



# Trend detection and depletion effects evidence in time series of groundwater levels in the southern sector of the left bank of the Tagus-Sado Basin (Portugal, Iberian Peninsula)

Mariana Ferreira Branco<sup>1</sup> · Sofia Verónica Barbosa<sup>1,2</sup> · João Xavier Matos<sup>3</sup>

Received: 10 July 2023 / Accepted: 8 March 2024  
© The Author(s) 2024

## Abstract

In the last 20 years in Portugal, water resources have been affected to the point that water storage has decreased by 20% since 2000. Creating strategies to manage water resources requires a comprehensive understanding of the factors influencing water storage and their effects over time. This study is focused on the evolution of Groundwater Deep Levels (GDL) by applying a two-phase trend analysis methodology to examine the dynamic changes in GDL within a series of monitoring wells located in the Central and Southern sectors of the Left Bank of the Tagus-Sado Cenozoic age Basin, situated in Portugal. In the initial phase of trend analysis, Factorial Analysis of Mixed Data (FAMD) was employed and posteriorly the Hierarchical Classification Analysis (HCA). These techniques enabled us to identify distinct GDL trend profiles and generate interpretative maps illustrating their spatial distribution. In the second phase, the non-parametric Mann–Kendall Analysis (MKA) and Innovative Trend Analysis (ITA) were applied, allowing for a quantified confirmation of the different trend profiles previously detected. These techniques allowed the identification of positive and negative hydrodynamic trends in distinct sections of the Basin. In the SE sector they are characterized by a significative increase of GDL associated with overexploitation and in the Central sector with a decrease of GDL. Nevertheless, significant depletion effects can result from natural factors such as prolonged droughts, and in certain regions, changes in geological and hydrothermal dynamics, such as Alpine-age faults, graben, and horst structures, may account for these alterations.

**Keywords** Tagus-Sado Cenozoic Basin · Groundwater level monitoring · Groundwater hydrodynamic trends · Factorial Analysis of Mixed Data · Hierarchical Classification Analysis · Mann–Kendall Analysis · Şen's T Analysis

## Introduction

Groundwater reservoirs play a crucial role in both the environment and the development of our 21st-century society. The preservation of these reservoirs is increasingly challenging due to a range of factors such as natural ones like dry seasons combined with low precipitation ranges and non-natural ones such as over-exploitation, contamination, and climate change. This preservation dilemma has become a prominent global concern. According to the latest report by the Intergovernmental Panel on Climate Change (IPCC), southern European countries are at high risk of extreme drought and with the potential for flooding, with a temperature increase equal to or greater than 1.5 °C, defined as a target value not to be exceeded (Pörtner et al. 2022). With these temperature increases, the hydrological cycle will be affected, and consequently, the aquifer systems will be too.

---

✉ Sofia Verónica Barbosa  
svtb@fct.unl.pt

Mariana Ferreira Branco  
mf.branco@campus.fct.unl.pt

João Xavier Matos  
joao.matos@lneg.pt

<sup>1</sup> Department of Earth Sciences, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, NOVA FCT, 2829-516 Caparica, Portugal

<sup>2</sup> Department of Earth Sciences, Faculdade de Ciências e Tecnologia, GeoBioTec—GeoBioSciences, GeoTechnologies and GeoEngineering & Universidade Nova de Lisboa, NOVA FCT, 2829-516 Caparica, Portugal

<sup>3</sup> LNEG—Laboratório Nacional de Energia e Geologia, I.P., Bairro da Vale d'Oca, Apartado 14, 7601-909 Aljustrel, Portugal

In order to find a solution, the initial step involves researching the evolution of groundwater systems through time and identifying factors that can either enhance or hinder their effective management.

Analysis of GDL evolution through time is a crucial key-issue for the understanding of the dynamics and the evolutionary conditions of groundwater systems. The GDL consist of the difference between the elevation of the top of the well and the groundwater level. If GDL increases, it means that the groundwater level decreases. If GDL decreases, the groundwater level increases. This gives us direct information of the health of the groundwater system, compared to de precipitation and temperature information.

When considering the application of GDL, it is crucial to take into consideration two assumptions. Firstly, the GDL time series are influenced by both local and regional conditions and so GDL time-series result in non-stationary behaviour. Therefore, it is necessary to explore alternative options for its trend analysis. Traditional trend identification and significance tests are restrictive and make assumptions that may not be applicable for monitoring hydrodynamic parameters such as GDL. Understanding and determining the factors that influence the spatial and temporal variation of groundwater levels and hydrochemical parameters is essential to outline action and monitoring measures, allowing the preservation of this resource.

As a reference many publications reflect the use of multivariate statistical analysis in the evolution of salinization conditions in coastal areas (Ferchichi et al. 2018; Soltani and Mellah 2023), in the study of water quality (Kantiranis et al. 2017; Wang et al. 2018; De Andrade Costa et al. 2020; Varol 2020; Ghashghaie et al. 2022; Ghaemi and Noshadi 2022; Wali et al. 2022), and in the variation of piezometric levels (Machiwal, and Singh 2015; Adamovic et al. 2016; Rabiei et al. 2022). These techniques thus allow the analysis of relationships between variables in a comprehensive way and their quantification (Shiker 2012; Javadinejad et al. 2019a).

Factorial Analysis (FA) is a set of multivariate statistical analysis techniques that helps understand the relationship between variables. By transforming data and simplifying the relationships between variables it is possible to highlight a smaller number of intrinsic characteristics, designated as factors (Pagès 2004). It is, therefore, also known as Common Factor Analysis (CFA) (Seth 2022).

Factors capture much of the information of a set of variables in the dataset by providing an understanding of the underlying concepts of its interrelations. The FA can be applied to group the samples in a dataset based on their similarity for a certain characteristic. It can explore deeper factors that may not be evident in the dataset, reducing its multidimensionality. This technique can be handy for exploring relationships in a particular dataset category. The explored concepts or causes may not be immediately evident but may

represent peculiarities or certain tendencies that may be difficult to identify or measure in a simpler way. The FA is beneficial for the interpretation of groundwater-quality data and for understanding specific hydrogeochemical processes, many distinct studies regarding multivariate statistics and FA applied in distinct contexts of groundwater study can be found in the related literature (Ruiz et al. 1990; Cerón et al. 2000; Love et al. 2004; Celestino et al. 2018; Krishnan et al. 2019; Panda et al. 2019; Fatahi et al. 2021).

The most widely used FA technique is Principal Component Analysis (PCA), which enables the relationships between quantitative variables to be assessed. There is another technique that simply assesses the relationship between qualitative variables, which is the Multiple Correspondence Analysis (MCA). There are several examples in the literature where factors identified with FA are considered in subsequent analyses, such as regression or unsupervised classification (cluster analysis). K-means Cluster Analysis and Hierarchical Classification Analysis (HCA) are the most common methods considered in this context (Celestino et al. 2018; Oh et al. 2020; Barbosa et al. 2021). However, in the presence of variables that contain both quantitative and qualitative information, it is not possible to use PCA or MCA.

Since this case study falls into this situation, FAMD technique and its potentialities were experimented. FAMD is an extension of the Factor Analysis (FA) method (Abdi 2003; Adamovic et al. 2016),. it is a multivariate statistical analysis method that is used to analyse datasets with both continuous and categorical variables. It is, in practical terms, a combination of FA and MCA techniques, FAMD is a multivariate statistical technique that is used to identify underlying dimensions or factors that explain the observed variation in a set of quantitative and qualitative variables.

In groundwater studies, the Hierarchical Classification Analysis (HCA) method, can be combined or not with FA, PCA or FAMD factors, allowing a clearer grouping of data (Wang et al. 2015; Jiang et al. 2015; Bayo and Lopez-Castellano 2016; Al Naeem et al. 2019; Rao and Chaudhary 2019; Subba et al. 2019; Visbal-Cadavid et al. 2020; Rahbar et al. 2020; Batdelger et al. 2023).

The other assumption is that GDL trends may change significantly through time due to the dependence on distinct factors that cause variations in hydrodynamic conditions, some of these causes may act independently in time or could be a part of complex combined trending effects. In these cases, verification of positive or negative slope tendencies may be difficult or even impossible using simple linear approaches. In geosciences, parametric methods are widely used, although often that the data under analysis does not meet the necessary normality criteria.

In hydrological studies it is wrong to assume that the data is stationary and independent of the time series, as this does not correspond to reality (Helsel 1987; Anderson 2008;

Mumby 2002; Riaz et al. 2016; Mirabbasi et al. 2020). Furthermore parametric methods are hypersensitive in to the presence of outliers in the data series, unlike non-parametric methods (Mirabbasi et al. 2020). Thus, it is natural to resort to non-parametric methods, like, Mann–Kendall Analysis (MKA) and Şen's T Analysis, also known as Innovative Trend Analysis (ITA). According to Şen (2012, 2014) ITA is proposed as an improved technique of the classical MKA, there are some differences between these two techniques.

In MK analysis it is assumed that time series are independent and have no serial correlation. In ITA it is considered the opposite situation, in this technique it is considered that serial correlations in small series are usually significant in the interpretations of the phenomena as an all (Şen 2017).

ITA allows the identification of monotonous and holistic trends, as is the case with MK, however, ITA is a graphical method that allows trend patterns to have low, medium, and high values in the data, therefore applying to non-stationary datasets (Şen 2012, 2014, 2017). This method was originally developed for hydro-climatological time series and has been increasingly used due to problems of non-stationary, high magnitude and variability which are increasingly more often detected in distinct type of climate and environmental data time series. Examples of its applicability can be found in recent works (Caloiero 2019; Javadinejad et al. 2019b; Ali-fujiang et al. 2020; Güçlü 2020; Minea et al. 2020; Achite et al. 2021; Zakwan 2021; Buri et al. 2022; Swain et al. 2022; Umar et al. 2022). Nevertheless, Şen et al. (2019) proposed another innovative method, defined was Innovative Polygon Trend Analysis (IPTA), which has been used in several studies and investigations in Hydrogeology (Caloiero et al. 2018; Sanikhani et al. 2018; Kuriqi et al. 2020; Harkat and Kisi 2021; Ahmed et al. 2022), in this method, polygon patterns are obtained using information such as the mean, minimum, maximum, standard deviation or skewness parameters of the data at different time scales (daily, monthly, etc.) (Akçay et al. 2022). When compared to the ITA this method allows for the information associated to one year to be symbolized and can retain the trend, the magnitude and slope of trend transitions between successive segments (e.g., days, months) (Akçay et al. 2022).

In the present study, due the two characteristics of GDL trends, FAMD was applied first to identify the temporary evolution of each well and to verify existence of effects of depletion or rising in the groundwater levels and their spatial distributions along the basin. After that, the results were combined with HCA to group wells by the results. Lastly, ITA, IPTA and MK algorithm were applied with the aim of having a synthesis of positive and negative trends of GDL dataset which are a direct consequence of distinct hydrodynamic conditions. Examples of application of ITA to groundwater levels data series can be found in the references Minea et al. (2020), and Zakwan (2021).

## Materials and methods

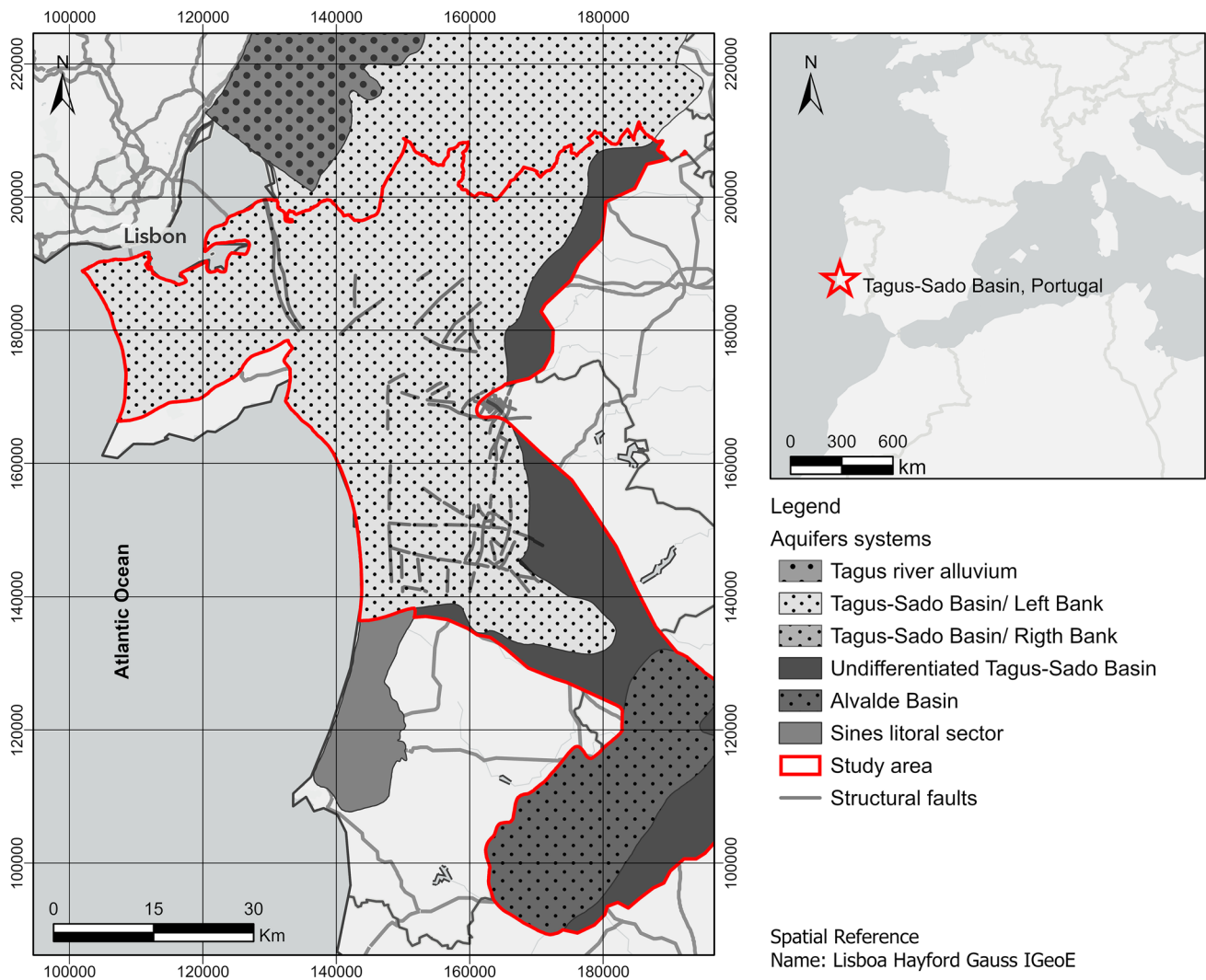
### Geological and hydrogeological characteristics of study area

The area of study integrates the Central and Southern sectors of the aquifer systems of the Left Bank of the Tagus-Sado Basin (Portugal, Iberian Peninsula, Fig. 1). The Tagus-Sado Basin is a wide sedimentary basin, formed by Cenozoic and Quaternary sediments, it consists of a long depression with NE-SW direction, limited in the W and N by Mesozoic formations, and NE, E and SE by Paleozoic substrate (Almeida et al. 2000).

The studied wells reach groundwaters from the Upper Miocene, which includes various types of aquifers (semi-confined and confined systems). It is a complex multilayer aquifer system (Costa 1994) mainly composed by alternating sedimentary sequences of continental and marine facies, resulting from transgressions and regression phenomena, controlled by the reactivation of graben and horst structures during the Alpine orogeny, promoting different hydrogeological behaviours and high spatial variability, thus allowing the creation of free, semi-confined and confined aquifer systems (Costa 1994; Fernandes and Silva 1998; Simões 1998; Simões and Legoinha 2014). This system is connected vertically with the free aquifer of Quaternary Pliocene–Pleistocene age (Zbyszewski et al. 1976; Antunes 1983; Manuppella et al. 1999; Pais et al. 2006), which is mainly constituted by fluvial deposits from deltaic systems associated to last regression of the basin, these allow some parts of the basin have a length of 800 m deeper. Most of the fault structures in the basin area are associated with the tectonic inversion of the Lusitanian Basin (Mesozoic) and the Paleozoic punch, as a result of the convergence of the African continent, relatively, to the Iberian continental block (Kullberg et al. 2000).

In the North sector, the basin was controlled by structures with NE-SW direction (Carvalho et al. 1983–85; Cunha 1992). However, one the most important structures, Pinhal Novo fault presents a general orientation NNW-SSE and covers a wide zone of deformation (about 1.5 km) (Correia 2017), in which it presents a pattern of branching and anastomosing faults.

In the South sector, the basin was controlled by the reactivation of the late-Hercynian faults of the Iberian Pyrite Belt, which have a general NW–SE orientation, showing cleavage and folding planes compatible with a strong vergence to SW (Matos 2021). During this time the fragmentation of the Paleozoic background happened due to alpine and/or late-variscan faults with NNE-SSW orientation NE-SW, NW–SE and E-W, creating stratigraphic variations between the North and South sector of the basin, as



**Fig. 1** Study area, located in the Central and Southern sectors of the aquifer systems of the Left Bank of the Tagus-Sado Basin (Portugal)

well as the diffraction from other aquifers systems associated to Tagus-Sado Basin (Oliveira et al. 1998, 2001).

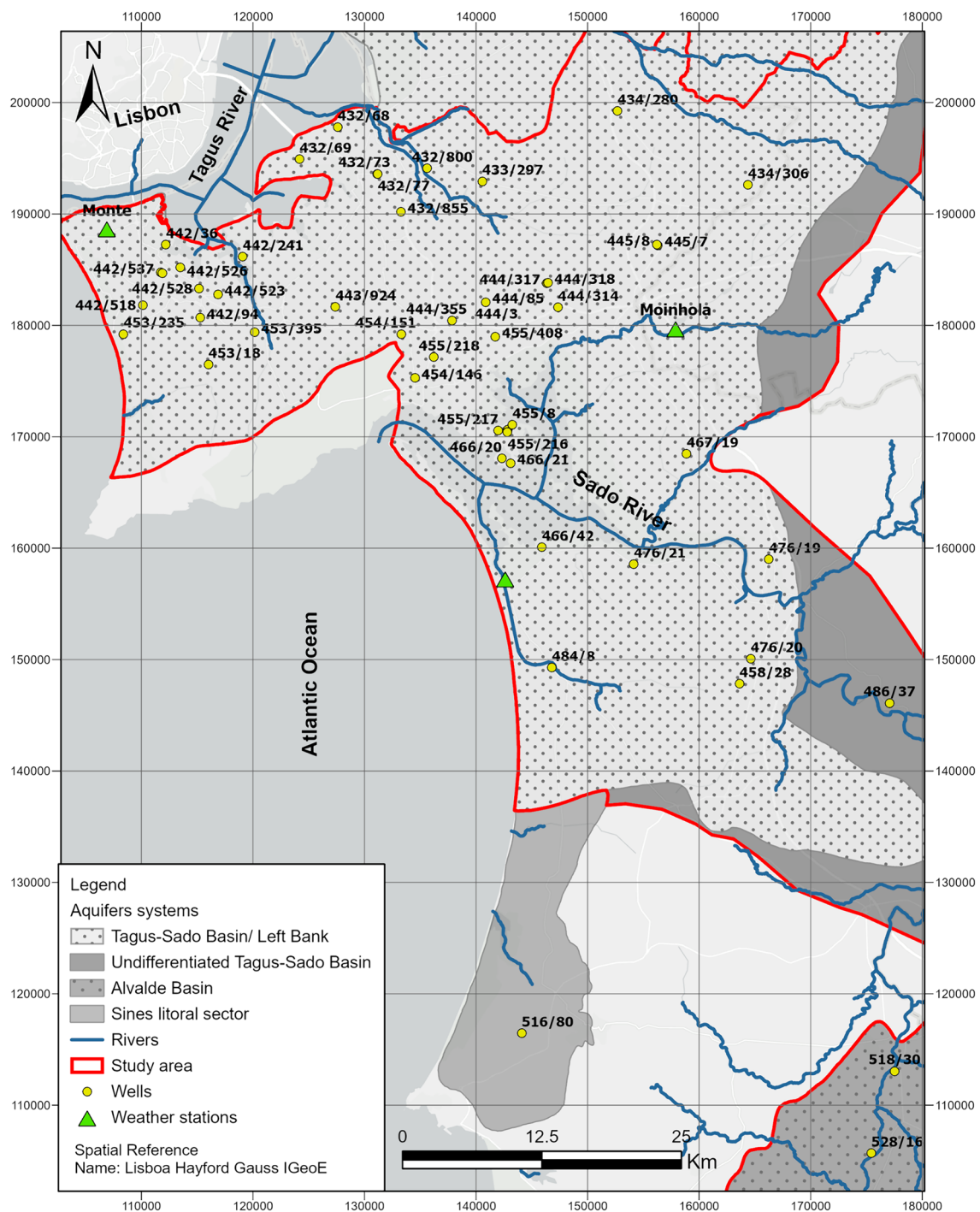
The main method of drainage of the basin occurs extensively in depth and the recharge of the aquifer system depends mostly in precipitation in which the infiltration of water flows into the soil, with descending and lateral flows, which then communicate with the water lines. The flow is conditioned by the two main water lines (the Tagus River to the North and the Sado River to the South), where the discharge of the system occurs. In the coastal sectors, the flow goes towards the Atlantic Ocean (Almeida et al. 2000). There are other types of flow, like local flows, whose discharge areas are the adjacent water lines and the recharge areas are the interfluves. These are flows where descending and lateral flow directions predominate. Also, there could occur intermediate groundwater flows, associated with one or more basins of the main tributaries (Almeida et al. 2000).

