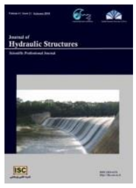
Journal of Hydraulic Structures
J. Hydraul. Struct., 2019; 6(1): 90- 104
DOI: 10.22055/jhs.2020.32747.1133

OPEN ACCESS



The Impact of outlier detection to estimate groundwater fluctuations using GRACE satellite data; Case Study: Khuzestan Province, Iran

Mohammad Mehdi Riyahi[1*]
Mohammad Jafarpour[1]
Lotfollah Emadali[2]
Mohammad Ali Sharifi[3]

Abstract

Groundwater aquifers are one of the most significant freshwater resources in the world. Hence, the monitoring of these resources is particularly important for available water resources planning. Piezometric wells have traditionally been used to monitor groundwater. This approach is costly and pointwise, which is not feasible for places with steep topography and mountainous areas. Nowadays, remote sensing techniques are widely used in various fields of engineering as appropriate alternatives to traditional methods. In water resources management, the Gravity Recovery and Climate Experiment (GRACE) satellites can monitor groundwater changes with acceptable accuracies. This paper applied the GRACE satellite data for a 40-month period to assess the variation of the groundwater level in Khuzestan province. The Global Land Data Assimilation System (GLDAS) model was used to counteract the soil moisture effect in final results. The observed data from piezometric wells were pre-processed to detect outliers using the Mahalanobis algorithm in Khuzestan province. At last, the outputs of GRACE were compared with these processed observed data. Despite the relatively small size of the area in question, the results indicated the efficiency of GRACE data (RMSE = 0.8, NRMSE = 0.2) for monitoring the groundwater level changes.

Keywords: Groundwater Monitoring, GRACE data, GLDAS Model, Outlier Detection, Mahalanobis Algorithm.

Received: 29 February 2020; Accepted: 17 April 2020

---

[1] Department of Civil Engineering, Civil Engineering and Architecture Faculty, Shahid Chamran University of Ahvaz, Ahvaz, Iran, 61357-83151(**Corresponding Author**,E-mail:mo_riyahi@yahoo.com*; E-mail: m.jafarpour13@gmail.com)
[2] Department of Civil Engineering, Engineering Faculty, Khatam Al-Anbia University of Technology, Behbahan, Iran, 63616-47189 (E-mail: bemad1390@alumni.ut.ac.ir)
[3] Department of Geodesy, Faculty of Surveying and Spatial Information, College of Engineering, University of Tehran, Tehran, Iran, 14155-6619 (E-mail: sharifi@ut.ac.ir)

1. Introduction

Most of the Earth's surface is covered by water and ice; only about 3 percent of the available water is fresh to supply human needs for drinking, agriculture, and industry, as well as essential for living creatures. In recent years, freshwater resources have been declined, which is more drastic in areas facing drought and natural hazards, including arid and semi-arid regions. Iran is located in southwest of Asia with hot and semi-arid climate which is subjected to the water shortage. Since a long time ago, due to the lack of surface water, groundwater has been traditionally used for water supply such as the Qanat system in central regions like Yazd.

A comprehensive study on groundwater resources is very important because it's consumed by more than half of the population around the world. Unfortunately, in recent years, due to the lack of rainfall and excessive extraction of groundwater, freshwater resources in the world have been declined and thus, the scarcity of groundwater has become a major challenge. The depletion of groundwater resources may have many adverse effects, such as lowering groundwater level in aquifers, desertification expansion and drying out of farms and gardens, decreasing in water discharge of Qanat and spring, drying up of rivers, lakes and wetlands, reduction of groundwater quality, soil salinity, environmental degradation, drying of plains and aggravation of land erosion and expansion of fields for producing the haze, land subsidence and disappearance of water storage reservoirs and natural water monitoring, and damage of the agricultural lands, infrastructures, roads, and etc., increasing costs and reducing agricultural production, which lead to widespread poverty and unemployment, economic, social, and political problems as well as migration of farmers to the cities.

An obvious example of this crisis is the occurrence of numerous and massive pits and land subsidence in various parts of the world as well as many plains of Iran such as Kaboudar Ahang plain of Hamadan and Varamin plain in the southeast of Tehran [1].

