (m a.s.l.) of the linear regressions between GWLs and SPIs. ....	85
Table 11 - Wells, associated meteorological gauging stations, maximum correlation coefficients between February-April GWLs and SPEIs and their accumulation periods and lags in months. Slopes and intercepts (m a.s.l.) of the linear regressions between GWLs and SPEIs. ....	85
Table 12 - Calculated hydrological parameters for the Salento aquifer. ....	97
Table 13 - Statistical summary of STL results for Salento boreholes.....	99
Table 14 - Statistical summary of parameters considered in MVSA applications. ....	108
Table 15 - Median values of physical and chemical parameters for each cluster derived by HCA. ....	109
Table 16 - Classification of groundwater based on Cl <sup>-</sup> values.....	110
Table 17 - Varimax rotated factor loadings, eigenvalues, and total and cumulative variance related to the Tiziano Project dataset. ....	111
Table 18 - Ordinary Kriging parameters for F1 factor scores.....	114
Table 19 - Ordinary Kriging parameters for HFE substages.....	114
Table 20 - Varimax rotated factor loadings, eigenvalues, and total and cumulative variance related to the Maggiore Project dataset. ....	121
Table 21 - Frequency values of nitrate concentration threshold exceedances. ....	129
Table 22 - Percentage areas corresponding to equal intervals of probability exceedance of the thresholds calculated for the two reference periods. ....	133
Table 23 - Summary of annual precipitation change scenarios. ....	152
Table 24 - Summary of annual maximum and minimum change scenarios.....	158

Table A1 - Main features of the Salento meteorological gauge station.....	194
Table A2 - Main features of the monitoring wells over the Salento aquifer including in the Tiziano Project. ....	195
Table A3 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs precipitation data for the period 1971-2005. ....	196
Table A4 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs maximum temperature data for the period 1971-2005.....	197
Table A5 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs minimum temperature data for the period 1971-2005.....	198

## **Annex**



Table A1 - Main features of the Salento meteorological gauge station.

<b>Code</b>	<b>Name</b>	<b>LAT</b>	<b>LONG</b>	<b>Elevation (m AMSL)</b>	<b>Precipitation (P) and Air Temperature (AT) Monitoring Period</b>
SAL_MS1	S. Pietro Vernotico	40.4675	18.0008	49	P (1953-2019) AT (1951-2019)
SAL_MS2	Novoli	40.3669	18.0506	51	P (1951-2019) AT (2013-2019)
SAL_MS3	Lecce	40.3503	18.1667	51	P (1950-2019) AT (1950-2019)
SAL_MS4	Masseria Monteruga	40.3339	17.8172	92	P (1951-2019) AT (1979-2019)
SAL_MS5	Copertino	40.2667	18.0500	34	P (1951-2019) AT (2013-2019)
SAL_MS6	Melendugno	40.2672	18.2672	36	P (1971-2019) AT (2013-2019)
SAL_MS7	Canale dell'Asso	40.1842	18.0375	88	P (2010-2019)
SAL_MS8	Nardò	40.1669	18.0333	54	P (1951-2019) AT (1951-2019)
SAL_MS9	Corigliano	40.1506	18.3003	85	P (2008-2019) AT (2013-2019)
SAL_MS10	Otranto	40.1503	18.4842	27	P (1951-2019) AT (1951-2019)
SAL_MS11	Galatina	40.1342	18.1678	73	P (1951-2019) AT (2013-2019)
SAL_MS12	Maglie	40.1008	18.2839	102	P (1950-2019) AT (1951-2019)
SAL_MS13	Minervino di Lecce	40.0836	18.4167	103	P (1951-2019) AT (1951-2019)
SAL_MS14	Collepasso	40.0667	18.1678	120	P (1971-2019) AT (2013-2019)
SAL_MS15	Gallipoli	40.0544	17.9944	31	P (1951-2019) AT (1951-2019)
SAL_MS16	Vignacastri	40.0006	18.4000	99	P (1951-2019) AT (1951-2019)
SAL_MS17	Taviano	39.9675	18.0836	65	P (1951-2019) AT (1951-2019)
SAL_MS18	Ruffano	39.9675	18.2667	134	P (1951-2019) AT (2013-2019)
SAL_MS19	Alessano	39.8842	18.3172	166	P (2018-2019)
SAL_MS20	Presicce	39.8839	18.2667	105	P (1951-2019) AT (1951-2019)
SAL_MS21	S. Maria di Leuca	39.7904	18.3459	26	P (1953-2019) AT (1953-2019)