## Data set information

The selected statistical techniques were applied to understand common characteristics underlying the GDL trends of a set of wells that are used for public monitoring purposes by the Portuguese National Water Resources Information System Entity, SNIRH (“Sistema Nacional de Informação de Recursos Hídricos”, Fig. 2).

The considered wells were selected following a previous rigorous work, the wells that were considered for the study are those that represent undoubtedly the chronostratigraphic units selected for the case study and presented data series with sufficient representativeness for the chosen period of study, that is, from 1999 to 2021, which are define in the Table 1.

It can be said that the high increases in GDL found in some monitoring wells, particularly in the South sector of the basin, were the initial motivation for this study (Fig. 3).



**Fig. 2** GDL Monitoring wells considered for the study, located in the Central and Southern sectors of the aquifer systems of the Left Bank of the Tagus-Sado Basin (Portugal)

To gain a comprehensive understanding of the results, the climatological evaluation of the area between 2000 and 2020 was also analyzed, which are represented in the Fig. 4.

The data was collected from the Comporta, Moinhola, and Monte de Caparica weather stations (SNIRH). For the initial understanding of the climatological evolution

of data between years, data was analyzed every five years since the hydrological year 2000. From that data there is some evidence of a decrease of precipitation, in the South area according to weather station of Comporta.

In Monte and Moinhola stations, the rate of precipitation is decreasing but it has some fluctuation. However, the

**Table 1** Characteristics of the wells for the representatives analysis

Wells	Elevation (m)	Depth (m)	Levels observe (m)	Lithology
420/105	30	6	0–6	Sand
420/12	60	12	0–12	Sand
420/8	44	300	160–198	Sandy clay/coarse sand
420/9	44	95	72–85	Sandy clay
432/68	9	139.2	50–120	Clay/coarse sand
432/800	11	258	63–78; 99–228; 234–249	Limestone/sandy clay/carbonate sand
432/855	30	63	42–50	Sand
433/297	28	96	85–95	Medium sand
434/280	74	254	78–93; 108–150; 162–231	Sand /coarse sand/ sandy clay
434/306	78	137	74–79; 89.5–105; 119–135	Arkose sand/ clay/arkosic sand with clay
442/241	10	14.3	0–14.3	Sand
442/36	30	150	93–129	Arkose
442/537	21	142.45	67–84; 108–123; 126–129	Sandy clay / arkose/ limestone
442/94	17	234	88.5–99.5; 125–140; 143.5–145.5	Sandy clay/ arkose/limestone
443/924	59	279.6	96–102 and 132–138; 159–216; 222–261	Argillaceous arkose/ clay intercalates with sand/ calcareous arkose
444/317	44	252	189–243	Calcareous arkose
444/318	44	168.5	67–78; 84–108; 117 to 159	Coarse sand/ arkose/conglomerate
444/355	31	42	30–34	Arkose
444/85	37	10.3	0–10.3	Sand
445/7	76	162	80–162	Clay/arkose
453/18	70	387	Approximately 100–350	Arkose/marlstone
453/395	65	256	Approximately 100–250	Sandy clay and sand with carbonate sediment/Marlstone
454/146	20	10.9	0–10.9	Sand
454/151	149	16.7	0–16.7	Sand
466/21	94	268	153–175; 185–250	Siliceous Limestone/Calcareous Arkose sand
476/19	27	150	79.7–88.5; 100.3–130.9; 129.8–138.6	Arkose/Limestone/calcareous arkose
476/20	10	150	51.1–85.6	Arkose
476/21	20	150	73.8–112.1 and 118–141.6	Calcareous arkose/arkose
484/8	44	150	64.5–91	Arkose
489/37	24	100	69.8–75.7; 81.6–108.1	Calcareous coarse sand
516/80	64.86	36	24–30	Sandstone
518/30	51	100	73.3–94	Sand
519/300	75	30	12–22	Sand
528/16	44	100	73.5–94	Sand

average temperature has an erratic tendency, and doesn't have correlation with the precipitation rate. Therefore, this suggests the influence of other factors that promote this decrease in the last 20 years.

To understand how these tendencies manifest over time, for each technique, the selections of wells were considered according to the relevance of the information and the quantity of data.

### Multivariate analysis – factorial analysis of mixed data

FAMD consists of a multivariate factorial analysis technique in which it is assumed that there is no dependency between

the variables (independent variables), focusing simply on the relationship between the data.

The method combines the principles of Factorial Analysis (FA) with Multiple Correspondence Analysis (MCA), data is first transformed using MCA to create a set of synthetic variables that capture the underlying structure of the categorical variables then, FA is applied to the synthetic and continuous variables to extract the common factors that explain the variance in the dataset.

Quantitative variables are scaled to unit variance, while qualitative variables are transformed into a disjunctive data table (Husson et al. 2017, 2020). Thus, this technique allows data to be analysed and the balance and influence of the two

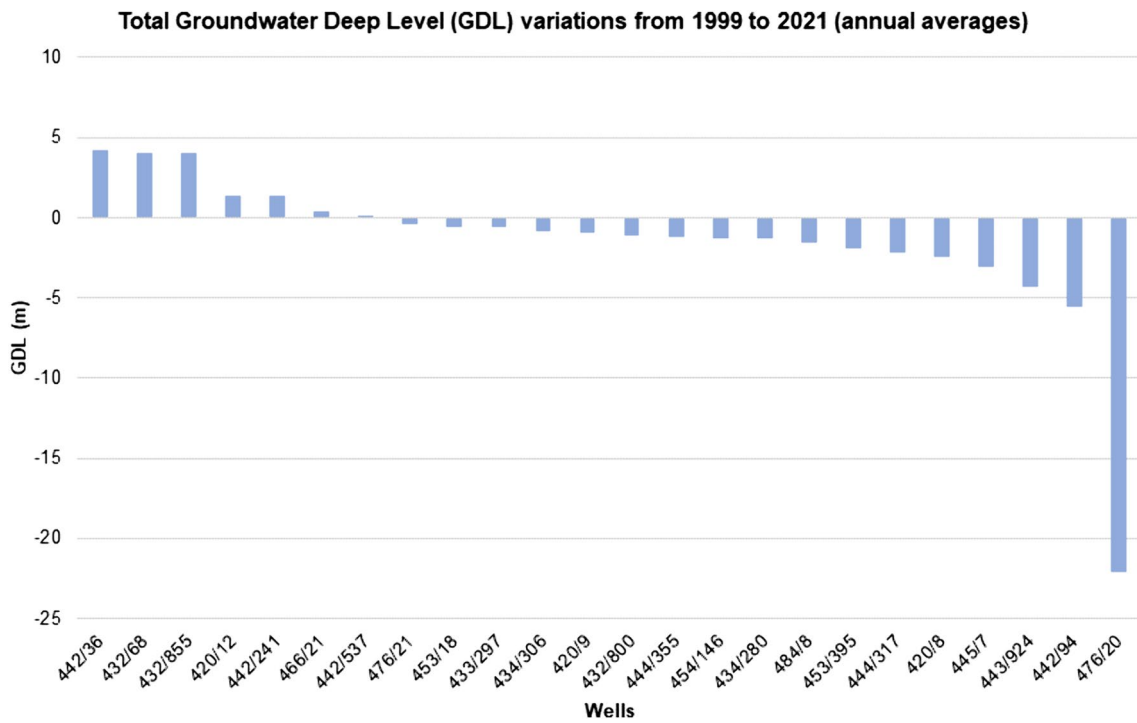
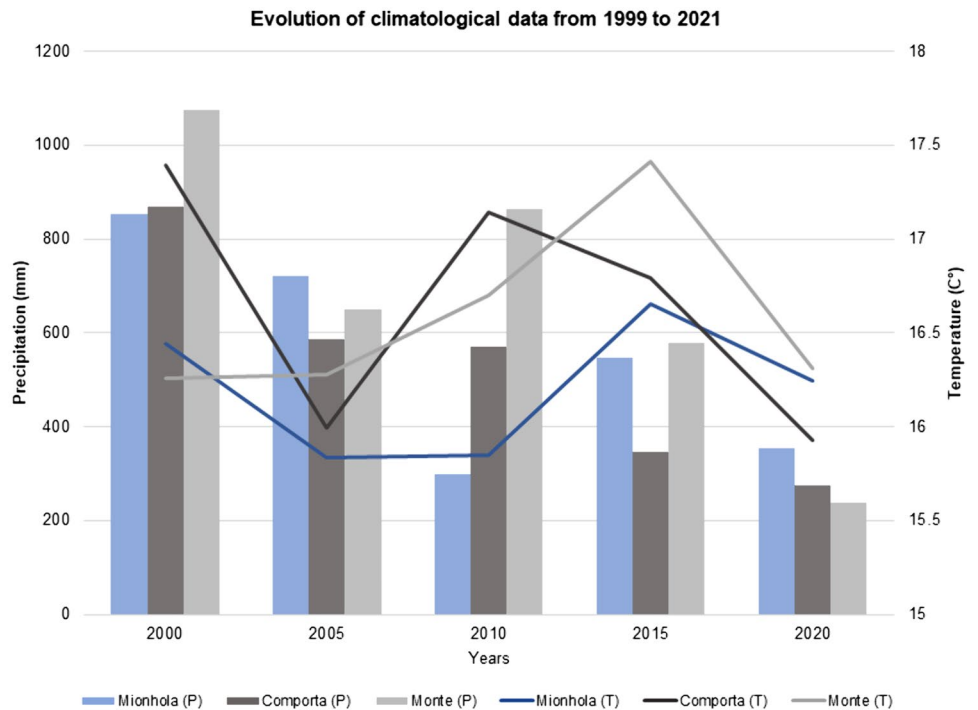


Fig. 3 Total GDL variations in distinct monitoring wells of the Left Bank of the Tagus-Sado Basin

Fig. 4 Evolution of climatological parameters – total annual precipitation (P) and annual mean of temperature (T) of each hydrogeological year, in distinct monitoring weather wells, from 1999 to 2021



types of variables to be assessed (Abdi 2003; Pagès 2004; Adamovic et al. 2016).

This method provides a comprehensive analysis of the dataset by uncovering the relationship among the variables

and identifying the most important factors that drive the variation.

According to Abdi (2003), Pagès (2004) and Adamovic et al. (2016), FAMD can be applied: (a) to data with few

qualitative variables compared to quantitative variables, (b) when the number of individuals in the population under study is generally low.

In the analysis of results, the representation of individuals in the data population is performed directly from factors, as a projection on the first two dimensions, where quantitative variables are represented through the circle of correlations associated with the PCA analysis and qualitative variables are represented in the same way as in MCA, in which the categorical variable is represented at the centroid of the individual who has it (Adamovic et al. 2016).

By performing this procedure, indicators “cos2” are obtained, these determine the representativeness of a variable and consist of the measure between the square of the cosine and the vector issued from the position of the variable and its projection on the axis (Husson et al. 2017, 2020; Lê et al. 2008). For indicator “cos2” values are close to 1 when a variable is well represented in the projection. According to Adamovic et al. (2016), usually, after performing this technique, it is advisable to proceed with HCA to complete the classification of individuals into groups that represent distinct well trend patterns.

As the mathematical concept, according to Audigier et al. (2016), the first step of FAMd consists of coding the categorical variables using the indicator matrix of dummy variables. For that, we have to initially define the information to the respective parameter, where  $I$  is the number of individuals,  $K_1$  is the number of continuous variables,  $K_2$  is the number of categorical variables, and  $K = K_1 + K_2$  is the total number of variables. Each continuous variable is a constructed matrix, where  $X_{ixj}$  is the matrix where  $(x_j)_{1 \leq j \leq K_1}$  are continuous variables  $(x_j)_{1 \leq j \leq K_1}$  and are dummy variables. The total number of columns is  $J = K_1 + \sum_{k=K_1+1}^K q_k$  where  $L$  is the number of categories of the variable  $k$ .

The PCA in the FAMd can be represented as  $((X - M)D_\Sigma^{-\frac{1}{2}})$ , where  $M_{I \times J}$  is the matrix with each row being the vector of the means of each column of  $X$  and  $D_\Sigma$  is the diagonal matrix  $diag(\sigma_{x_1}^2, \dots, \sigma_{x_{K_1}}^2, \dots, p_{K_1+1}, \dots, p_j, \dots, p_J)$ , being the standard deviation of the continuous variable  $x_j$  and  $p_j$  being the proportion of individuals in the category  $j(j = K_1 + 1, \dots, 1, \dots, J)$ . The matrix  $D_\Sigma$  is the metric used to compute distances between rows. The loss function (known as the reconstruction error), which in the PCA is minimized in matrix  $X$  is:

$$\|X - U\Lambda^{1/2}V'\|^2 \tag{1}$$

Thus, FAMd can be defined as minimizing:

$$\|(X - M)D_\Sigma^{-\frac{1}{2}} - U\Lambda^{1/2}V'\|^2 \tag{2}$$

FAMd provides the best low-rank ( $S < (J - K_2)$ ) approximation of the matrix  $(X - M)D_\Sigma^{-\frac{1}{2}}$  in the least square sense. The solution is given by the singular value decomposition (SVD) of the matrix  $(X - M)D_\Sigma^{-\frac{1}{2}}$ , with  $\hat{U}_{I \times S}$  the left singular vectors and  $\hat{V}_{J \times S}$  the right-singular vectors associated with the  $S$  largest singular values gathered in the matrix  $(\hat{\Lambda}_{S \times S})^{1/2} = diag(\sqrt{\hat{\lambda}_1}, \dots, \sqrt{\hat{\lambda}_S})$ . Notice that the maximum number of non-null eigen values is  $(J - K_2)$  because of the linear restrictions on the columns for the categorical variables (the row sum for each variable equals 1).

The specific weighting implies that the distances between two individuals in the initial space,  $i$  and  $i'$  (before approximating the distances by keeping the first  $S$  dimensions obtained from the SVD), is:

$$d^2(i, i') = \sum_{k=1}^{K_1} \frac{(x_{ik} - x_{i'k})^2}{\hat{\sigma}_{x_k}^2} + \sum_{j=K_1+1}^J \frac{1}{p_j} (x_{ij} - x_{i'j})^2, \tag{3}$$

Weighting by  $\frac{1}{\hat{\sigma}_{x_k}^2}$  ensures that units of continuous variables do not influence the (square) distance between individuals. Weighting by  $\frac{1}{p_j}$  unequivocally establishes that two individuals who belong to different categories for the same variable are significantly more distant from each other when one of them is in a rare category compared to when both of them belong to frequent categories. The marginal frequencies of the categorical variables play a crucial role in this method. Categories with a small frequency have a greater inertia than the others, and consequently, rare categories have a greater influence on the construction of the principal components. The specific weighting also implies that, in FAMd, the principal components maximize the associations with both continuous and categorical variables. More precisely, the first principal component  $f_1$ , maximizes:

$$\sum_{k=1}^K R^2(x_k, f_1) + \sum_{k=K_1+1}^K \eta^2(z_k, f_1), \tag{4}$$

With  $(z_k)_k = K + 1, \dots, K$  the categorical variables. The first principal component is the synthetic variable, which has the highest correlation with both the continuous variables (measured by the coefficient of determination,  $(R^2)$ ) and the categorical variables (measured by the squared correlation ratio  $(\eta^2)$ ). The second principal component is the synthetic variable, which maximizes the criterion among variables orthogonal to the first principal component and to the rest of the other principal component.



### Multivariate analysis – hierarchical classification analysis

In support of FAMD results, HCA was executed, it is a method of clustering data based on the similarity of the variables or cases.

In the case of FA, HCA can be used to identify groups of variables or cases that share similar factor scores. This can be particularly useful in identifying patterns or subgroups within a dataset that might not be immediately apparent from the initial FA.

Unsupervised classification is a technique used to determine homogeneous groups, defined as clusters, whose data presents similarities among themselves in a population. For this purpose, it includes techniques that use successive iterative processes, which process all objects in a set through divisive or agglomerative methods.

Agglomerative methods start with each case or variable as its own cluster and then merges clusters together based on their similarity, while divisive methods start with all cases or variables in one cluster and then split them into smaller clusters based on their dissimilarity. This process ends when all objects are processed (Almeida et al. 2007).

The choice of method will depend on the characteristics of the dataset and the research question being addressed. The HCA results are often presented in a dendrogram, where each linking step in the clustering process is represented by a connecting line (Smoliski et al. 2002; Granato et al. 2018). One common approach to conducting HCA in the context of Factorial Analysis is to use Ward's method, as in our case-study, which aims to minimize the sum of squared distances between the cases or variables within each cluster.

Other methods, such as single linkage or complete linkage, can also be used, depending on the research question and the characteristics of the dataset.

According to Nielsen (2016), this method groups all potential groups in pairs, in which the differences between the sum of the square of the differences of each object to the centroid is the smallest when comparing before and after the junction. The Ward linkage function is characterized as: To merge  $X_i(n_i = |X_i|)$  with,  $X_j(n_j = |X_j|)$  where consider the following Ward criteria:

$$\Delta(X_i, X_j) = \frac{n_i n_j}{n_i + n_j} \|c(X_i) - c(X_j)\|^2, \tag{5}$$

where  $c(X')$  denotes the centroid of the subset  $X' \subseteq X : c(X') = \frac{1}{|X'|} \sum_{x \in X'} x$ . Observe that the distance between two elements induced from the sub-set distance  $\Delta$  is merely half of the squared Euclidean distance:

$$\Delta(\{x_i\}, \{x_j\}) = D(x_i, x_j) = \frac{1}{2} \|x_i - x_j\|^2, \tag{6}$$

According to the equation in each iteration, the centroid of each group is calculated, and the sum of squared error (SSE) is calculated as the Euclidean distance between each object and the centroid of the group. Then, for all possible pairwise groupings, the hypothetical centroid resulting from the merger of the two groups is calculated, along with the sum of the Euclidean distances from each object to the hypothetical centroid. For each evaluated group merger, the value resulting from the sum of the two distances calculated separately is compared with the distance calculated already as a group (it is always an increment), and the similarity matrix is updated with these increments. Once the matrix is fully populated, the pair of groups leading to the smallest increment is merged. At the beginning of the process, each object is a group with zero dispersion, so the sum of dispersions for all groups is zero. In the end, only one group will remain, so the total dispersion is the sum of the squares of the differences between each value and the mean of the respective variable. The distance recorded in the dendrogram is the sum of the dispersions of the involved groups.