Monitoring of the groundwater fluctuations is essential for the investigation of variation in the aquifer water level. Since a long time ago, piezometric wells have been locally used to measure and estimate the groundwater level changes at specific and discrete points; however, it may be difficult and costly to construct, maintain and access the wells, as well as to observe accurately the wells, Also, data acquisition of the wells can be associated with multiple error sources. Hence, Since the 1970s, regarding the wide area of data coverage, cheapness and availability of data, remote sensing methods have been developed to investigate, predict and monitor hydrological phenomena like changes in groundwater levels. In this context, The Gravity Recovery and Climate Experiment (GRACE) project is the first satellite project that can be directly used to determine the groundwater changes in any topographical conditions[2]. The GRACE is a joint mission of National Aeronautics and Space Administration (NASA) and the Deutsches Zentrum für Luft-und Raumfahrt (DLR) which was designed to measure the Earth's gravity field anomalies and launched from Plesetsk station, Russia, on 17 March 2002 [3]. Having the appropriate spatial and temporal coverage and acceptable accuracy, GRACE along with other hydrological models can be used as an alternative tool to study changes in groundwater resources.

Wahr et al. [4] used data from GRACE Satellite to study the variations of water storages in three watersheds including Mississippi, Amazon, and Bay of Bengal [4]. Ramillien et al. [5] applied the GRACE satellite data to investigate groundwater level changes in the Amazon River basin. Their results indicated that filtering of the GRACE data can be improved the final results [5]. Awange et al. [6] examined the fluctuations of Lake Victoria in Africa as the second-largest freshwater lake in the world. They used a 45-months period of GRACE data from April 2001 to

April 2006. The results showed the lake dwindle at a rate of 1.8 km^3 per year [6].

Longuevergne et al. [7] examined changes in groundwater levels for a water basin using the GRACE data from 2003 to 2007. In this study, the Gaussian filter with a radius of 300 km was used to improve GRACE results [7]. Voss et al. [8] studied the trend of groundwater fluctuations in the north-central Middle East using the GRACE data. The results demonstrated the decreasing rate of the water storage during 2003-2009 [8]. Ferreira et al. [9] studied the groundwater changes between 2005 and 2010 in the Volta River Basin in West Africa. In their research, both GRACE and precipitation data were utilized [9]. Ahmed et al. [10] used 10-year GRACE data along with other remote sensing data to measure the spatiotemporal variations of the groundwater resources in the continent of Africa between 2003 and 2012. The results indicated that a large portion of the Africa Continent has significant variations from 44 mm/year to 15 mm / year, which is consistent with previous findings [11].

Rodell et al. used the GRACE observations to recognize 34 trends in changes of groundwater storage between 2002 and 2016 for classifying based on natural internal variations, unstable groundwater consumptions, and climate changes [12].

There has been no comprehensive research on the practical use of GRACE data for estimating groundwater level changes in limited areas of Iran; in addition, according to the importance of groundwater in Khuzestan province with very hot and dry climate, this study aims to investigate the capability of GRACE data to estimate groundwater level changes in Khuzestan province. In this study, first, the piezometric observations have been preprocessed to detect outliers; then, the outliers were replaced using the spline function fitted to the data. On the other hand, the changes in level of underground water have been evaluated using 40 months of GRACE data and applying Gaussian filter. Afterward, the Global Land Data Assimilation System (GLDAS) model has been employed to compute soil moisture for correcting the GRACE results. Finally, the comparison of the GRACE model with piezometric observation wells has been used to verify the outputs.

2. Materials and methods

2.1. GRACE Satellite

The GRACE project consists of two identical satellites orbiting at an altitude of about 500 kilometers from the Earth's surface. the mission uses a microwave ranging system named KBR (K-Band Ranging System) to accurately measure changes in the speed and distance between two identical spacecraft flying in a polar orbit about 220 kilometers (140 mi) apart. The ranging system is sensitive enough to detect separation changes as small as 10 micrometers (approximately one-tenth the width of a human hair) over a distance of 220 kilometers. In fact, KBR is a key observation in the GRACE mission, referred as the High Accuracy Inter-Satellite Ranging System (HAIRS). As the twin GRACE satellites circle the globe 15 times a day, they sense slight variations in Earth's gravitational field.

The primary mission of these satellites was to obtain temporal gravity anomalies of the Earth over a five-year period [13]; nevertheless, in June 2010, NASA and DLR decided to extend the GRACE mission to 2017 [14]. While GRACE reached the end of its science mission, the next-generation of this satellite, namely Gravity Recovery and Climate Experiment Follow-On (GRACE-FO) was launched by NASA in 22 May 2018 for a minimum five-year design lifetime and entered the science phase of its mission on 28 January 2019.