Table A2 - Main features of the monitoring wells over the Salento aquifer including in the Tiziano Project.

<b>Name</b>	<b>LAT</b>	<b>LONG</b>	<b>Elevation [m AMSL]</b>	<b>Well-Head Thickness [m]</b>	<b>Well Depth [m]</b>	<b>Considered Monitoring Period</b>
LE_19/IIS	40.293	18.299	34.557	0.605	227.0	1/10/2007 - 30/09/2011
LE_12/IIIS	40.295	18.065	42.065	0.646	62.5	1/10/2007 - 30/09/2011
LE_NC4	40.245	17.978	48.195	0.976	170.0	1/10/2007 - 30/09/2011
LE_LS21LE	40.026	18.217	154.621	0.503	230.0	1/10/2007 - 30/09/2011
LE_PS24LE	39.968	18.306	106.059	0.575	250.0	1/10/2007 - 30/09/2011
LE_SG3	40.112	18.061	90.88	0.000	124.0	1/10/2008 - 30/09/2011
LE_PS17LE	40.149	18.229	76.685	0.566	250.0	1/10/2008 - 30/09/2011
LE_P1TAU	39.979	18.237	148.925	0.865	240.0	1/10/2008 - 30/09/2011
LE_4/IIS	40.420	18.165	22.681	0.810	100.0	1/10/2009 - 30/09/2011
LE_1/LR	40.241	18.166	51.412	0.508	210.0	1/10/2009 - 30/09/2011
LE_2/BS	40.085	18.280	86.984	0.463	113.0	1/10/2009 - 30/09/2009

Table A3 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs precipitation data for the period 1971-2005.