### Innovative trend analysis and innovative polygon trend analysis

ITA (Jiang et al. 2015) divides the time series into two equal subseries based on the average of its values, which are consequently classified separately in ascending order. Each subseries is subsequently distributed along two axes, the  $X_i$  subseries on the X axis, and the  $X_j$  subseries on the Y axis. With the subseries defined, the arithmetic means of each subseries are obtained, allowing for the determination of the slope value, according to Eq. (1) where n is the number of total time series data.

$$E(S_a) = \frac{2}{n} [E(\overline{X_2}) - E(\overline{X_1})], \tag{7}$$

If  $E(\overline{X_1}) = E(\overline{X_2})$ , then  $E(s) = 0$ , and the centroid of the trend line falls on the 1:1 line, indicating that there is no trend. This corresponds with the null hypothesis,  $H_0$ , where it is assumed that there is no trend if the calculated slope value,  $S_a$ , remains below a critical value,  $S_{cr}$ . Otherwise, an alternative hypothesis is applied,  $H_a$ , in which  $S_a > S_{cr}$  (Şen 2017). In this case, if  $X_1 > X_2$ , there is a negative trend, that is, there is a decrease, and if  $X_1 < X_2$  there is a positive trend, that is, there is an increase (Şen 2017).

This can also be verified by the variance  $V(S_a)$ , in which,

$$V(S_a) = E(S_a^2) - E^2(S_a), \tag{8}$$

In this situation the null hypothesis  $H_0$  assumes that  $V(S_a) = E(S_a^2)$  according to Eq. (3).

$$E(S_a^2) = \frac{E(\overline{X_2^2}) - 2E(\overline{X_2X_1}) + E(\overline{X_1^2})}{n^2/4}, \tag{9}$$

In the situation that  $E(\overline{X_2^2}) = E(\overline{X_1^2})$ , variance mean is equal to:

$$E(S_a^2) = \frac{E(\overline{X_2^2}) - E(\overline{X_2X_1})}{n^2/8}, \tag{10}$$

The correlation coefficient between two values is given according to its autocorrelation which is equal to:

$$\rho_{\overline{X_1}, \overline{X_2}} = \frac{E(\overline{X_2X_1}) - E(\overline{X_2})E(\overline{X_1})}{S_{\overline{X_1}} S_{\overline{X_2}}}, \tag{11}$$

where  $S_{\overline{X_1}} = S_{\overline{X_2}} = S/\sqrt{n}$ . Hence, the equation can be written according as:

$$E(S_a^2) = \frac{(1 - \rho_{\overline{X_1}, \overline{X_2}})}{n^2/8} * \frac{S_a^2}{n}, \tag{12}$$

where the correlation coefficient corresponds to the arithmetic means between the two subseries. Thus, the standard deviation of the sampling slope value can be obtained from Eq. (7).

$$\sqrt{E(S_a^2)} = \frac{2\sqrt{2}}{n\sqrt{n}} S \sqrt{1 - \rho_{\overline{X_1}, \overline{X_2}}}, \tag{13}$$

Under these circumstances, the third-order moment of the slope variable is also equal to zero, and the same is true for all odd-order moments. For this reason, the slope respects the normal (Gaussian) Probability Distribution Functions (PDF) with zero mean and standard deviation, thus being the basic criterion of this method (Almazroui and Şen 2020). As previously mentioned, this technique makes it possible to distinguish between series values in “low”, “medium” and “high”. Individualization and characterization of data is possible to be considered into 9 subareas, according to Fig. 5, with the next following interpretative means (Almazroui and Şen 2020):

1. Trendlines running parallel to the 1:1 line (45°) implies an increase or decrease, which is constant (B and C), while partial lines (D, E, I and L), encompass ratings below (“low”, “medium” or “high”). If the centroid is on the 1:1 straight line (45°), then there is a trend;

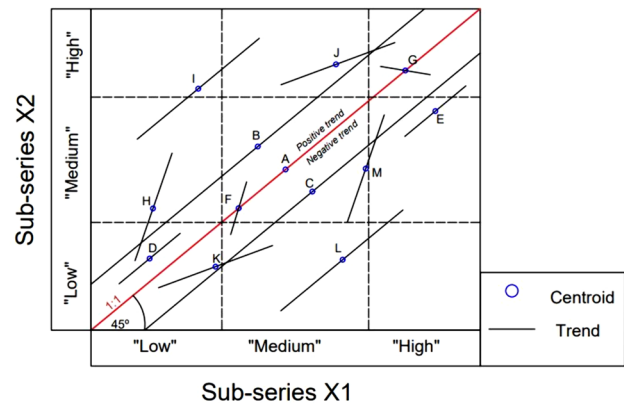


Fig. 5 Different partial trends considered by ITA technique (adapted from Almazroui and Şen 2020)

2. Non-parallel trend lines (F, G, H, J, K or M) to the 1:1 (45°) straight line implies a change in the deviation pattern over time;
3. Straight lines F and G have trends in standard deviation but not in arithmetic mean;
4. Trend lines, H (K) and J (M) imply an increasing (decreasing) standard deviation trend.

According to Şan et al. (2021), Şen et al. (2019) proposed the IPTA graphic method and revealed that it has certain advantages compared to other traditional trend methods in specific areas, such as the agricultural field and in hydrological studies. The ability to identify trends in a sequence is one of its significant characteristics that provides a highly productive basis for better linguistic and numerical interpretation and deduction.

If IPTA is applied to monthly data, according to Şen et al. (2019) and Şan et al. (2021), five processing steps are required. Step (1): monthly (e.g. Jan., Feb., Mar...) time series is divided into two equal periods. Step (2): basic statistics (e.g. mean, max) or desired criteria (e.g. uncertainty) for each month are calculated in both periods. Step (3): the first (second) period is placed on the horizontal (vertical) axis in the scatter chart, and 12 points are marked for 12 months. Step (4): the points of consecutive months are connected by straight lines forming a polygon (Fig. 2). Step (5): the slope and size of the line between consecutive points are calculated.

$$|AB| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{14}$$

$$s = \frac{y_2 - x_2}{y_1 - x_1} \tag{15}$$

where  $s$  is the trend slope,  $|AB|$  is the trend length,  $x_1$  and  $x_2$  are two consecutive points in the first part in horizontal,  $y_1$  and  $y_2$  are two consecutive points in the second part. In the Cartesian coordinate system, a 1:1 (45°) line is drawn, where the points are below (above) the line representing a decreasing (increasing) trend (Şen 2012). Straight lines connecting the points provide information about the changes between the successive months. If the slopes of the lines between consecutive months are far from each other, the contribution of the changes between months to the average change in the hydrometeorological series is significant and vice versa (Şan et al. 2021).

### Mann–Kendall analysis

According to diverse authors (Hamed 2008; Güçlü 2020; Mirabbasi et al. 2020) the MKA developed by Mann (1945) and revised by Kendall in 1948 (NLJ 1948), assumes that the trend value to be obtained is calculated according to the Eq. (8), and (9).

$$\text{sgn}(x_j - x_i) = \begin{cases} 1; & \text{If } x_j > x_i \\ 0; & \text{If } x_j = x_i \\ -1; & \text{If } x_j < x_i \end{cases}, \quad (16)$$

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (17)$$

Each pairwise of a data series is classified in one of the three subsets (-1, 0, 1) according to their differences (Eq. (8)). In Eq. (9)  $x_i$  and  $x_j$  are the data values for time  $i$  and  $j$ , and  $n$  is the dataset length. When the  $S$  value is positive, it indicates an increasing trend, otherwise, it verifies a decreasing trend. When a data length of a series is higher than  $n > 10$ , the data distributions approach a normal distribution law, with a mean equal to zero, where the variance [ $\text{Var}(S)$ ] assumes the following expression (Eq. (10)):

$$\text{Var}(S) = \frac{n - (n - 1)(2n + 5) - C}{18}, \quad (18)$$

where  $C$  is a factor for modified variation and  $n$  represents the number of defined groups. In the presence of data series with the occurrence of successive data, the parameter  $C$  is calculated by the following formula:

$$C = \sum_{i=1}^m t_i(t_i - 1)(2t_i - 5), \quad (19)$$

where  $t_i$  defines the number of linked data in the group. If there are no groups, it automatically skips this process. After

calculating the variance of the series data, the standard  $Z$  value is calculated according to the following equation:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}}; & \text{If } S > 0 \\ 0; & \text{If } S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}}; & \text{If } S < 0 \end{cases}, \quad (20)$$

In the MK analysis, the null hypothesis  $H_0$  represents “no significant trend in the time series”, and it is accepted if the confidence level,  $\alpha$ , is:

$$-Z_{\left(1-\frac{\alpha}{2}\right)} \leq Z \leq Z_{\left(1+\frac{\alpha}{2}\right)}. \quad (21)$$

If this hypothesis is not verified,  $H_0$  is rejected, and the alternative hypothesis (that is, existence of a significant trend in terms of importance) is accepted (Dinpashoh et al. 2014).

In the graphical results of MKA information is divided into two different lines: the prograde and the retrograde line. The prograde line corresponds to  $Z(S)$  values, and the retrograde line represents  $Z^*(S)$ . When these two lines touch each other, we are under an event that represents an inversion of the tendency. With this, it is possible to determine at which point the transition occurs. The arrangement of lines gives us information about the type of tendency and the confidence level interval value (CL). If the line prograde is above the other line, that indicates a positive tendency, which means an increase in GDL and if the opposite situation manifests, that indicates a decrease in GDL. However, for the analysis to be correct, it is necessary for the two lines to be inside the CL value of 95%. In the case where just one of the lines is inside that interval, the tendency is represented by the line inside the CL if 95% prior to the inversion. For the null hypothesis to occur, the two lines must touch each other constantly with a minimal gap in between.

## Results

### FAMD and HCA

Information was analysed between 2006 and 2021 for the monitoring wells defined in Table 2.

In this study, the years represent the quantitative variables, and the wells represent the qualitative variables. Figure 6 presents the distribution of the qualitative variables (that is, the monitoring wells) in Dimension 1 (Dim 1), and Dimension 2 (Dim2) axes, it is possible to observe the projection of the contribution of each qualitative variable for

**Table 2** Average value of water level depth (m) from 2006 to 2021 in distinct monitoring well

Wells	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
420/105	0.9	0.9	0.9	0.9	0.9	1.1	1.1	1.1	1.1	1.1	1.1	1.4	1.5	1.1	1	1.1
420/12	6.8	6	6.2	7.1	6.2	5.5	6.9	5.9	5.5	6.7	6.2	7	6.8	7.3	7.1	6.9
420/8	28.1	26	26.7	29.4	27.5	27.2	27.1	27.9	26.9	27.3	25.7	33.3	26	28	27.4	29
420/9	27.1	28.5	26	28.6	26.8	26.7	26.4	27.1	26.1	26.4	25.1	32.4	28	27.1	26	28.3
432/68	9.8	9.5	9.7	10.4	9.8	9.4	9.4	9.9	9	11.7	8.8	11.8	9.7	9.9	10.1	9.8
432/800	2.6	1.9	1.2	3.1	2.3	2	1.9	2.5	1.8	2.3	1.1	3.3	3	3.2	3.1	4.2
432/855	21.3	21.1	21	21.6	21.4	20.7	20.4	19.9	18.9	18.6	18	18.1	17.6	17.7	17.5	16.8
433/297	13.7	13.3	13.3	14.2	13.7	13.1	13	13.5	13.2	13.1	12.7	14	14.2	14.3	14.1	14.2
434/280	35.1	36.5	35.1	36.6	35.5	35.3	35.1	35.7	35.3	35.4	34.5	35.3	36.1	36.5	36.3	37.4
434/306	46.9	47	46.9	47.1	47	47	47	47	47.2	47.1	46.9	47.3	47.2	47.3	47.3	47.2
442/241	3.7	3.3	3.7	3.8	2.9	3	3.2	3.4	3.1	3.2	3.4	3.9	3.9	4	4	4.4
442/36	11.4	8.8	8.8	8	6.3	5.4	5.7	6.2	5.4	5.6	4.3	6.3	5.7	6.7	7.5	8.2
442/537	33.7	33.6	32.9	31.5	32.9	33.1	31.9	31.7	31.1	30.5	31.2	30.1	31	32.4	31.2	33.4
442/94	20.1	20.2	20.3	19.8	20.8	20.8	20.1	19.9	19.8	19.4	18.7	19.3	19.2	20	19.9	22
443/924	52.8	52.8	52.7	53.6	53.4	53.2	52.7	53.2	52.6	53	51.3	53.7	53	53.4	54.4	55.5
444/317	25.5	25.1	25.5	26.3	25.8	25.2	25.3	25.7	24.9	25.9	24.4	26.9	27	27.4	27.4	27.2
444/355	16.2	16.1	16.3	16.7	16	14.7	14.6	14.7	14	14.1	14.4	15.9	15.7	16.5	17	17.6
445/7	51.7	51.4	51.7	52.8	52	51.7	51.8	51.9	51.8	52.2	51	53.5	52.3	51.9	53.5	55.5
453/18	58.9	58.9	59	59.7	59.5	59.4	59.1	58.8	58.8	58.6	57.7	58.7	58	58.8	58.9	60.1
453/395	29.5	29.7	29.5	29.9	30	29.4	29	29.2	28.4	28.3	27.3	28.8	27.4	26.9	31.4	28.6
454/146	9.1	8.6	9.2	9.6	8.6	7.6	8.1	7.8	7.3	7.5	7.8	8.8	9.2	9.8	9.5	9.9
466/21	15	14.2	13.4	15.9	17.2	17.8	16.9	16.8	17.3	17.8	16.9	19.2	17.1	16.3	18.2	18
476/20	35.4	34.3	34.3	35.8	37.3	37.3	38.7	39.2	39.3	39.8	41.1	41.2	42.1	41.9	55.4	59.2
476/21	9.6	9	8.6	9.1	8.9	9.3	9.6	9.3	9.2	9.2	8.7	10.4	9.8	10.4	9.4	10.5
484/8	-0.1	-0.8	-0.9	-0.6	-0.8	-1	-1	-0.5	-1	-0.7	-0.3	-0.7	0	0.6	0.9	1.1
518/30	17.2	15.4	13.3	13.3	13.3	16.7	17.2	15.2	16.3	17.9	18.8	19.7	18.2	18.5	17.5	21.7
528/16	11.8	11	10.1	10.1	10.1	10.1	10.8	9.9	9.6	9.8	10.3	10.5	10.3	11.1	12.2	14.8

dimensions 1 and 2 (axes values) and its classification for the indicator “cos2” simultaneously (colour gradient scale).

For a better understanding of the projection results and on the meaning of each dimension, the well “476/20” was not considered in the initial analysis, due to its behaviour, this well shows a very marked downward trend with a difference of GDL close to 24 m between 2006 and 2021 (Table 2) having a behaviour of extreme outlier.

In the performed analysis, Dim 1 and Dim 2 represents about 72% of the variance. The distribution of qualitative variables is dispersed, indicating the existence of subpopulations with distinct tendencies in time. Dim1 is conditioned by the monitoring wells with the most irregular trends, from lower depths of groundwater levels (contribution close to -5, well 484/8) to higher depths of groundwater levels (contribution close to 9, well 453/18) (Fig. 6). Monitoring wells with the most

constant trends are positioned in Dim 1 close to “0”, along Dim 2 (Fig. 6).

The Fig. 7 presents the FAMD results considering well “476/20” and other two wells included in a posterior stage of the analysis (518/30, and 528/16), it is to say that, although the outlier behaviour of variable “476/20”, the variance results for Dim1 and Dim2 don’t vary significantly. For this reason, and due to the importance of this monitoring well in the performed analysis, it was decided to consider FAMD results with the outlier “476/20”.

In these conditions (Fig. 7), the monitoring wells with the highest scores in Dim2 present more accentuated non-constant tendential patterns on the depths of groundwater levels. The Dim2 in these conditions is, therefore, conditioned by well “476/20” which evolves from constant stationary trends (negative score contributions in Dim2 and closer to “0” in Dim1) to less constant trends (increasing or decreasing

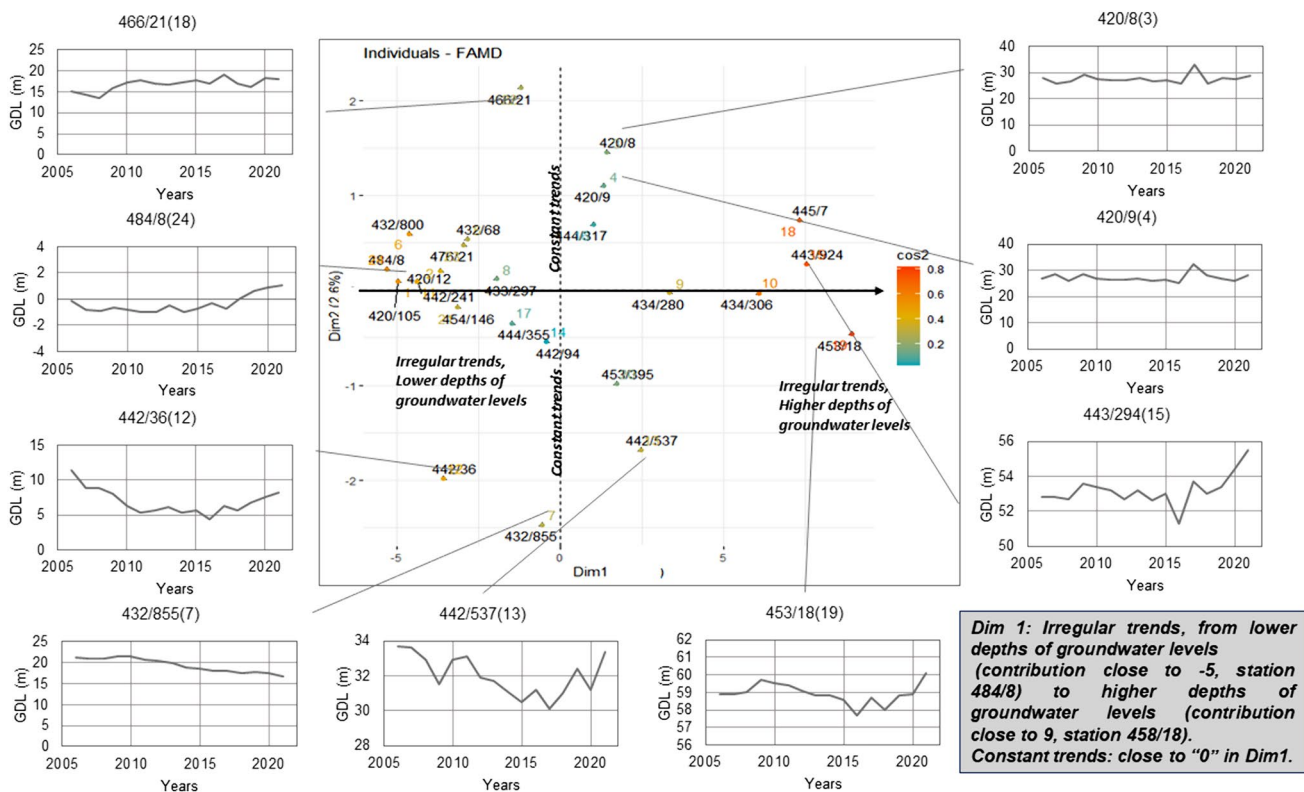


Fig. 6 Projection of the contribution scores of qualitative variables in Dim1 and Dim2 without the consideration of monitoring well “476/20”