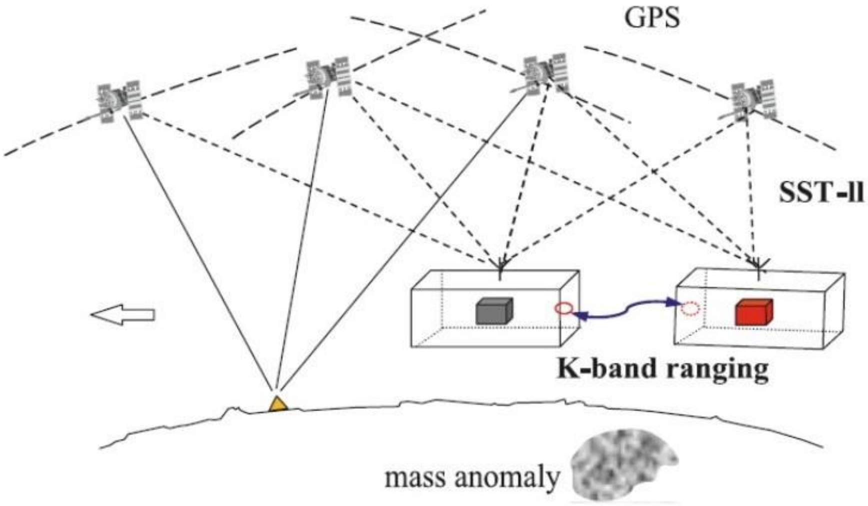


Figure 1. Low orbit GRACE satellites and KBR measurement system; High orbit GPS satellites determine the position of GRACE satellites.[15]

GRACE data are divided into three levels comprising 0, 1, and 2. The structure of data processing is that raw data from the level-0, are received and processed by satellite. Level-2 data consist of two sets of spherical harmonic coefficients. The first set represents the static part, and the other set shows the dynamic part of the gravity field. GRACE satellite data at level-2 are presented on a in the form of monthly geopotential models for the Earth's gravity field.

2.2. Methodology

Using the monthly Stokes coefficients obtained from the GRACE satellites, the monthly water level changes can be evaluated. So, according to the following steps, changes in the water level can be calculated [6].

(a) First, the monthly average of the Stokes coefficients should be determined; then these long-term temporal means are subtracted from the Stokes coefficients to obtain the residual Stokes coefficients of the jth monthly solution:

$$\begin{pmatrix}\delta\bar{C}_{lm}\\ \delta\bar{S}_{lm}\end{pmatrix}_j = \begin{pmatrix}\bar{C}_{lm}\\ \bar{S}_{lm}\end{pmatrix}_j - \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\bar{C}_{lm}\\ \bar{S}_{lm}\end{pmatrix}_i \quad (1)$$

Where, $\bar{C}_{lm}$and $\bar{S}_{lm}$are the monthly Stokes coefficients of degree l and order m obtained from the GRACE satellite, and N is the total number of monthly solutions of available GRACE data.

(b) The Residual surface density coefficients $\delta\check{C}_{lm}$ and $\delta\check{S}_{lm}$ are obtained from the residual Stokes coefficients $\delta\bar{C}_{lm}$ and $\delta\bar{S}_{lm}$, as follows:

$$\begin{pmatrix}\delta\check{C}_{lm}\\ \delta\check{S}_{lm}\end{pmatrix}_j = \frac{\rho_{ave}}{3\rho_w}\frac{2l+1}{1+k_l}\begin{pmatrix}\delta\bar{C}_{lm}\\ \delta\bar{S}_{lm}\end{pmatrix}_j \quad (2)$$

Where, ρ_{ave} is the average density of the earth (5517 kg/m3), ρ_w is the density of water (1000 kg/m3), and k_l is the Love number of degree l [16].

(c) The spatial variations of the surface density can be computed from the following equation:

$$\delta\sigma(\varphi.\lambda) = R\rho_w\left[\left(\delta\check{C}_{00}\right)_j + \sum_{l=0}^{\infty}\sum_{m=0}^{l}\bar{p}_{lm}\,(sin\varphi)\times\left(\delta\check{C}_{lm}cosm\lambda + \delta\check{S}_{lm}sinm\lambda\right)_j\right] \quad (3)$$

Where, R is the radius of the Earth (6378137 m), $\bar{p}_{lm}$ is the normalized Legendre function, and (φ, λ) is the latitude and longitude of the relevant point, respectively.