	AL_CN	CC_CN	CC_IC	CC_MP	HI_IC	RA_IC	RC_CN	RC_IC	RC_IP	RC_MP	RE_MP	WF_IP
<b>RMSE</b>												
CDW_hist	70.05	58.71	63.35	59.23	62.82	64.89	62.16	68.06	62.50	64.49	64.39	285.20
LS	56.22	61.70	69.89	62.48	67.06	62.90	57.58	64.63	68.53	62.81	67.38	59.46
EQM	55.28	59.57	67.18	60.89	65.30	65.77	60.41	69.55	61.69	65.91	64.14	104.29
PQM	77.27	72.13	80.58	72.95	75.55	78.92	72.45	79.11	79.81	75.83	80.63	76.76
GPQM	85.56	78.00	87.30	84.41	79.40	86.84	94.46	95.66	86.02	96.38	83.68	121.46
LOCI	69.29	64.48	69.87	64.07	67.98	65.73	61.12	67.75	69.60	65.00	69.78	69.74
PTR	61.70	62.18	69.78	64.31	65.32	67.32	62.87	68.52	69.38	66.84	68.28	68.48
DQM	77.75	74.05	90.32	70.22	71.51	77.05	83.57	79.93	77.05	112.17	73.57	138.71
QDM	126.08	68.81	76.90	72.74	68.71	85.74	93.05	86.47	78.29	88.45	75.20	197.54
<b>RHO</b>												
CDW_hist	-0.09	0.32	0.27	0.34	0.35	0.37	0.33	0.34	0.25	0.35	0.29	-0.10
LS	0.44	0.42	0.31	0.37	0.36	0.39	0.44	0.32	0.31	0.35	0.37	0.34
EQM	0.45	0.43	0.34	0.40	0.41	0.41	0.45	0.36	0.36	0.37	0.39	0.28
PQM	0.41	0.43	0.33	0.38	0.38	0.39	0.43	0.33	0.33	0.35	0.37	0.29
GPQM	0.40	0.43	0.33	0.39	0.39	0.39	0.43	0.33	0.33	0.35	0.38	0.26
LOCI	-0.05	0.32	0.27	0.34	0.34	0.37	0.32	0.33	0.25	0.35	0.29	-0.10
PTR	0.41	0.42	0.32	0.37	0.37	0.37	0.41	0.31	0.31	0.33	0.36	0.26
DQM	0.17	0.37	0.30	0.34	0.31	0.33	0.31	0.26	0.22	0.31	0.31	0.02
QDM	0.00	0.38	0.29	0.35	0.30	0.37	0.34	0.29	0.22	0.33	0.31	0.02
<b>NSE</b>												
CDW_hist	-2.16	-1.53	-1.29	-1.42	-0.86	-0.64	-0.81	-0.53	-3.15	-0.59	-1.30	-2.01
LS	-0.49	-0.29	-0.45	-0.60	-0.39	-0.64	-0.57	-0.68	-0.67	-0.65	-0.51	-0.76
EQM	-0.57	-0.38	-0.49	-0.58	-0.34	-0.47	-0.39	-0.40	-0.85	-0.43	-0.57	-0.27
PQM	-0.15	-0.11	-0.26	-0.31	-0.20	-0.27	-0.20	-0.31	-0.37	-0.30	-0.26	-0.31
GPQM	-0.14	-0.10	-0.21	-0.20	-0.13	-0.21	-0.11	-0.17	-0.30	-0.13	-0.23	-0.16
LOCI	-1.65	-0.56	-0.62	-0.70	-0.48	-0.63	-0.87	-0.58	-0.81	-0.61	-0.69	-1.98
PTR	-0.37	-0.29	-0.45	-0.55	-0.41	-0.51	-0.42	-0.57	-0.65	-0.55	-0.50	-0.54
DQM	-0.61	-0.24	-0.24	-0.38	-0.34	-0.39	-0.30	-0.33	-0.51	-0.10	-0.49	-0.67
QDM	-0.87	-0.29	-0.37	-0.32	-0.36	-0.30	-0.31	-0.25	-0.48	-0.18	-0.43	-1.03
<b>BIAS</b>												
CDW_hist	23.4	-25.2	-20.8	-19.0	-19.3	9.5	13.5	4.8	-41.0	1.1	-22.9	409.6
LS	1.6	-0.7	0.9	-0.8	-1.3	0.5	-0.1	-2.3	-1.6	-2.1	-0.6	-0.1
EQM	-1.8	3.9	7.8	5.6	10.0	9.3	10.2	13.5	-4.8	12.2	2.5	71.7
PQM	35.1	26.7	27.6	27.7	24.3	31.7	32.7	29.1	26.8	29.0	28.4	31.5
GPQM	39.7	30.0	30.9	31.7	25.4	36.1	45.5	37.5	29.7	41.1	27.9	62.3
LOCI	-1.3	-4.1	-1.6	-3.3	-4.4	-1.6	-1.1	-2.2	-2.7	-3.1	-3.0	-2.7
PTR	2.0	-0.7	1.0	-0.4	-1.5	0.4	-0.1	-2.0	-1.3	-2.2	-0.4	0.1
DQM	37.5	7.4	4.4	7.4	-4.5	22.3	44.4	16.6	9.4	25.7	2.0	138.5
QDM	122.8	10.9	7.6	16.3	-4.9	47.7	70.9	34.6	11.7	39.0	9.2	246.0

Table A4 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs maximum temperature data for the period 1971-2005.