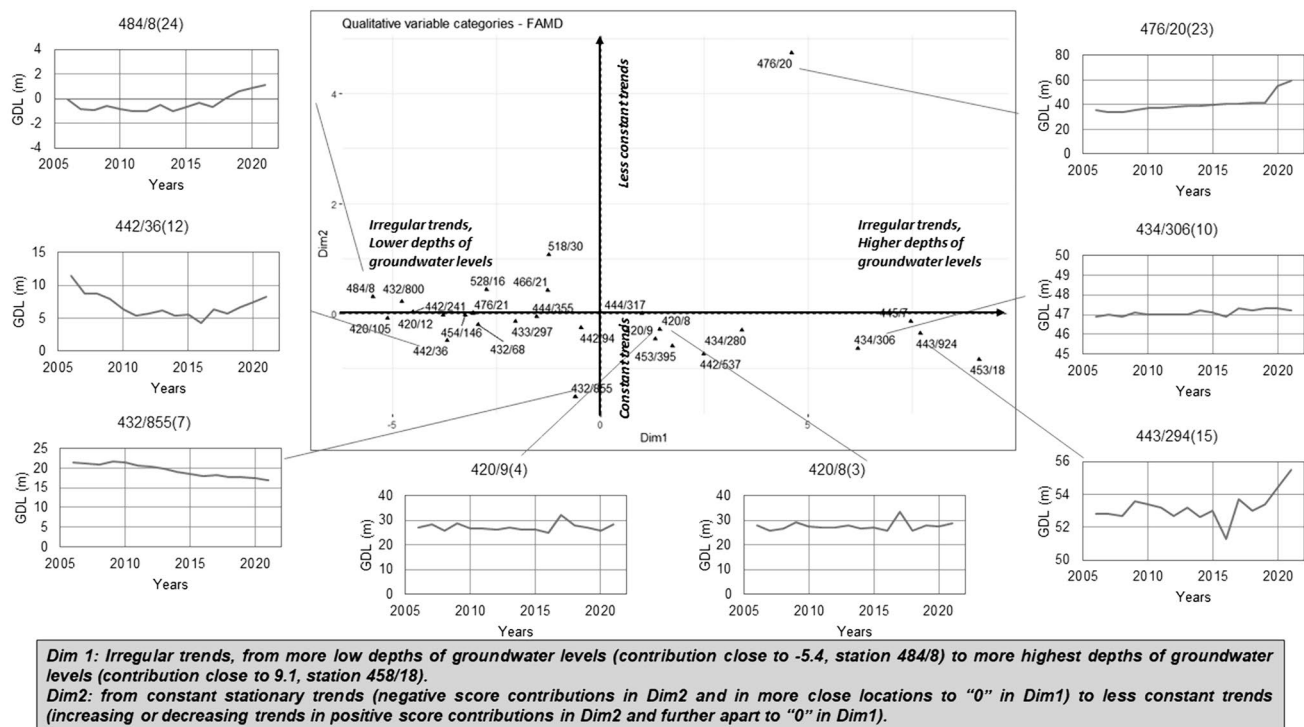
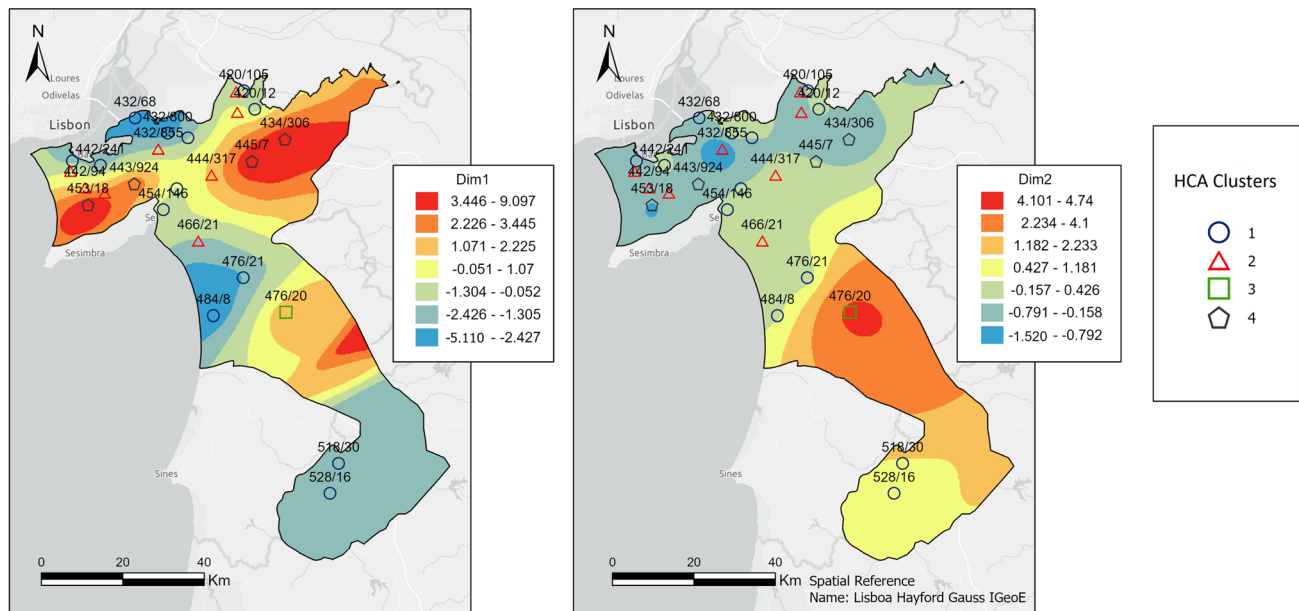


Fig. 7 Projection of the contribution scores of qualitative variables in Dim1 and Dim2 with the consideration of monitoring well “476/20”



**Fig. 8** Interpretational kriging maps of FAMD contribution scores of Dim1 and Dim2, and locations of HCA cluster classification

**Table 3** The scores of contributions to Dim 1 and Dim2 of each well station

Wells	Individuals	Dim1	Dim2
420/105	1	- 5.110	- 0.092
420/12	2	- 3.777	- 0.034
420/8	3	1.431	- 0.284
420/9	4	1.326	- 0.459
432/68	5	- 2.939	- 0.209
432/800	6	- 4.771	0.220
432/855	7	- 0.601	- 1.520
433/297	8	- 2.035	- 0.143
434/280	9	3.398	- 0.306
434/306	10	6.190	- 0.640
442/241	11	- 4.502	0.024
442/36	12	- 3.682	- 0.492
442/537	13	2.488	- 0.742
442/94	14	- 0.460	- 0.266
443/924	15	7.688	- 0.363
444/317	16	1.000	- 0.002
444/355	17	- 1.533	- 0.055
445/7	18	7.463	- 0.149
453/18	19	9.097	- 0.841
453/395	20	1.735	- 0.593
454/146	21	- 3.252	- 0.028
466/21	22	- 1.265	0.419
476/20	23	4.600	4.740
476/21	24	- 3.058	0.007
484/8	25	- 5.466	0.308
518/30	26	- 1.233	1.068
528/16	27	- 2.732	0.431

trends in positive score contributions in Dim2 and further apart to “0” in Dim1). Again, Dim1 is conditioned by the monitoring wells with the most irregular trends, from lower depths of groundwater levels (contribution of -5.4, well 484/8) to higher depths of groundwater levels (contribution of 9.1, well 458/18) (Fig. 7).

It is also observed an increasing tendency in the depth of groundwater levels in the monitoring wells with the highest score contributions for Dim1, like it is the case of wells 434/306, 442/94, 443/924, 445/7 and 453/18. The location of these monitoring wells can be observed in Fig. 8, which includes the interpretational kriging maps of the FAMD contribution scores (Table 3) for Dim 1 and Dim2 and the subsequent HCA classification of monitoring wells in 4 clusters. These results evidence the subsector in the Basin with deepest groundwater levels (scores with red and orange colours in Dim1), and the areas with groundwater levels closest to the ground surface (scores with blue and green colours in Dim1).

From the score mapping of Dim2 in Fig. 7, it is also possible to observe a more constant stationary behaviour in GDL of the monitoring wells located at the North sector of Basin while towards the South sector, there is a growing trend to greater oscillations, mainly with an increasing in deep of groundwater levels as is the case of wells 476/21, 476/20, 484/8, 518/30 and 528/16, which present positive values in Dim2.

Monitoring well 476/20 has the heights increasing tendency the deep of groundwater levels (with more than 20 m in depth). The score map of FAMD results for Dim 1 and Dim 2 (Fig. 8) gives a good perspective of the distinct

**Table 4** Interpretation of HCA results considering the classification in 4 clusters

Cluster number	Wells	Evolution	Trend
1	420/105, 420/12, 432/68, 432/800, 442/241, 442/297, 442/36, 444/355, 454/146, 476/21, 484/8, 518/30 and 528/16	Less constant	Low increasing and irregular
2	420/8, 420/9, 432/855, 434/280, 442/537, 442/94, 444/317, 453/395 and 466/21	Constant	Stationary increasing and decreasing and less irregular
3	470/20	Stationary to Less constant	High increasing and less irregular
4	434/306, 443/924, 445/7 and 453/18	Constant	High Increasing and irregular

GDL trend groups in the aquifer systems between 2006 and 2021.

For a better understanding of the different characteristics of the distinct well trend groups, HCA was applied to the contribution score values of obtained from FAMD. The data was subjected to an iterative process to find the right number of clusters which represent the most adequate for the population. From that process, the number of clusters considered was 4 (Table 4). From the observation of Fig. 8 it is possible to understand the localization of the HCA clusters in the study area. It is possible to verify that HCA applied to FAMD scores enhances the results of FAMD making it possible do have a finer discretization of the results.

### ITA and MK

For ITA method, monthly information of GDL was used, from each well, to detect small variations in trends. In this way, all samples from Table 1, which refers information from 2006 until 2021, were used. It also included 4 wells with data from 2011 to 2021 and 3 others from outside the main aquifer system to understand the tendencies in some areas, which was not possible to conclude in the other analysis. The achieved results are presented in Table 5 and also in the Appendix A (from Figs. 11, 12, 13, 14, 15, 16, 17, 18, 19).

ITA slope kriging results are presented in Fig. 9, it is possible to observe that there is a more pronounced increase in GDL in the South sector of the Basin as it has been verified previously with FAMD. There are also situations where a decrease in GDL is observed (432/855, 442/36, 442/537, 442/94, 444/355, 453/18, 453/395, 454/151 and 476/19), and in cases even a constant tendency is verifiable (444/318 and 454/146).

Through the analysis of the graphs obtained from ITA, it is possible to verify two situations, which are also reflected in FAMD analysis:

- There are monitoring wells with constant tendencies and monitoring wells with irregular tendencies. The monitoring wells with a more constant behaviour mostly show positive trends, defined in the confidence interval ( $\alpha$ ). The monitoring wells with irregular behaviours (10 wells) can present general negative trends of GDL (in 453/151, 432/855 442/36) or general positive trends, as is the case of 420/105, 432/68, 432/800, 476/20, 476/21, 484/8, and 528/16. Higher slopes (positive and negative) present a more pronounced anomalous trend, suggesting that anomalous behaviours register a very significant variation in short periods, which is the case of well 476/20 (Fig. 6).
- The monitoring wells with constant tendency behaviour have smaller slope standard deviation values. All found correlation values are quite high, greater than 0.8 and in most of the cases close to 0.99, except for one monitoring well (420/105) that has a low correlation value, less than 0.5, which results from an episode in which the groundwater depth, at first, increased sharply.

Through Fig. 9 it is possible to verify that in the Southern sector and in a smaller area of the North study area, increasing tendencies of GDL (that is, increases in groundwater depth levels) occur, while the more coastal sectors and areas close to estuaries zones present a certain slight tendency to decreases in GDL. However, the well 466/21 shows a different tendency, with an increasing of GDL, which could result from exploitation in the area. Hence, from this result it possible to visualize an initial transition to the decreasing of GDL as shown in the Appendix A.

MKA method was applied to complement ITA results, and to reverify the tendencies observed with the previous methods. Like ITA, the graphical results of MKA are also presented in Appendix A.

It is possible to observe that most of the wells display a positive trend of the GDL, except in the cases of 432/855,

**Table 5** ITA results

Wells	Number of data	Slope, s	Correlation, $\rho_{y1y2}$	Sig-nificance level	Lower CL	Upper CL	Hypothesis	Decision	Type of trend
420/105	189	0.0016	0.4940	0.10	-0.00025	0.00025	Ha	Yes	Increasing
420/12	189	0.0037	0.9800	0.10	-0.00022	0.00022	Ha	Yes	Increasing
420/8	189	0.0028	0.9788	0.05	-0.00085	0.00085	Ha	Yes	Increasing
420/9	189	0.0055	0.9866	0.05	-0.00085	0.00085	Ha	Yes	Increasing
432/68	189	0.0128	0.8217	0.10	-0.00170	0.00170	Ha	Yes	Increasing
432/800	189	0.0081	0.9909	0.10	-0.00032	0.00032	Ha	Yes	Increasing
432/855	189	-0.0316	0.9444	0.10	-0.00072	0.00072	Ha	Yes	Decreasing
433/297	173	0.0042	0.9882	0.10	-0.00022	0.00022	Ha	Yes	Increasing
434/280	189	0.0040	0.8590	0.05	-0.00137	0.00137	Ha	Yes	Increasing
434/306	172	0.0019	0.9731	0.01	-0.00018	0.00018	Ha	Yes	Increasing
442/241	190	0.0038	0.9883	0.10	-0.00010	0.00010	Ha	Yes	Increasing
442/36	190	-0.0334	0.9839	0.10	-0.00043	0.00043	Ha	Yes	Decreasing
442/537	189	-0.0216	0.9043	0.10	-0.00094	0.00094	Ha	Yes	Decreasing
442/94	190	-0.0057	0.9474	0.05	-0.00052	0.00052	Ha	Yes	Decreasing
443/924	190	0.0030	0.9892	0.01	-0.00058	0.00058	Ha	Yes	Increasing
444/317	177	0.0060	0.9760	0.05	-0.00049	0.00049	Ha	Yes	Increasing
444/318	71	-0.0009	0.9530	0.01	-0.00126	0.00126	H0	No	-
444/355	189	-0.0022	0.933810	0.05	-0.000644	0.000644	Ha	Yes	Decreasing
444/85	159	0.0046	0.963336	0.10	-0.001037	0.001037	Ha	Yes	Increasing
445/7	226	0.0077	0.975342	0.05	-0.000336	0.000336	Ha	Yes	Increasing
453/18	190	-0.0040	0.947998	0.01	-0.000630	0.000630	Ha	Yes	Decreasing
453/395	179	-0.0124	0.865201	0.10	-0.001740	0.001740	Ha	Yes	Decreasing
454/146	188	0.0002	0.955813	0.05	-0.000425	0.000425	H0	No	-
454/151	132	-0.0165	0.853812	0.10	-0.001721	0.001721	Ha	Yes	Decreasing
466/21	189	0.0167	0.982662	0.10	-0.000449	0.000449	Ha	Yes	Increasing
476/19	85	-0.0230	0.982020	0.10	-0.001035	0.001035	Ha	Yes	Decreasing
476/20	132	0.1407	0.979033	0.10	-0.003206	0.003206	Ha	Yes	Increasing
476/21	132	0.0159	0.982995	0.10	-0.000543	0.000543	Ha	Yes	Increasing
484/8	132	0.0156	0.964235	0.10	-0.000568	0.000568	Ha	Yes	Increasing
489/37	136	0.0072	0.938116	0.05	-0.000526	0.000526	Ha	Yes	Increasing
516/80	172	0.0091	0.949980	0.05	-0.000538	0.000538	Ha	Yes	Increasing
518/30	136	0.0371	0.978441	0.10	-0.001067	0.001067	Ha	Yes	Increasing
519/300	60	0.0027	0.899623	0.05	-0.001851	0.001851	Ha	Yes	Increasing
528/16	136	0.0312	0.902131	0.10	-0.001693	0.001693	Ha	Yes	Increasing

442/36, 442/537, 442/94, 444/355, 453/18, 453/395, 454/151 and 476/19, as it was stated in ITA.

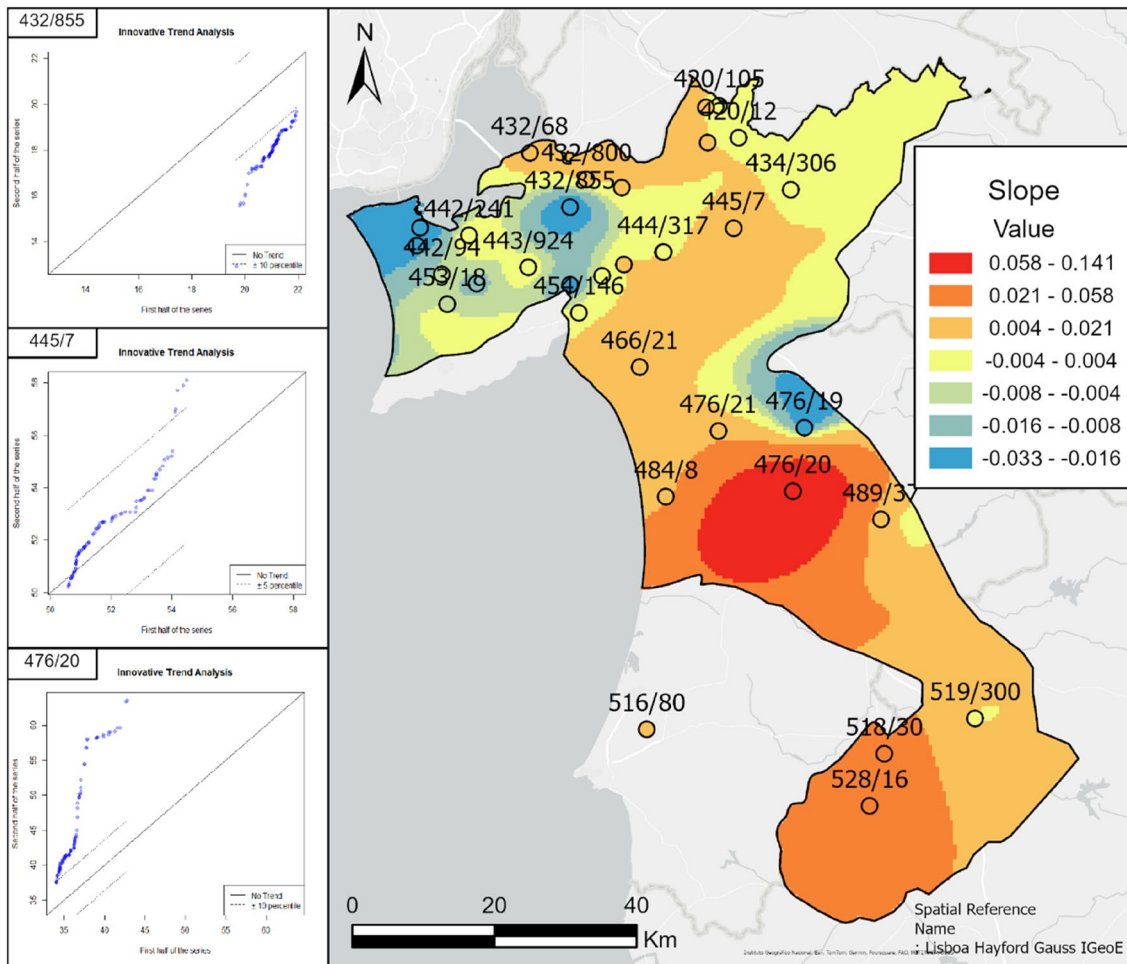
Positive trends of GDL are predominant between 2006–2016 and 2018–2021, where some wells that display a minor transition period between 2008–2009 and 2017–2018. In 2018–2021, the slope of both lines is significant, which shows that in the last years the rate of GDL increase was escalating.