(d) At last, the changes in the thickness of the water layer can be obtained as below:

$$EWT(\varphi.\lambda)_j = \frac{\delta\sigma(\varphi.\lambda)_j}{\rho_w} \quad (4)$$

Or

$$EWT(\varphi.\lambda)_j = \delta\sigma(\varphi.\lambda)_j \quad (5)$$

Equation 4 represents the changes in the thickness of the water layer in meters; Equation 5 can be extracted by multiplying Equation 4 by 1000 taking into account the density of water equal to 1000 kg⁄m3 which indicates the thickness of the groundwater layer in millimeters.

2.3. GRACE data

In order to estimate the groundwater changes in Khuzestan province, the surface dataset from two GFZ centers between 2007 and 2011 was used. In addition, the coefficient C_20 obtained from GRACE is replaced by the coefficient C_20 derived from the SLR geodetic satellite to improve the results. Then, the Stokes coefficients of the GRACE satellite were used into Equation 1 to obtain the residual Stokes coefficients. Using the residual Stokes coefficients in Equation 2, the residual surface density coefficients can be calculated. Having a gridding with 15' □ 15' cell size for the study area (Khuzestan province), the spatial variations of the surface density can be obtained from Equation 3. Since high-order and degree Stokes coefficients of GRACE data could be introduced some noise in the final results, applying filter is necessary to eliminate these noises. Although the resolution of the region increases with higher degree and order of Stokes coefficients, more errors appear in the results; on the other hand, lower degree and order decrease the errors in the results as well as the resolution of the region.

In this paper, the Gaussian filter was utilized to reduce the noise in higher degree and order of GRACE data. The Gaussian filter equation is as follow [16]:

$$W_0 = \frac{1}{l}, \quad W_1 = \frac{1}{l}\left(\frac{1+e^{-2b}}{1-e^{-2b}} - \frac{1}{b}\right), \qquad W_{l+1} = -\frac{2l+1}{b}W_l + W_{l-1} \quad (6)$$

Where, l is the degree of Stokes coefficient. b is also obtained from the following equation:

$$b = \frac{\ln(2)}{1-\cos\left(\frac{r}{R}\right)} \quad (7)$$

Where, r is the selected radius of applied filter; according to the selected radius of filter, the influence of the Stokes coefficients outside the r will be minimum. Based on the previous studies

[17], the best radius of Gaussian filter for hydrological applications of GRACE data is about 350 km. Thus, in pour research, the filtering radius has been considered to 350 km. Applying the Gaussian filter to the GRACE data, equation 3 will be transformed into the following equation:

$$EWT(\varphi.\lambda) = \frac{a\rho_{ave}}{3}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\frac{2l+1}{1+k_l}W_l\bar{p}_{lm}(sin\varphi) \times [\delta\bar{C}_{lm}cosm\lambda + \delta\bar{S}_{lm}sinm\lambda] \quad (8)$$

Consequently, using the residual Stokes coefficients obtained from Equation 1, the monthly changes of thickness of the groundwater layer will be calculated in millimeters.

2.4. Study area and piezometric data

Khuzestan province, located at the southwest of Iran, with an area of about 64,000 km2, extends between 29° 58^' and 33° 4^' latitude and 47° 41^' and 50° 39^' longitude and covers approximately 4% of the area of Iran. The north and east of Khuzestan are surrounded by the Zagros Mountains, whose altitudes decrease towards the southwest. From the topographical point of view, Khuzestan province can be divided into mountainous regions and lowlands. The mountainous area is located in the north and east of the province, and lowlands start from the south of Dezful, Masjed-Soleyman, Ramhormoz, and Behbahan, and continue to the banks of the Persian Gulf and Arvand Rud. Khuzestan province has different climates: semi-desert climate in some parts and hot steppe climate covers the other parts.

Piezometric wells are one of the most important measuring instruments for quantifying of an aquifer and planning for optimal usage of the groundwater. Having a network of observation wells and measuring the water level in these wells, the fluctuations of the groundwater level can be evaluated. In this study, the data obtained from 446 piezometric wells in 25 regions of Khuzestan province was used to verify the results of GRACE data processing. Figure 3 shows the location of the piezometric wells. The characteristics of the region and the number of wells in each region have been shown in Table 1. Since the piezometric data are collected manually, the personal error may introduce the outliers in the available dataset. Therefore, the Mahalanobis Distance-based data mining method is employed to detect the outliers; then, these identified outliers will be replaced by the correct data extracted from spline regression function. This method is explained in section 2-6.