	AL_CN	CC_CN	CC_IC	CC_MP	HI_IC	RA_IC	RC_CN	RC_IC	RC_IP	RC_MP	RE_MP	WF_IP
<b>RMSE</b>												
<b>Historical</b>	2.45	2.57	2.60	2.40	2.48	2.76	2.33	2.57	2.48	2.14	3.17	24.43
<b>EQM</b>	2.13	2.15	2.16	2.10	1.90	2.12	1.95	2.02	2.01	1.95	1.99	2.04
<b>LS</b>	2.14	2.21	2.21	2.12	1.88	2.10	1.93	2.00	2.00	1.93	1.98	2.10
<b>PQM</b>	2.14	2.08	2.07	2.02	1.97	2.16	2.00	2.08	2.05	2.01	2.02	1.98
<b>GPQM</b>	2.13	2.08	2.06	2.02	1.97	2.15	1.99	2.07	2.02	2.00	2.02	1.98
<b>DQM</b>	2.15	2.14	2.14	2.05	1.97	2.16	2.03	2.14	2.07	2.05	2.06	2.02
<b>QDM</b>	2.34	2.27	2.27	2.15	2.15	2.31	2.30	2.47	2.31	2.38	2.26	2.25
<b>VAR</b>	2.14	2.09	2.07	2.02	1.97	2.16	2.00	2.08	2.03	2.01	2.02	1.98
<b>RHO</b>												
<b>Historical</b>	0.93	0.93	0.93	0.93	0.93	0.93	0.94	0.94	0.93	0.94	0.91	0.92
<b>EQM</b>	0.94	0.94	0.94	0.94	0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.94
<b>LS</b>	0.94	0.93	0.94	0.94	0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.94
<b>PQM</b>	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
<b>GPQM</b>	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
<b>DQM</b>	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.93	0.94	0.94	0.94	0.94
<b>QDM</b>	0.92	0.93	0.93	0.93	0.94	0.93	0.92	0.92	0.93	0.92	0.93	0.93
<b>VAR</b>	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
<b>NSE</b>												
<b>Historical</b>	0.88	0.87	0.85	0.87	0.80	0.78	0.88	0.84	0.88	0.90	0.78	-22.69
<b>EQM</b>	0.89	0.89	0.89	0.90	0.91	0.89	0.91	0.90	0.90	0.91	0.90	0.90
<b>LS</b>	0.89	0.88	0.88	0.89	0.91	0.89	0.91	0.90	0.90	0.91	0.90	0.89
<b>PQM</b>	0.89	0.89	0.89	0.90	0.90	0.88	0.90	0.89	0.90	0.90	0.90	0.90
<b>GPQM</b>	0.89	0.89	0.89	0.90	0.90	0.88	0.90	0.89	0.90	0.90	0.90	0.90
<b>DQM</b>	0.88	0.89	0.89	0.90	0.90	0.88	0.90	0.88	0.89	0.89	0.89	0.90
<b>QDM</b>	0.86	0.87	0.86	0.88	0.88	0.86	0.86	0.84	0.86	0.85	0.88	0.87
<b>VAR</b>	0.89	0.89	0.89	0.90	0.90	0.88	0.90	0.89	0.90	0.90	0.90	0.90
<b>BIAS</b>												
<b>Historical</b>	-1.40	-3.10	-4.00	-1.00	-5.20	-6.90	-3.90	-6.60	0.00	-0.80	7.20	-118.30
<b>EQM</b>	1.20	1.30	1.20	1.90	1.30	1.70	0.20	0.10	0.00	0.30	1.60	0.10
<b>LS</b>	1.00	1.30	1.20	1.60	1.40	1.50	0.10	0.20	0.20	0.20	1.70	0.90
<b>PQM</b>	1.00	1.20	1.10	1.50	1.60	1.60	0.20	0.20	-0.20	0.20	1.80	1.10
<b>GPQM</b>	1.00	1.30	1.10	1.50	1.60	1.60	0.20	0.20	0.30	0.20	1.80	1.10
<b>DQM</b>	1.00	1.00	1.10	1.40	1.30	1.20	0.10	0.20	0.10	0.20	1.70	0.60
<b>QDM</b>	0.30	0.20	0.50	0.00	1.70	0.20	-0.10	0.20	0.10	0.00	2.20	1.60
<b>VAR</b>	1.00	1.20	1.10	1.50	1.50	1.60	0.20	0.20	0.20	0.20	1.80	0.90