Negative trends in GDL began to manifest after the beginning of 2007 and in 2010 and 2011. The wells 432/855 and

454/151 display a GDL negative trend between 2015 and 2021. From the previous analyses, the well 476/20 represents a very significant positive trend since the beginning of 2019. It is also possible to verify that wells 434/280, 434/306, and 443/924, have three to five breaking points at different periods, which reinforces the already irregular profiles detected. In general, wells with anomalous tendencies have significant slopes in ITA after the breaking points in MKA.

After analysing the ITA and MK, was applied the IPTA methodology to complement the analyses of ITA as well.





**Fig. 9** Map of ITA slope kriging results

Ten wells, were analysed, 5 in the south sector and the other 5 in the North sector. The results obtained, which are in Appendix B present some clarifications. It is possible to visualise that, in a general way, almost wells that show an increase in GDL manifest this trend for most months of the year, especially in the wells located in the southern sector of the study area. The wells with GDL decreasing behaviour show the opposite behaviour. It suggests that depletion and rising effects have been occurring consistently over the years.

By careful analysis, it becomes evident that some wells share a similar pattern in their monthly evolution. In wells 476/20 and 528/16 between the months Dec-Jan, there is an inflexion, whose slope is accentuated and coincides with the maximum trend line, indicating a tendency for the depth to increase in the month of Dec over the years in comparison to the month Jan. This suggests a decrease in the recharge

rate in the southern sector in the winter months, with other factors of a structural nature in the basin. The wells 445/7 and 466/21 show a more constant trend of increasing depth over the months, except for the month of Aug for the well 445/7, where there is a decrease in GDL, indicating the possibility of recharge in the system by anthropogenic means (agricultural areas), to mitigate the effects of drought. The well 476/21, in comparison to the remaining wells, observes an inversion between the months Sep-Oct, corresponding to one of the maximum trend lines, followed by the months Apr-May. In addition, the month of Jan shows no trend relative to the other months. This could be due to changes in the recharge effect and rising sea levels during the dry season. In 442/241, there is an increase in GDL over the months, especially in the summer months, which is conditioned by the recharge rate.

In the wells 453/18 and 432/855 show irregular behaviour. The significant decrease observed in the well 432/855, could result from the combination of the rise of sea level and possibly from precipitation. In the well 453/18, the rise of sea level could also explain that effect in the summer combine with the precipitation but not so significant as the well 432/855. The well 476/19, on the other hand, shows a constant trend over the months, and in the months of Aug, Sep and Oct, this effect is significant due to the rise of sea level. The well 420/12 shows variations in the tendency, where the months Jan, Apr, May and Aug present a decrease in GDL, Feb, Mar, Jun, Jul, Nov and Dec, present a decrease in GDL, and last Sep and Oct don't show tendency at all. This suggests a combination of factor like anthropic and the climate change.

## Discussion and conclusions

The evaluation of annual and long-term changes in groundwater reserves, namely the variations in groundwater depth levels (GDL), assumes an enormous relevance in the context of hydrogeological resource management. Among others, the GDL trend analysis allows to estimate indirectly the groundwater recharge rate effectiveness, and allows to determine its gradients, to understand dynamic evolutions of aquifer systems and to design efficient and sustainable exploitation of groundwater wells.

In general, FAMD associated with HCA allows to obtain more reliable results in cases where the amounts of available data are small.

However, there were some difficulties and incongruencies in the clustering analysis process, probably due to the reduced number of observation wells when compared to the dimension of the study area, and the extreme outlier behaviour associated with the well "476/20" which affected the results.

When comparing the FAMD and ITA, there are correspondences in the detected trends observed with both methods, without a doubt, ITA enables significantly a more efficient analysis of periods and can effectively leverage a larger pool of observation data in comparison to FAMD.

In this case-study, variations in GDL are likely relatable to deficiencies of the recharge rate of the aquifer system. From ITA it was possible to observe that the high slope values, that is, the most demarcated trends, occur central area of the Basin and the inverse situations in the bay areas.

With the IPTA analyses it's also possible to conclude that in most of the years, the increases and decreases of the GDL

occur almost in all months. The evolution along the month reveals also that in some locations the combination of different factors, severely conditionate the transition between months, creating irregular polygons, showing the evolution of GDL is not constant.

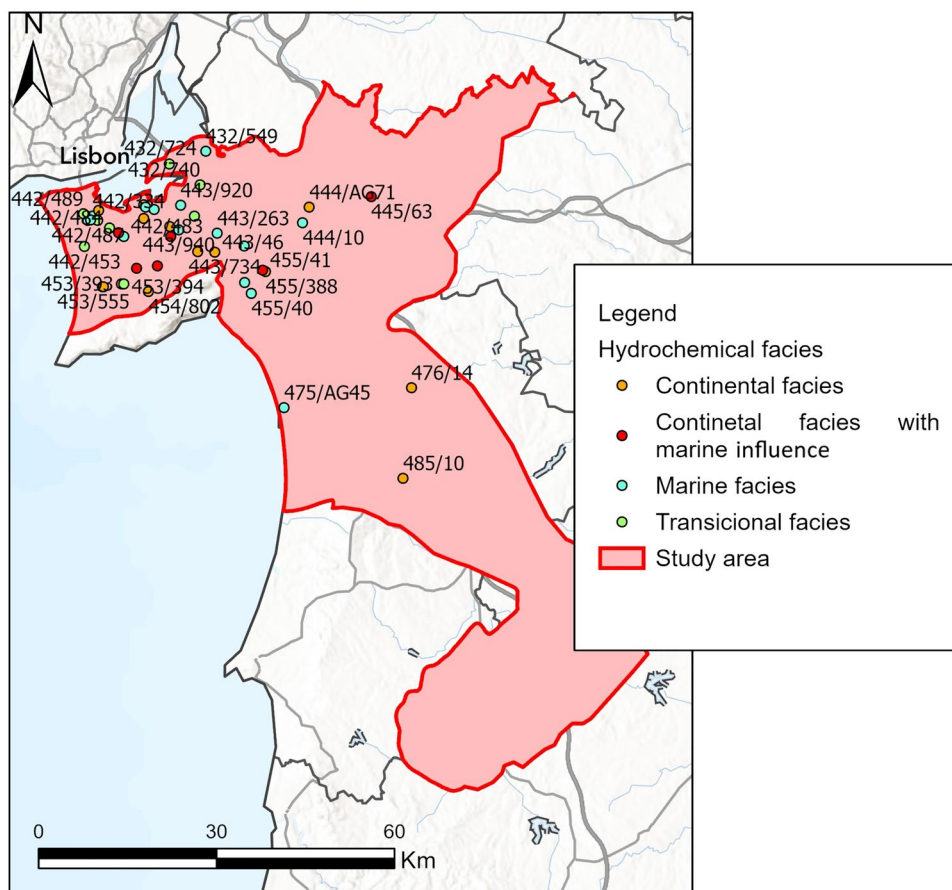
When ITA results are complemented with MKA, it is possible to conclude with a certain level of accuracy that, in the period 2006–2010, there was an initial increase of GDL in the center of the Basin, and from 2010 to 2015, the opposite occurred in the bay areas. In the South sector, there are significant variations of GDL in the last five years, in which the well 476/20 represents the highest increase in GDL. This change is undoubtedly visible in both ITA and MKA.

The higher increase tendencies in GDL can be related to insufficient recharge effects, caused by anthropic action, which is reflected by the intensification of human activity in the NE and SE sectors of the study area. In the North sector, industrialization, urban areas, and population growth may be inducing some of the detected localized increases in GDL.

In the SE sector, besides the influence of agricultural activities, the generalized increase in GDL suggests the probable influence of other natural factors such as more pronounced drought effects, which has been evidenced by recent investigation studies in the area (Novo et al. 2020), and by the evolution of temperature and precipitation according to Fig. 4. Another reason the increase in GDL is associated with the probability of some localized hydrodynamic changes due to the influence of Variscan faulting and hydrothermal systems present in the Paleozoic basement massif that underlies the studied aquifer systems (Oliveira et al. 1998, 2001; Matos 2008, 2021; Matos et al. 2009). The southern sector of the Tagus-Sado Basin is affected by Paleozoic structures corresponding to the Iberian Pyrite Belt where VMS deposits are present (e.g., Lagoa Salgada) (Matos 2021). Distinct Alpine age faults intersect the area, compartmentalizing the Paleozoic basement in distinct horst and graben structures and it is likely that some of these faults are in an active tectonic regime (Oliveira et al. 1998, 2001; Matos 2008, 2021; Matos et al. 2009).

It is also important to mention that in the SE sector of the basin, some areas are classified as high rock temperatures until a depth 5 of km (Anderson et al. 2011). Under this geological context, hydrodynamic conditions of the aquifer systems may suffer in specific locations with slight or well demarked alterations in groundwater dynamics and hydrochemical composition This information is corroborated by the analysis of hydrochemical facies from data ranging from 2000 to 2010 shown in Fig. 10, in which

**Fig.10** Hydrochemical facies map distribution



there is a progression of the tendency from the ocean to the continent in the NW sector, with the appearance of transition facies, and to the south the increase of continental facies.

Due to some missing values in some data ranges, a certain level of uncertainty is presented in some of the results, which can mislead the observation of accurate tendencies in some few wells. It is to say that this is an exploratory study. Considering the detected tendencies, and the extension and complexity of the geological and hydrogeological

conditions, it is advisable to increase the official ongoing monitoring plan, and future research works at regional scale.

### Appendix A

See Figs. [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#).

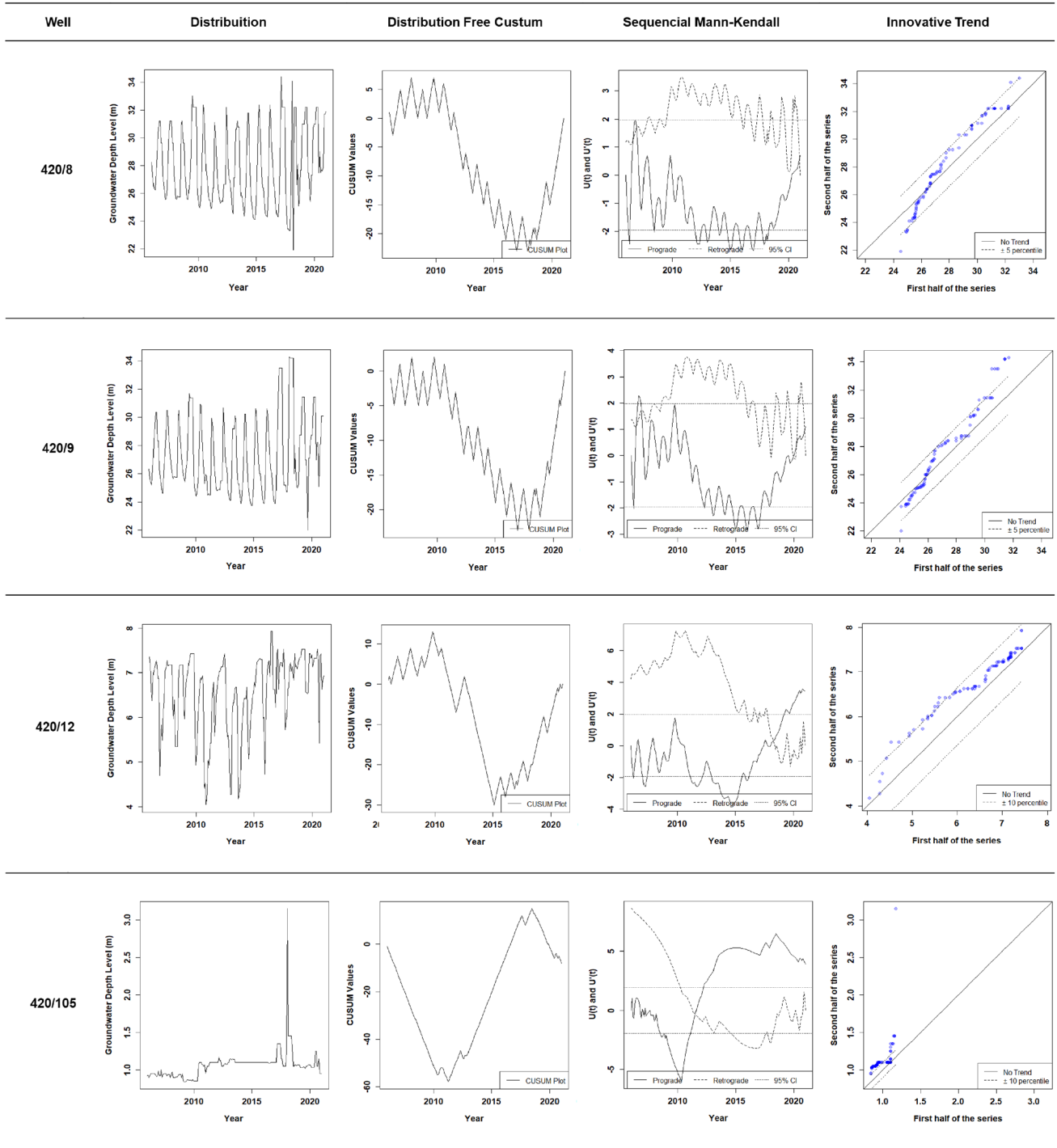


Fig. 11 MKA and ITA Graphical Results (part 1)

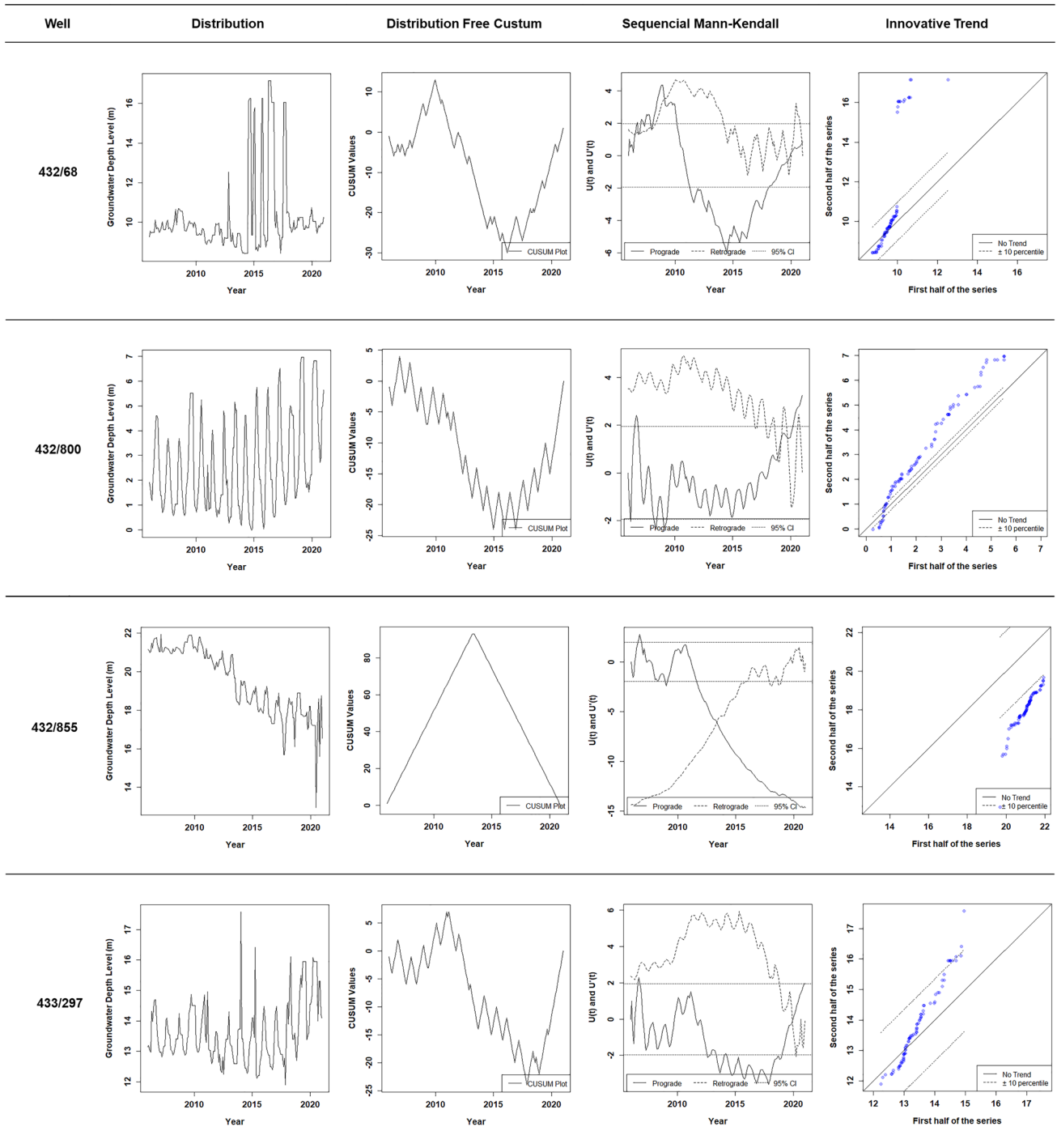


Fig. 12 MKA and ITA Graphical Results (part 2)

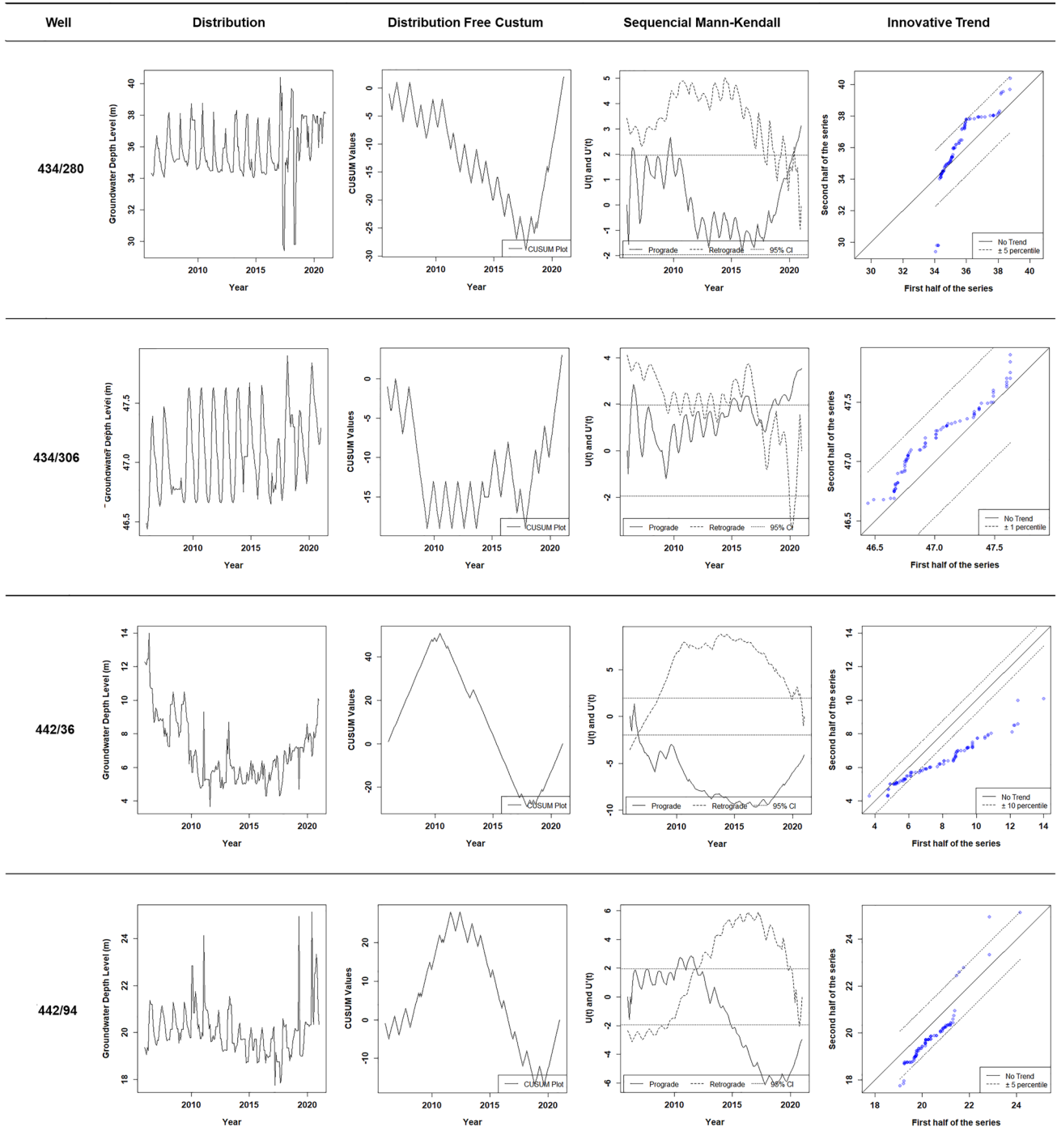


Fig. 13 MKA and ITA Graphical Results (part 3)

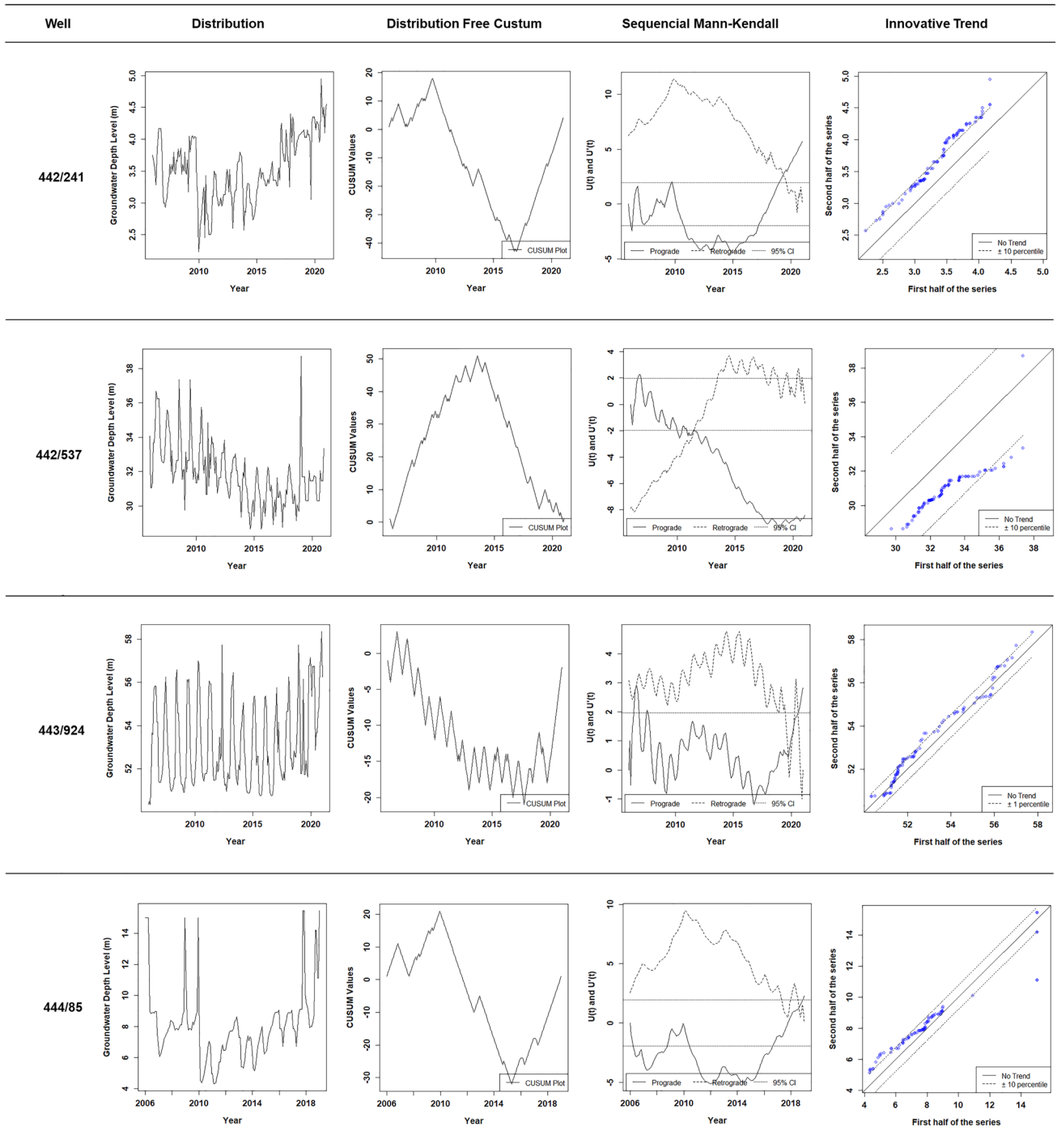


Fig. 14 MKA and ITA Graphical Results (part 4)

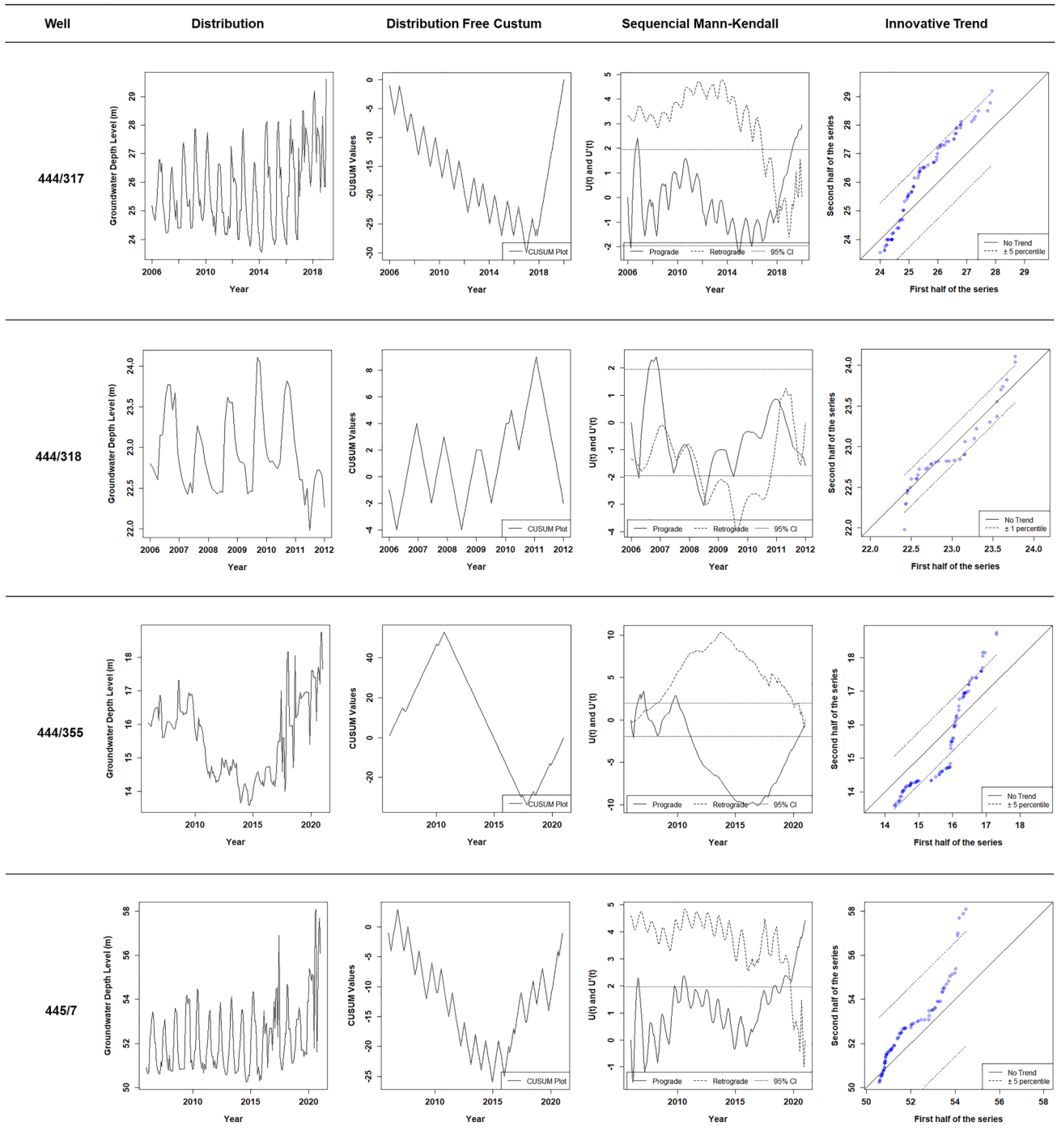


Fig. 15 MKA and ITA Graphical Results (part 5)



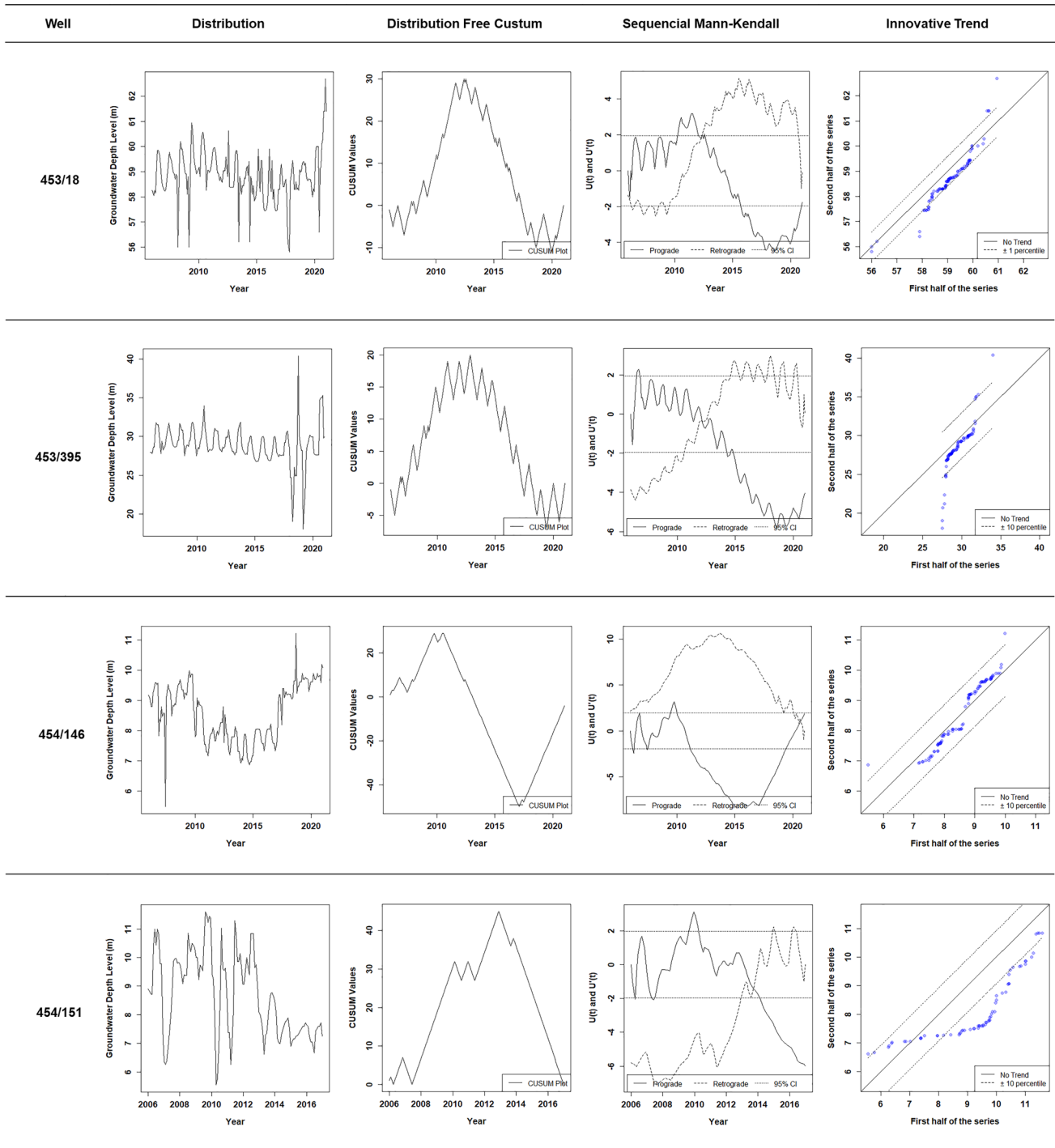


Fig. 16 MKA and ITA Graphical Results (part 6)

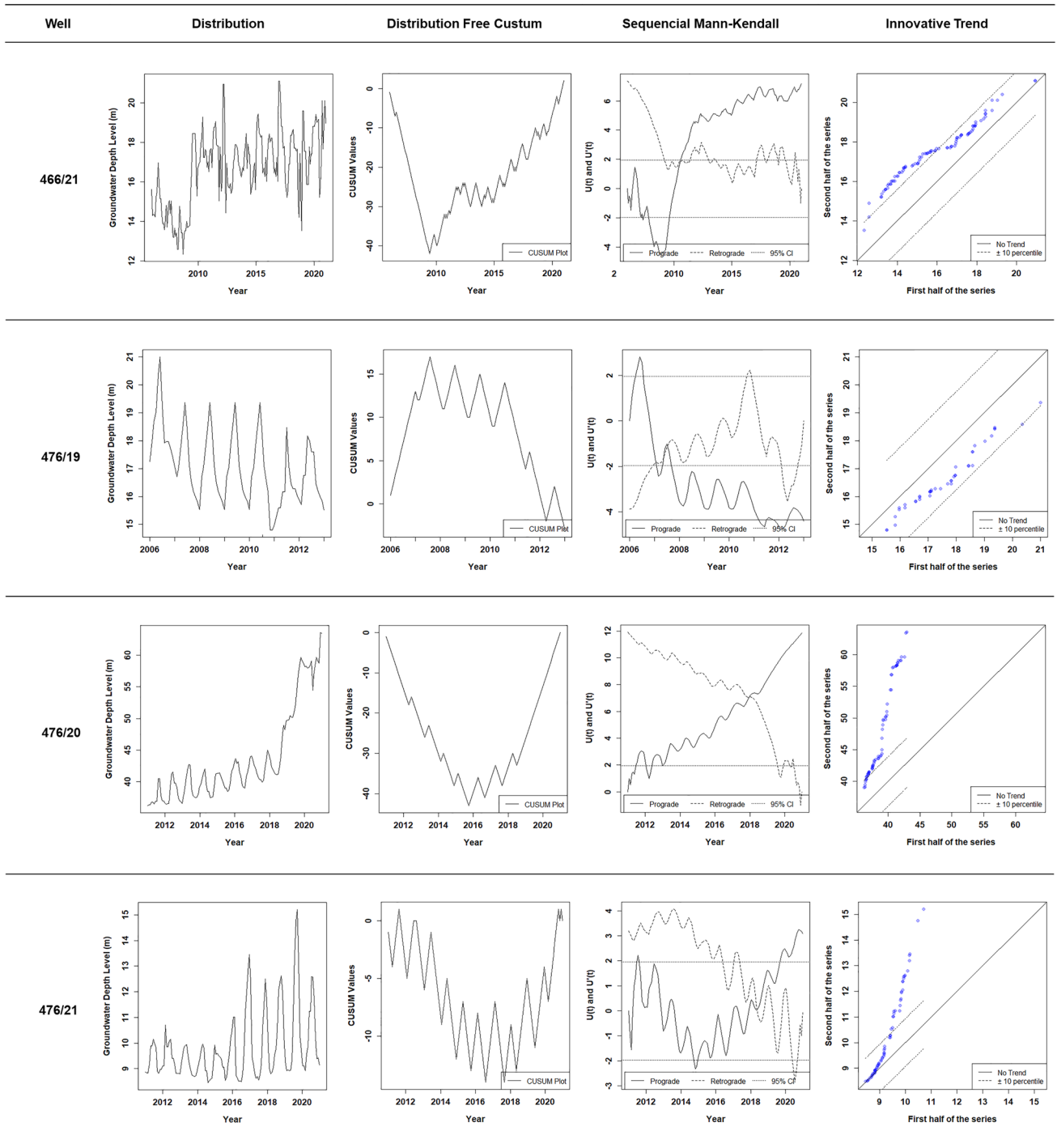


Fig. 17 MKA and ITA Graphical Results (part 7)

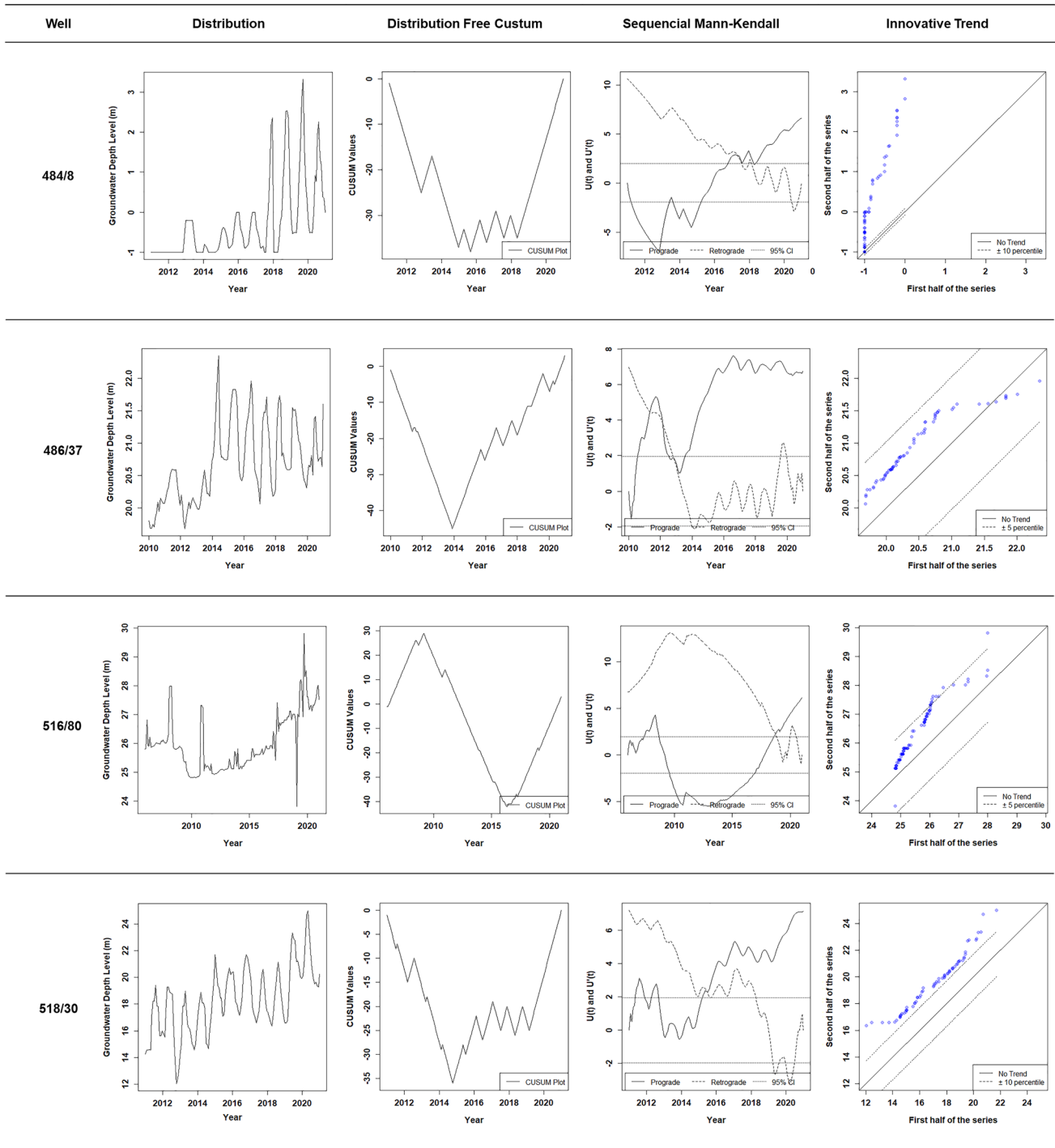


Fig. 18 MKA and ITA Graphical Results (part 8)