Table A5 - Time series-based evaluation of the ensemble monthly raw and bias-corrected RCMs minimum temperature data for the period 1971-2005.

	AL_CN	CC_CN	CC_IC	CC_MP	HI_IC	RA_IC	RC_CN	RC_IC	RC_IP	RC_MP	RE_MP	WF_IP
<b>RMSE</b>												
<b>Historical</b>	4.01	2.08	2.10	2.11	2.16	3.34	3.10	4.03	3.02	2.75	2.32	20.70
<b>EQM</b>	1.89	1.90	1.98	1.92	1.77	1.95	1.74	1.81	1.76	1.83	1.77	2.35
<b>LS</b>	1.91	1.91	1.97	1.91	1.74	1.91	1.76	1.83	1.80	1.85	1.72	2.43
<b>PQM</b>	1.82	1.90	1.97	1.94	1.85	1.99	1.72	1.80	1.83	1.82	1.86	1.89
<b>GPQM</b>	1.82	1.90	1.97	1.93	1.84	1.98	1.72	1.80	1.74	1.82	1.85	1.91
<b>DQM</b>	1.87	1.95	2.01	1.97	1.86	2.02	1.77	1.85	1.81	1.89	1.89	1.99
<b>QDM</b>	2.01	2.10	2.11	2.10	1.96	2.15	1.92	2.03	2.00	2.08	2.06	1.92
<b>VAR</b>	1.82	1.90	1.97	1.94	1.84	1.99	1.72	1.80	1.74	1.82	1.86	1.90
<b>RHO</b>												
<b>Historical</b>	0.90	0.92	0.92	0.92	0.92	0.89	0.92	0.92	0.91	0.91	0.92	0.88
<b>EQM</b>	0.93	0.93	0.93	0.93	0.94	0.93	0.94	0.94	0.94	0.93	0.94	0.92
<b>LS</b>	0.93	0.93	0.93	0.93	0.94	0.93	0.94	0.94	0.94	0.93	0.94	0.90
<b>PQM</b>	0.94	0.93	0.93	0.93	0.94	0.93	0.94	0.94	0.94	0.94	0.93	0.93
<b>GPQM</b>	0.94	0.93	0.93	0.93	0.94	0.93	0.94	0.94	0.94	0.94	0.93	0.93
<b>DQM</b>	0.93	0.93	0.93	0.93	0.93	0.92	0.94	0.93	0.94	0.93	0.93	0.93
<b>QDM</b>	0.93	0.92	0.92	0.92	0.93	0.92	0.93	0.93	0.93	0.92	0.92	0.93
<b>VAR</b>	0.94	0.93	0.93	0.93	0.94	0.93	0.94	0.94	0.94	0.94	0.93	0.93
<b>NSE</b>												
<b>Historical</b>	0.62	0.87	0.84	0.84	0.78	0.55	0.66	0.26	0.71	0.70	0.78	-11.60
<b>EQM</b>	0.87	0.88	0.87	0.88	0.89	0.87	0.89	0.88	0.89	0.88	0.89	0.83
<b>LS</b>	0.87	0.87	0.86	0.87	0.89	0.87	0.89	0.88	0.89	0.88	0.89	0.81
<b>PQM</b>	0.88	0.87	0.86	0.87	0.88	0.86	0.90	0.89	0.89	0.88	0.88	0.88
<b>GPQM</b>	0.89	0.88	0.87	0.87	0.88	0.86	0.90	0.89	0.90	0.88	0.88	0.87
<b>DQM</b>	0.88	0.87	0.86	0.87	0.88	0.86	0.89	0.88	0.89	0.88	0.88	0.86
<b>QDM</b>	0.86	0.84	0.83	0.84	0.86	0.84	0.87	0.87	0.87	0.86	0.85	0.87
<b>VAR</b>	0.88	0.87	0.86	0.87	0.88	0.86	0.90	0.89	0.90	0.88	0.88	0.88
<b>BIAS</b>												
<b>Historical</b>	-22.10	0.70	-2.10	4.50	-1.50	-17.70	-18.10	-26.40	-16.30	-14.00	7.10	-158.50
<b>EQM</b>	1.00	-1.00	-1.10	-0.50	-0.70	0.50	0.80	1.00	0.60	0.60	-0.20	-5.70
<b>LS</b>	-0.70	-0.70	-0.80	-0.20	-0.60	-0.20	-0.10	-0.10	-0.10	-0.10	-0.10	-1.50
<b>PQM</b>	-0.70	-0.70	-0.80	-0.20	-0.20	0.00	-0.10	-0.10	-1.60	0.00	0.20	-1.40
<b>GPQM</b>	-0.70	-0.70	-0.70	-0.20	-0.30	0.10	-0.10	-0.10	-0.10	0.00	0.20	-1.50
<b>DQM</b>	-0.80	-1.00	-1.00	-0.70	-0.70	-0.40	-0.30	-0.40	-0.40	-0.40	-0.50	-2.20
<b>QDM</b>	-3.30	-1.40	-1.20	-1.50	-0.50	-1.80	-2.20	-2.90	-2.10	-1.90	-0.60	1.10
<b>VAR</b>	-0.80	-0.70	-0.80	-0.20	-0.30	0.00	-0.10	-0.10	-0.10	-0.10	0.20	-1.50