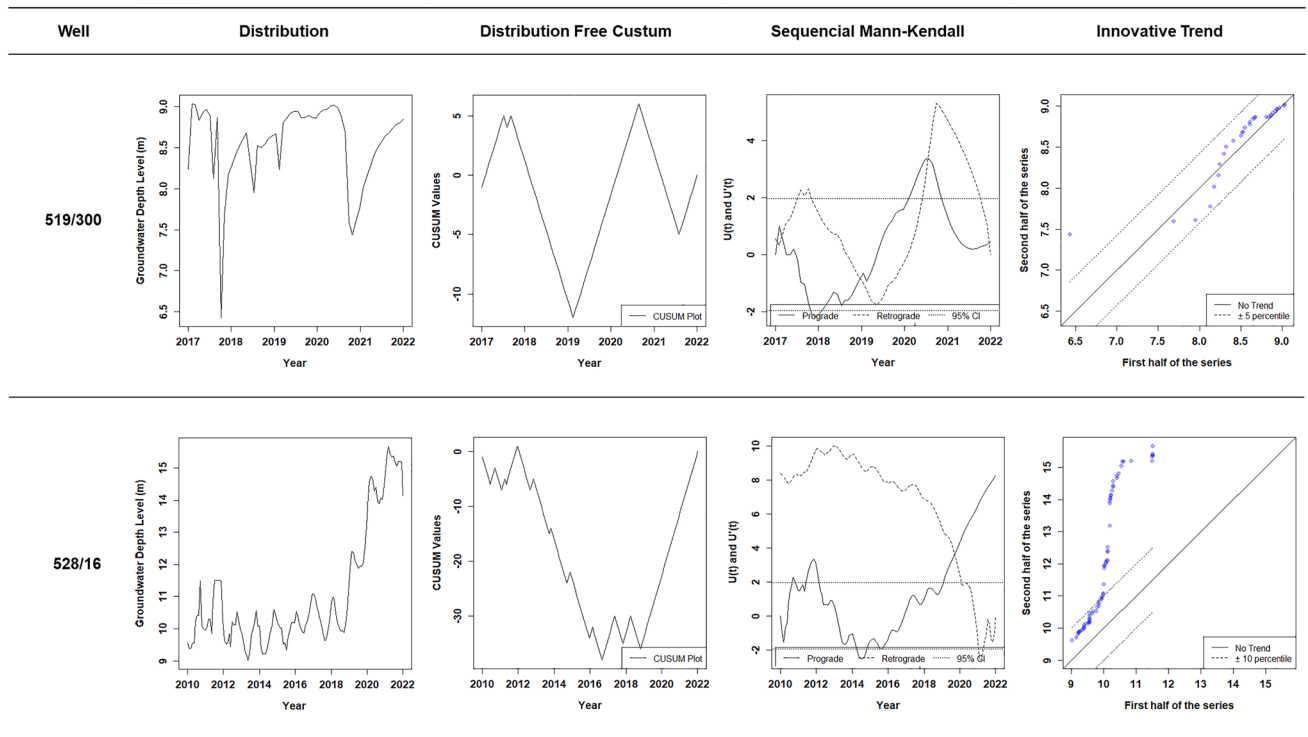


Fig. 19 MKA and ITA Graphical Results (part 9)

### Appendix B

See Figs. 20, 21.

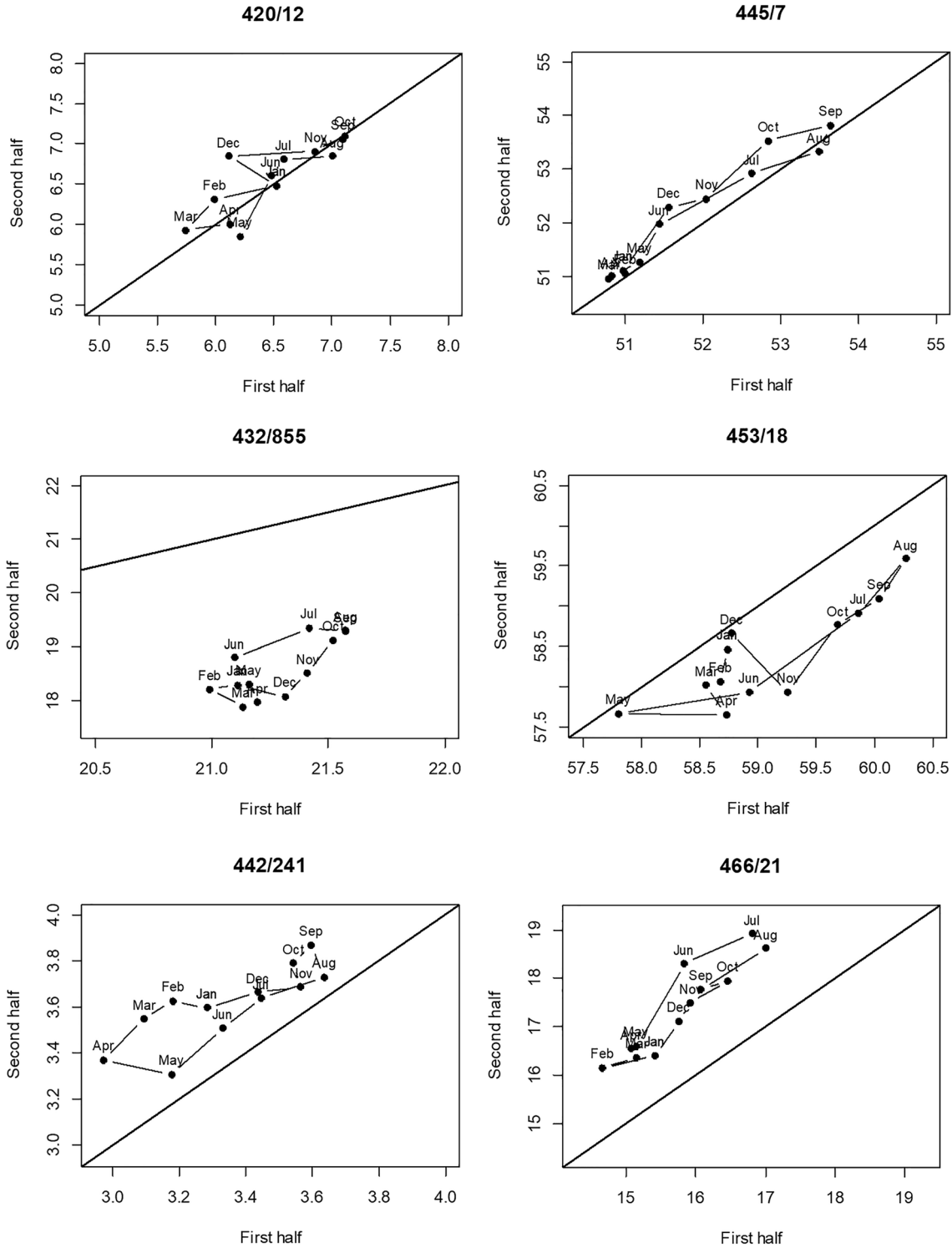


Fig. 20 IPTA Graphical Results (part 1)

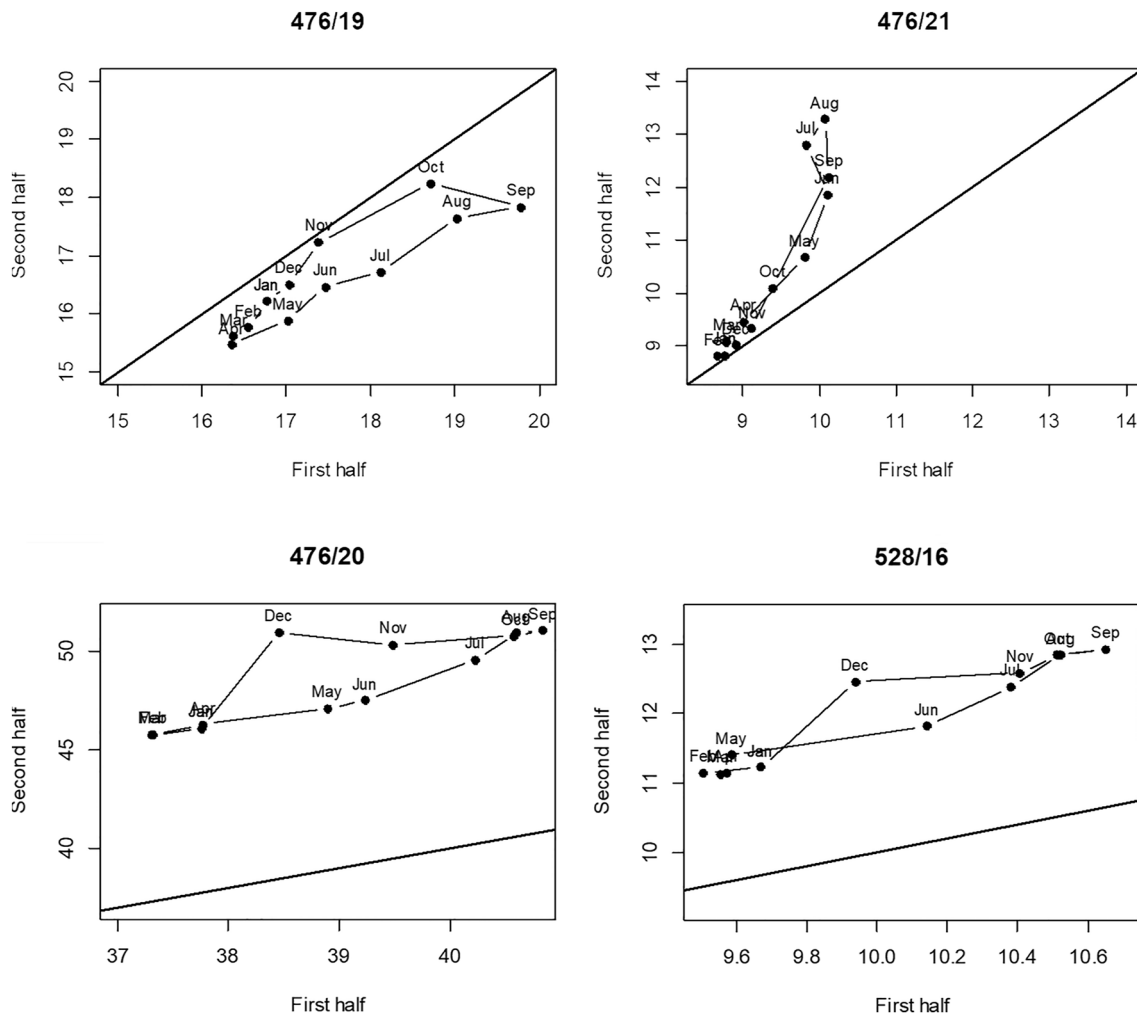


Fig. 21 IPTA Graphical Results (part 2)

**Acknowledgements** The authors acknowledge the support of the Department of Earth Sciences of FCT-NOVA, GeoBioTec, and Mineral Resources and Geophysics Unit of LNEG.

**Author contributions** Conceptualization, methodology, funding acquisition, S.B.; software, M.B.; investigation, S.B., M.B., and J.X.M.; resources, S.B. and J.X.M.; data curation, M.M.; validation, M.B., S.B., and J.X.M.; writing—original draft preparation, M.B. and S.B.; writing—review and editing, S.B. and J.X.M. All authors have read and agreed to the published version of the manuscript.

**Funding** Open access funding provided by FCTIFCCN (b-on). The research was partially developed for a master thesis, developed in the Earth Sciences Department of the Faculty of Sciences and Technology (FCT), NOVA University of Lisbon, Portugal. This research was partially funded by FCT-Fundação para a Ciência e a Tecnologia, Portugal, grant number UIDB/04035/2020.

**Data availability** Not applicable.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abdi H (2003) Factor Analysis of mixed data: methods and applications. In: Tashakkori A, Teddlie C (eds) Handbook of Mixed methods in social & behavioral research. Sage Publications, Thousand Oaks, CA, pp 401–425

- Achite M, Ceribasi G, Ceyhunlu AI, Wałęga A, Caloiero T (2021) The innovative polygon trend analysis (IPTA) as a simple qualitative method to detect changes in environment—example detecting trends of the total monthly precipitation in semiarid area. *Sustainability* 13(22):12674. <https://doi.org/10.3390/su13212674>
- Adamovic M, Branger F, Braud I, Kralisch S (2016) Development of a data-driven semi-distributed hydrological model for regional scale catchments prone to Mediterranean flash floods. *J Hydrol* 541:173–189. <https://doi.org/10.1016/j.jhydrol.2016.03.032>
- Ahmed N, Wang G, Booij MJ, Ceribasi G, Bhat MS, Ceyhunlu AI, Ahmed A (2022) Changes in monthly streamflow in the Hindu-kush–Karakoram–Himalaya region of Pakistan using innovative polygon trend analysis. *Stochastic Environ Res Risk Assess* 36(3):811–830. <https://doi.org/10.1007/S00477-021-02067-0/FIGURES/11>
- Akçay F, Bingölbali B, Akpınar A, Kankal M (2022) Trend detection by innovative polygon trend analysis for winds and waves. *Front Mar Sci*. <https://doi.org/10.3389/fmars.2022.930911>
- Al Naeem MF, Yusoff I, Ng TF, Maity JP, Alias Y, May R, Alborsh HA (2019) A study on the impact of anthropogenic and geogenic factors on groundwater salinization and seawater intrusion in Gaza coastal aquifer, Palestine: an integrated multi-techniques approach. *J Afr Earth Sci* 156:75–93. <https://doi.org/10.1016/j.jafrearsci.2019.05.006>
- Alifujiang Y, Abuduwaali J, Maihemuti B, Emin B, Groll M (2020) Innovative trend analysis of precipitation in the lake Issyk-Kul basin Kyrgyzstan. *Atmosphere* 11(4):332. <https://doi.org/10.3390/atmos11040332>
- Almazroui M, Şen Z (2020) Trend analyses methodologies in hydro-meteorological records. *Earth Syst Environ* 4(4):713–738. <https://doi.org/10.1007/s41748-020-00190-6>
- Almeida JS, Barbosa L, Pais AACC, Formosinho SJ (2007) Improving hierarchical cluster analysis: a new method with outlier detection and automatic clustering. *Chemom Intell Lab Syst* 87(2):208–217. <https://doi.org/10.1016/j.chemolab.2007.01.005>
- Almeida C, Mendonça JLL, Jesus MR, Gomes AJ (2000) Sistemas aquíferos de Portugal Continental. Centro de Geologia da Fac. Ciências Univ. Lisboa Instituto da Água, vol. II e III. [https://snirh.apambiente.pt/snirh/download/aquiferos\\_PortugalCont/Ficha\\_T3.pdf](https://snirh.apambiente.pt/snirh/download/aquiferos_PortugalCont/Ficha_T3.pdf)
- Anderson MJ (2008) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26(1):32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Anderson EW, Antkowiak M, Butt R, Davis J, Dean J, Hillesheim M, Hotchkiss E, Hunsberger R, Kandt A, Lund J, Massey K, Robichaud R, Stafford B, Visser C (2011) Broad overview of energy efficiency and renewable energy opportunities for department of defense installations. (Report No NREL/TP-7A20–50172). National Renewable Energy Laboratory. <https://doi.org/10.2172/1023698>
- Antunes MT (1983) Notícia Explicativa da Folha 39-C (Alcácer do Sal) da Carta Geológica de Portugal, na escala 1/50 000. Serviços Geológicas de Portugal, Lisboa, pp 39–47. [https://geoportal.ineg.pt/pt/dados\\_abertos/cartografia\\_geologica/cgp50k/](https://geoportal.ineg.pt/pt/dados_abertos/cartografia_geologica/cgp50k/)
- Audigier V, Husson F, Josse J (2016) A principal component method to impute missing values for mixed data. *Adv Data Anal Classif* 10:5–26. <https://doi.org/10.1007/S11634-014-0195-1/FIGURES/9>
- Barbosa S, Pinto M, Almeida JA, Carvalho E, Diamantino C (2021) Hydrochemical contamination profiling and spatial-temporal mapping with the support of multivariate and cluster statistical analysis. *World Acad Sci Eng Tech Intern J Geol Environ Eng* 15(9):243–252
- Batdelger O, Tsujimura M, Tran DA, Zorigt B, Thuc PTB (2023) Identification of hydrogeochemical processes and controlling factors in groundwater and surface water using integrated approaches, Tuul River Basin (Ulaanbaatar, Mongolia). *Environ science and Eng. Springer Nature, Cham*, pp 167–198. [https://doi.org/10.1007/978-3-031-17808-5\\_12](https://doi.org/10.1007/978-3-031-17808-5_12)
- Bayo J, López-Castellanos J (2016) Principal factor and hierarchical cluster analyses for the performance assessment of an urban wastewater treatment plant in the Southeast of Spain. *Chemosphere* 155:152–162. <https://doi.org/10.1016/j.chemosphere.2016.04.038>
- Buri ES, Keesara VR, Loukika NK (2022) Long-term trend analysis of observed gridded precipitation and temperature data over Munneru River basin, India. *J Earth Syst Sci*. <https://doi.org/10.1007/s12040-022-01864-7>
- Caloiero T (2019) Evaluation of rainfall trends in the South Island of New Zealand through the innovative trend analysis (ITA). *Theoret Appl Climatol* 139(1–2):493–504. <https://doi.org/10.1007/s00704-019-02988-5>
- Caloiero T, Coscarelli R, Ferrari E (2018) Application of the innovative trend analysis method for the trend analysis of rainfall anomalies in Southern Italy. *Water Resour Manag* 32(15):4971–4983. <https://doi.org/10.1007/S11269-018-2117-Z/TABLES/1>
- Celestino AEM, Cruz DP, Sánchez E, Reyes FV, Soto D (2018) Groundwater quality assessment: an improved approach to K-means clustering, principal component analysis and spatial analysis: a case study. *Water* 10(4):437. <https://doi.org/10.3390/w10040437>
- Cerón JC, Jiménez-Espinosa R, Pulido-Bosch A (2000) Numerical analysis of hydrogeochemical data: a case study (Alto Guadalentín, southeast Spain). *Appl Geochem*. [https://doi.org/10.1016/s0883-2927\(99\)00105-5](https://doi.org/10.1016/s0883-2927(99)00105-5)
- Correia J (2017) Contributo para a modelação 3D de horizontes geológicos com o auxílio de sísmica de reflexão e dados de sondagens e poços: um caso de estudo na Bacia Cenozóica do Baixo Tejo. Master dissertation, Nova School of Science and Technology. <http://hdl.handle.net/10362/22223>
- Costa EF (1994) Notícia explicativa das folhas 7 e 8 da Carta Hidrogeológica de Portugal, Instituto Geológico e Mineiro, Lisboa, pp 83. [https://geoportal.ineg.pt/pt/dados\\_abertos/cartografia\\_geologica/carta\\_hidrogeologica\\_200k](https://geoportal.ineg.pt/pt/dados_abertos/cartografia_geologica/carta_hidrogeologica_200k)
- Cunha, PP (1992) Estratigrafia e sedimentologia dos depósitos do cretácio superior e terciário de Portugal Central, a leste de Coimbra. Doctoral dissertation, University of Coimbra. <https://estudogeral.sib.uc.pt/handle/10316/2015>
- De Andrade CD, De Azevedo JA, Santos MAD, Assunção RDSFV (2020) Water quality assessment based on multivariate statistics and water quality index of a strategic river in the Brazilian Atlantic forest. *Sci Rep*. <https://doi.org/10.1038/s41598-020-78563-0>
- Dinpashoh Y, Mirabbasi R, Jhajharia D, Abianeh HZ, Mostafaeipour A (2014) Effect of short-term and long-term persistence on identification of temporal trends. *J Hydrol Eng* 19(3):617–625. [https://doi.org/10.1061/\(asce\)jhe.1943-5584.0000819](https://doi.org/10.1061/(asce)jhe.1943-5584.0000819)
- Fatahi NR, Yaghoobi P, Reaisi VH, Ostad-Ali-Askari K, Nouri J, Maghsoudlou B (2021) Eco-hydrologic stability zonation of dams and power plants using the combined models of SMCE and CEQUALW2. *Appl Water Sci*. <https://doi.org/10.1007/s13201-021-01427-z>
- Ferchichi H, Hamouda MB, Farhat B, Mammou AB (2018) Assessment of groundwater salinity using GIS and multivariate statistics in a coastal Mediterranean aquifer. *Int J Environ Sci Tech* 15(11):2473–2492. <https://doi.org/10.1007/s13762-018-1767-y>
- Fernandes GP, Silva OM (1998) Contribuição para a caracterização hidrogeológica do sistema aquífero da bacia do sado. In 4 ° Congresso da Água, A água como recurso estruturante do desenvolvimento, Centro de congressos da FIL, Lisboa. APRH. [https://www.aprh.pt/congressoagua98/files/c\\_titul.htm](https://www.aprh.pt/congressoagua98/files/c_titul.htm)
- Ghaemi Z, Noshadi M (2022) Surface water quality analysis using multivariate statistical techniques: a case study of Fars Province