# CURRICULUM VITAE



## Maria Rosaria Alfio

**Nationality:** Italian **Date of birth:** 20/03/1993 **Gender:** Female

**Email address:** [mariarosaria.alfio@poliba.it](mailto:mariarosaria.alfio@poliba.it)

**ORCID:** [0000-0001-8991-9571](https://orcid.org/0000-0001-8991-9571)

## EDUCATION

---

### **Polytechnic University of Bari**

2020 – 2023

PhD in Risk and Environmental, Territorial and Building Development

### **Polytechnic University of Bari**

2016 – 2019

MSc in Civil Engineering (D.M. 270/04) [LM-23]

Thesis in Geotechnics: “FEM analysis of soil-structure interaction: influence of seismic signal scaling techniques”

### **Polytechnic University of Bari**

2016 – 2019

BSc in Civil and Environmental Engineering (DM 270/04) [LT-07]

Thesis in Hydraulic Construction: “Estimation of the erosion process in the Galeso and Citrello basins”

## PROFESSIONAL EXPERIENCES

---

### **Polytechnic University of Bari, Research fellow**

2019 – 2020

Title: “T.E.S.A. - Tecnologie innovative per l'affinamento Economico e Sostenibile delle Acque reflue depurate rivenienti dagli impianti di depurazione di Taranto Bellavista e Gennarini”. Scientific Responsible: Prof. A.F. Piccini

## SCIENTIFIC PUBLICATIONS

---

**Alfio, M.R.**, Balacco, G., Parisi, A., Totaro, V., Fidelibus, M.D. (2020). Drought Index as Indicator of Salinization of the Salento Aquifer (Southern Italy). *Water*, 12, 1927.  
<https://doi.org/10.3390/w12071927>

Balacco, G., **Alfio, M.R.**, Fidelibus, M.D. (2022a). Groundwater Drought Analysis under Data Scarcity: The Case of the Salento Aquifer (Italy). *Sustainability*. 2022; 14(2):707.  
<https://doi.org/10.3390/su14020707>

**Alfio, M.R.**, Balacco, G., Delle Rose, M., Fidelibus, C., Martano, P. (2022). A Hydrometeorological Study of Groundwater Level Changes during the COVID-19 Lockdown Year (Salento)

- Peninsula, Italy). *Sustainability* 2022, 14, 1710. <https://doi.org/10.3390/su14031710>
- Balacco, G., **Alfio, M.R.**, Parisi A., Panagopoulos, A., Fidelibus M.D. (2022). Application of short time series analysis for the hydrodynamic characterization of a coastal karst aquifer: the Salento aquifer (Southern Italy). *Journal of Hydroinformatics*; jh2022135. <https://doi.org/10.2166/hydro.2022.135>
- Parisi, A., **Alfio, M.R.**, Balacco, G., Güler, C., Fidelibus, M.D. (2023). Analyzing spatial and temporal evolution of groundwater salinization through Multivariate Statistical Analysis and Hydro-geochemical Facies Evolution-Diagram. *Science of the Total Environment – ELSEVIER*. DOI: 10.1016/j.scitotenv.2022.160697
- Balacco, G., Fiorese, G.D., **Alfio, M.R.** (2023). Assessment of groundwater nitrate pollution using the Indicator Kriging approach. *Groundwater for Sustainable Development*, 21, 100920. <https://doi.org/10.1016/j.gsd.2023.100920>
- Balacco, G., Fiorese, G.D., **Alfio, M.R.**, Totaro, V., Binetti, M., Torresi, M., Stefanizzi, M. (2023). PaT-ID: A tool for the selection of the optimal pump as turbine for a water distribution network. *Energy*, 282, 128366. DOI: 10.1016/j.energy.2023.128366
- Alfio, M.R.**, Pisinaras, V., Panagopoulos, A., Balacco, G. (2023). A comprehensive assessment of RCP4.5 projections and bias- correction techniques in a complex coastal karstic aquifer in the Mediterranean. *Front. Earth Sci.* 11:1231296. <https://doi.org/10.3389/feart.2023.1231296>

## **PUBLICATIONS UNDER REVIEW**

---

- Alfio, M.R.**, Pisinaras, V., Panagopoulos, A., Balacco, G. (2023). Groundwater level response to precipitation at the Hydrological Observatory of Pinios (central Greece). *Groundwater for Sustainable Development*.



## Abstract

This research investigates a set of data-driven methods for characterising coastal aquifers under data scarcity using available short hydrogeological time series and chemical surveys. Correlation indices, time series analyses, wavelet analyses, and seasonal and trend decomposition, are extensively analysed using precipitation and groundwater levels (GWLs) to define the hydrodynamic mechanism of an aquifer system in response to climate factors. They can provide valuable insights into various aspects, including the nature of the aquifer, the influence of climatic conditions, and significant abstractions. Multivariate statistical analysis is instead a valuable approach handling multiple geochemical and physical parameters to reveal spatial and temporal variations in groundwater quality, identify key hydrochemical processes, and assess how they change over time. Together with the Hydrogeochemical Facies Evolution-Diagram, these techniques allow to explore the salinisation process. Through Geostatistic, the salinisation process and nitrate pollution are investigated in space and compared in time, allowing the identification of the areas more vulnerable. Finally, a comprehensive assessment of climate change projections and bias-correction techniques, employing historical and regional climate data, examine the potential impacts of weather projections on a coastal aquifer. The Thesis focuses on the coastal karst aquifer of Salento (Southern Italy), which presents numerous challenges, i.e., geomorphological complexity, regional size, limited surface water resources, and significant water withdrawals for various human activities. The aim is to investigate the hydrodynamic mechanism of such aquifer and discuss the potential problematics to which groundwater resources are exposed due to climate change and human pressure. The study encourages for the establishment and consistent implementation of a comprehensive and strategic monitoring plan for GWLs, water quality parameters, and other relevant variables aimed at ensuring the long-term sustainability and availability of groundwater resources.