- rivers Iran. *Environ Monit Assess.* <https://doi.org/10.1007/s10661-022-09811-1>
- Ghashghaie M, Eslami H, Ostad-Ali-Askari K (2022) Applications of time series analysis to investigate components of Madiyan-rood River water quality. *Appl Water Sci.* <https://doi.org/10.1007/s13201-022-01693-5>
- Granato D, Santos JS, Escher GB, Ferreira BL, Maggio RM (2018) Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective. *Trends Food Sci Technol* 72:83–90. <https://doi.org/10.1016/j.tifs.2017.12.006>
- Güçlü YS (2020) Improved visualization for trend analysis by comparing with classical Mann-Kendall test and ITA. *J Hydrol* 584:124674. <https://doi.org/10.1016/j.jhydrol.2020.124674>
- Hamed KH (2008) Trend detection in hydrologic data: the Mann-Kendall trend test under the scaling hypothesis. *J Hydrol* 349(3–4):350–363. <https://doi.org/10.1016/j.jhydrol.2007.11.009>
- Harkat S, Kisi O (2021) Trend analysis of precipitation records using an innovative trend methodology in a semi-arid Mediterranean environment: Chelif watershed case (Northern Algeria). *Theoretical Appl Climatol* 144(3–4):1001–1015. <https://doi.org/10.1007/S00704-021-03520-4/FIGURES/10>
- Helsel DR (1987) Advantages of nonparametric procedures for analysis of water quality data. *Hydrol Sci J* 32(2):179–190. <https://doi.org/10.1080/02626668709491176>
- Husson F, Lê S, Pagès J (2017) Exploratory multivariate analysis by example using R. Chapman and Hall/CRC eBooks. <https://doi.org/10.1201/b21874>
- Husson F, Josse J, Lê S, Mazet J (2020) Package ‘FactoMineR’: multivariate exploratory data analysis and data mining. <https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>
- Javadinejad S, Hannah D, Ostad-Ali-Askari K, Krause S, Zalewski M, Boogaard F (2019a) The impact of future climate change and human activities on hydro-climatological drought, analysis and projections: using CMIP5 climate model simulations. *Water Conserv Sci Eng* 4:71–88. <https://doi.org/10.1007/s41101-019-00069-2>
- Javadinejad S, Ostad-Ali-Askari K, Eslamian S (2019b) Application of multi-index decision analysis to management scenarios considering climate change prediction in the Zayandeh Rud river basin. *Water Conserv Sci Eng* 4(1):53–70. <https://doi.org/10.1007/S41101-019-00068-3/TABLES/4>
- Jiang Y, Guo H, Jia Y, Cao Y, Hu C (2015) Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin. *Inner Mongolia Chem Der Erde* 75(2):197–205. <https://doi.org/10.1016/j.chemer.2014.12.002>
- Kantiranis N, Mattas C, Pavlou A, Patrikaki O, Voudouris K (2017) Multivariate statistical analysis for the assessment of groundwater quality under different hydrogeological regimes. *Environ Earth Sci.* <https://doi.org/10.1007/s12665-017-6665-y>
- Krishnan RA, Ansari J, Sundararajan M, John C, Saharuba PM (2019) Chapter 9 – groundwater quality assessment using multivariate statistical methods for chavara aquifer system, Kerala, India. In: Venkatramanan S, Prasanna MV, Chung SY (eds) GIS and geostatistical techniques for groundwater science. Elsevier, pp 113–131. <https://doi.org/10.1016/B978-0-12-815413-7.00009-2>
- Kullberg MC, Kullberg JC, Terrinha P (2000) Tectónica da Cadeia da Arrábida. Tectónica das regiões de Sintra e Arrábida. *Memórias Geociências Do Museu Nacional História Natural Da Universidade De Lisboa* 2:35–84
- Kuriqi A, Ali R, Pham QB, Gambini JM, Gupta V, Malik A, Linh NTT, Joshi Y, Anh DT, Nam VT, Dong X (2020) Seasonality shift and streamflow flow variability trends in central India. *Acta Geophys* 68:1461–1475. <https://doi.org/10.1007/s11600-020-00475-4>
- Lê S, Josse J, Rennes A, Husson F (2008) FactoMineR: an R package for multivariate analysis. *JSS J Statist Softw* 25:1–18
- Love DJ, Hallbauer D, Amos A, Hranova R (2004) Factor analysis as a tool in groundwater quality management: two southern African case studies. *Phys Chem Earth Parts a/b/c* 29(15–18):1135–1143. <https://doi.org/10.1016/j.pce.2004.09.027>
- Machiwal D, Singh PK (2015) Understanding factors influencing groundwater levels in hard-rock aquifer systems by using multivariate statistical techniques. *Environ Earth Sci* 74(7):5639–5652. <https://doi.org/10.1007/s12665-015-4578-1>
- Mann HB (1945) Nonparametric tests against trend. *Econ Soc* 13(3):245–259. <https://doi.org/10.2307/1907187>
- Manuppella G (Coord), Antunes MT, Pais J, Cardoso JL, Ramalho M, Rey J (1999) Notícia Explicativa da Folha 39-B (Setúbal) da Carta Geológica de Portugal, na escala 1/50 000. Serviços Geológicos de Portugal, Lisbon, pp 143. [https://geoportal.lneg.pt/pt/dados\\_abertos/cartografia\\_geologica/cgp50k/](https://geoportal.lneg.pt/pt/dados_abertos/cartografia_geologica/cgp50k/)
- Matos JX, Sousa P, Ricardo J (2009) MAEPA - Área de Marateca. Caracterização Geológica, Geofísica e Geoquímica da Região de Palma-Serrinha-Cordeira, Definição de Alvos de Sondagem. Technical Report, DPMM INETI pp. 66.
- Matos JX (2008) MAEPA - Área de Marateca. Caracterização Geológica e Mineira Preliminar das Janelas Paleozóicas de Vela, Serra do Loureiro, Palma, Serrinha, Clérigos e Cordeira. Technical Report DPMM LNEG/INETI, pp. 17
- Matos JX (2021) Alteração hidrotermal ácido-sulfato associada aos jazigos de sulfuretos maciços de Lagoa Salgada, Caveira, Lousal, Aljustrel e São Domingos (Faixa Piritosa Ibérica). PhD Thesis. In Geology, specialization metallogeny, geology department, science faculty University of Lisbon, Lisbon pp 435.
- Minea I, Boicu D, Chelariu OE (2020) Detection of groundwater levels trends using innovative trend analysis method in temperate climatic conditions. *Water* 12(8):2129. <https://doi.org/10.3390/w12082129>
- Mirabbasi R, Ahmadi F, Jhahharia D (2020) Comparison of parametric and non-parametric methods for trend identification in groundwater levels in Sirjan plain aquifer. *Iran Hydrol Res* 51(6):1455–1477. <https://doi.org/10.2166/nh.2020.041>
- Mumby PJ (2002) Statistical power of non-parametric tests: a quick guide for designing sampling strategies. *Mar Pollut Bull* 44(1):85–87. [https://doi.org/10.1016/s0025-326x\(01\)00097-2](https://doi.org/10.1016/s0025-326x(01)00097-2)
- Nielsen F (2016) Hierarchical clustering. In: Nielsen F (ed) Introduction to HPC with MPI for data science. International Publishing, Cham, pp 195–211. [https://doi.org/10.1007/978-3-319-21903-5\\_8](https://doi.org/10.1007/978-3-319-21903-5_8)
- NLJ (1948) Rank correlation methods. By Maurice G. Kendall, M. [Pp. vii + 160. London: Charles Griffin and Co. Ltd., 42 Drury Lane, 1948. 18s.]. *J Inst Actuar* 75(1):140–141. <https://doi.org/10.1017/s0020268100013019>
- Novo ME, Martins T, Henriques JM (2020) BINGO project: Impacts of climate change in groundwater in the lower Tagus - Coupling outputs from climate and re-charge models with aquifer modeling. Report No 245/2020 – DHA/NRE. Laboratório Nacional de Engenharia Civil. <http://repositorio.lnec.pt:8080/xmlui/handle/123456789/1012984>
- Oh J, Kim H, Yu S, Kim K, Yun ST (2020) Delineating the impacts of poultry burial leachate on shallow groundwater in a reclaimed agro-livestock farming area, using multivariate statistical analysis of hydrochemical data. *Environ Sci Pollut Res* 28(7):7742–7755. <https://doi.org/10.1007/s11356-020-08178-5>
- Oliveira V, Matos JX, Bengala M, Sousa P (1998) Principais alinhamentos vulcânicos a norte da Falha de Grândola, sob formações da Bacia Terciária do Sado e sua potencialidade mineira no contexto da Faixa Piritosa Ibérica. In *Actas V Congresso Nacional de Geologia Com. IGMT.* 84 F. 2, pp.F15–18.



- Oliveira V, Matos JX, Rosa C (2001) The NNW sector of the Iberian pyrite belt — new exploration perspectives for the next decade. In Geode workshop — massive sulphide deposits in the Iberian pyrite belt: new advances and comparison with equivalent systems, pp.34–35, Aracena Spain.
- Pages J (2004) Analyse factorielle de donnes mixtes. *Revue de Statist Appliq* 52(4) :93–111. [http://www.numdam.org/item?id=RSA\\_2004\\_\\_52\\_4\\_93\\_0](http://www.numdam.org/item?id=RSA_2004__52_4_93_0)
- Pais J, Moniz C, Cabral J, Cardoso JL, Legoinha P, Machado S, Morais MA, Loureno C, Ribeiro ML, Henriques P, Fal P (2006) Notcia Explicativa da Folha 34-D (Lisboa) da Carta Geolgica de Portugal, na escala 1/50 000. Departamento de Geologia, INETI - Instituto Nacional de Engenharia, Tecnologia e Inovao., Lisbon, pp. 74. [https://geoportal.ineg.pt/pt/dados\\_abertos/cartografia\\_geologica/cgp50k/](https://geoportal.ineg.pt/pt/dados_abertos/cartografia_geologica/cgp50k/)
- Panda B, Radha VD, Chidambaram S, Arindam M, Thilagavathi R, Manikandan S, Thivya C, Ramanathan AL, Ganesh N (2019) Chapter 22 - Fluoride contamination in groundwater—a GIS and geostatistics reappraisal. In: Venkatramanan S, Prasanna MV, Chung SY (eds) GIS and geostatistical techniques for groundwater science. Elsevier, Cham, pp 309–322
- Prtner H, Roberts D, Tignor M, Poloczanska E, Mintenbec K, Alegria A, Craig M, Langsdorf S, Lschke S, Mller V, Okem A, Rama B, Belling D, Dieck W, Gtze S, Kersher T, Mangele P, Maus B, Mhle A, Weyer N (2022) Climate change 2022: impacts, adaptation and vulnerability working group II contribution to the sixth assessment report of the intergovernmental panel on climate change. Cambridge, UK and New York, NY, USA
- Rabiei J, Sadat KM, Bagherpour S, Ebadi N, Karimi A, Ostad-Ali-Askari K (2022) Investigation of fire risk zones using heat-humidity time series data and vegetation. *Appl Water Science*. <https://doi.org/10.1007/s13201-022-01742-z>
- Rahbar A, Vadiati M, Talkhabi M, Nadiri AA, Nakhaei M, Rahimian M (2020) A hydrogeochemical analysis of groundwater using hierarchical clustering analysis and fuzzy C-mean clustering methods in Arak plain, Iran. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-020-09064-6>
- Rao NK, Chaudhary M (2019) Hydrogeochemical processes regulating the spatial distribution of groundwater contamination, using pollution index of groundwater (PIG) and hierarchical cluster analysis (HCA): a case study. *Groundw Sustain Dev* 9:100238. <https://doi.org/10.1016/j.gsd.2019.100238>
- Riaz M, Tahir M, Arslan M (2016) Non-parametric versus parametric methods in environmental sciences. *Bull Environ Stud* 1(1):28–30
- Ruiz F, Gomis V, Blasco P (1990) Application of factor analysis to the hydrogeochemical study of a coastal aquifer. *J Hydrol* 119(1–4):169–177. [https://doi.org/10.1016/0022-1694\(90\)90041-u](https://doi.org/10.1016/0022-1694(90)90041-u)
- an M, Akay F, Linh NTT, Kankal M, Pham QB (2021) Innovative and polygonal trend analyses applications for rainfall data in Vietnam. *Theoretical Appl Climatol* 144(3–4):809–822. <https://doi.org/10.1007/S00704-021-03574-4/FIGURES/6>
- Sanikhani H, Kisi O, Mirabbasi R, Meshram SG (2018) Trend analysis of rainfall pattern over the Central India during 1901–2010. *Arabian J Geosci* 11(15):1–14. <https://doi.org/10.1007/S12517-018-3800-3/TABLES/8>
- en Z (2012) Innovative trend analysis methodology. *J Hydrol Eng* 17(9):1042–1046. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000556](https://doi.org/10.1061/(asce)he.1943-5584.0000556)
- en Z (2014) Trend Identification simulation and application. *J Hydrol Eng* 19(3):635–642. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000811](https://doi.org/10.1061/(asce)he.1943-5584.0000811)
- en Z (2017) Innovative trend significance test and applications. *Theoretical Appl Climatol* 127(3–4):939–947. <https://doi.org/10.1007/s00704-015-1681-x>
- en Z, iman E, Dabanli I (2019) Innovative polygon trend analysis (IPTA) and applications. *J Hydrol* 575:202–210. <https://doi.org/10.1016/J.JHYDROL.2019.05.028>
- Seth N. (2022). What is Principal Component Analysis (PCA) vs. Factor Analysis?. In: Blogs & Updates on Data Science, Business Analytics, AI Machine Learning. <https://www.analytixlabs.co.in/blog/factor-analysis-vs-pca/>. Accessed 21 May 2022
- Shiker MAK (2012) Multivariate statistical analysis public policies-new smart settings in public management view project transportation problems view project British. *J Sci* 6(1):55–77
- Simes M, Legoinha P (2014) Unidades hidrostratigrficas no Miocnico da Bacia do Tejo na regio de Almada. *Comunicaes Geolgicas* 101(Especial II):717–720
- Simes MM (1998) Contribuio para o Conhecimento Hidrogeolgico do Cenozico na Bacia do Baixo Tejo. PhD Thesis, FCT- NOVA Faculdade de Cincias e Tecnologia da Universidade Nova de Lisboa, Caparica. <http://hdl.handle.net/10362/1152>
- Smoliski A, Walczak B, Einax JW (2002) Hierarchical clustering extended with visual complements of environmental data set. *Chemom Intell Lab Syst* 64(1):45–54. [https://doi.org/10.1016/S0169-7439\(02\)00049-7](https://doi.org/10.1016/S0169-7439(02)00049-7)
- Soltani L, Mellah T (2023) Exploring farmers' adaptation strategies to water shortage under climate change in the Tunisian semi-arid region. *Environ Manag* 71(1):74–86. <https://doi.org/10.1007/s00267-022-01604-z>
- Subba Rao N, Srihari C, Deepthi Spandana B, Sraavathi M, Kamallesh T, Abraham Jayadeep V (2019) Comprehensive understanding of groundwater quality and hydrogeochemistry for the sustainable development of suburban area of Visakhapatnam, Andhra Pradesh, India. *Hum Ecol Risk Assess: Int J* 25(1–2):52–80. <https://doi.org/10.1080/10807039.2019.1571403>
- Swain S, Sahoo S, Taloor AK, Mishra S, Pandey A (2022) Exploring recent groundwater level changes using innovative trend analysis (ITA) technique over three districts of Jharkhand, India. *Groundw Sustain Dev* 18:100783. <https://doi.org/10.1016/j.gsd.2022.100783>
- Umar S, Lone MA, Goel N, Zakwan M (2022) Trend analysis of hydro-meteorological parameters in the Jhelum River basin, North Western Himalayas. *Theoretical Appl Climatol* 148(3–4):1417–1428. <https://doi.org/10.1007/s00704-022-04014-7>
- Varol M (2020) Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: a case study. *Environ Pollut* 266:115417. <https://doi.org/10.1016/j.envpol.2020.115417>
- Visbal-Cadavid D, Mendoza AM, De La Hoz-Dominguez EJ (2020) Use of factorial analysis of mixed data (FAMD) and hierarchical cluster analysis on principal component (HCPC) for multivariate analysis of academic performance of industrial engineering programs. *Xinan Jiaotong Daxue Xuebao*. <https://doi.org/10.35741/issn.0258-7274.55.5.34>
- Wali SU, Alias N, Harun S, Umar K, Gada MA, Dankani IM, Kaoje IU, Usman AA (2022) Water quality indices and multivariate statistical analysis of urban groundwater in semi-arid Sokoto Basin, Northwestern Nigeria. *Groundw Sustain Dev* 18:100779. <https://doi.org/10.1016/j.gsd.2022.100779>
- Wang H, Jiang X, Wan L, Han G, Guo H (2015) Hydrogeochemical characterization of groundwater flow systems in the discharge area of a river basin. *J Hydrol* 527:433–441. <https://doi.org/10.1016/j.jhydrol.2015.04.063>
- Wang Y, Zhu GY, Yu R (2018) Assessment of surface water quality using multivariate statistical techniques: a case study in China. *Irrigat Drainage Syst Eng*. <https://doi.org/10.4172/2168-9768.1000214>

- Zakwan M (2021) Trend analysis of groundwater level using innovative trend analysis. Springer eBooks, pp 389–405
- Zbyszewski G, Antunes MT, Ferreira ODV (1976) Notícia Explicativa da Folha 39-A (Águas de Moura) da Carta Geológica de Portugal, na escala 1/50 000. Serviços Geológicos de Portugal, Lisbon, pp.

59. [https://geoportal.lneg.pt/pt/dados\\_abertos/cartografia\\_geologica/cgp50k/](https://geoportal.lneg.pt/pt/dados_abertos/cartografia_geologica/cgp50k/)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.