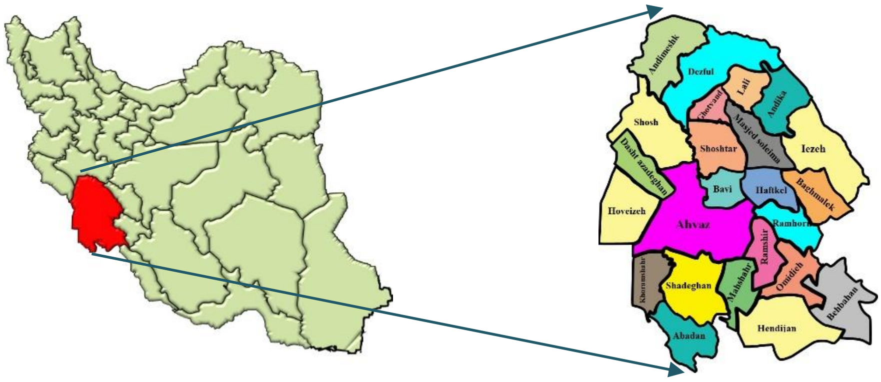


Figure 2. Location of the study area

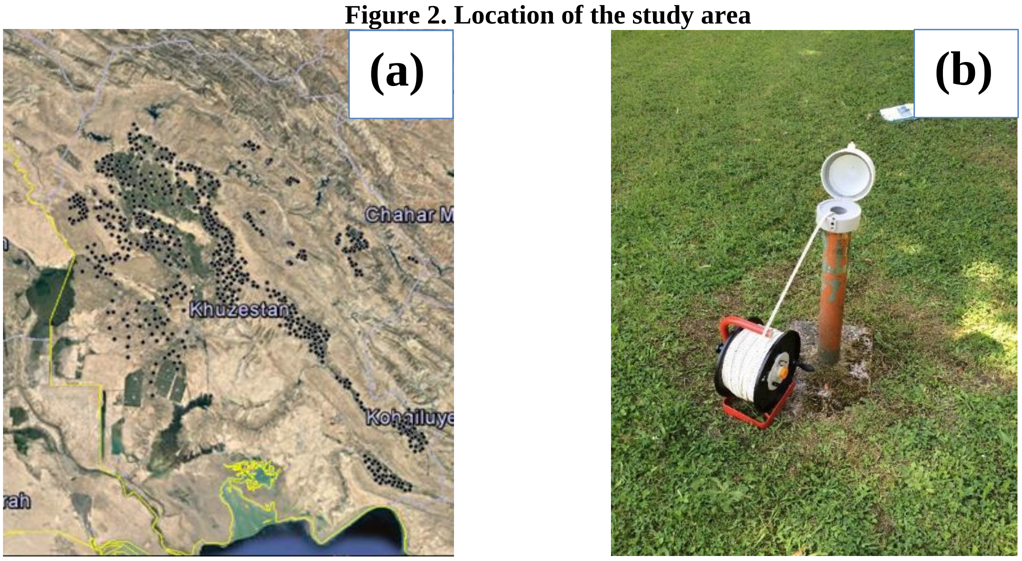


Figure 3. Piezometric wells; (a) spatial distribution of piezometric wells in Khuzestan province; black dots indicate well location. (b) reading process of the piezometer

Table 1. Specifications and number of piezometric wells used in each area

Row	Name of the area	Number of wells	Row	Name of the area	Number of wells	Row	Name of the area	Number of wells
1	Yashmineh zar	6	10	Chahar tang	5	19	Tembi	3
2	Dezful-Andimeshk	65	11	Andika	9	20	Masjed-soleyman	2
3	khowis	59	12	Harkaleh	6	21	Shoeibieh	10
4	Abdolkhan	59	13	Yali	8	22	Kharan	21
5	Dehdez	3	14	Aghili	22	23	Mianab-Shooshtar	47
6	Izeh	21	15	Gotvand	21	24	North Ahwaz	23
7	Ghaleeh tol	9	16	Batvand	11	25	South Ahwaz	16
8	Notreghi	5	17	Golgir	8	Total No. of wells		446
9	Hendijan	3	18	Simili	4			

2.5. The global GLDAS model

The Global Land Data Assimilation System (GLDAS) is a collaborative working between NASA, the Goddard Space Flight Center (GSFC), National Oceanic and Atmospheric Administration (NOAA), and the National Centers for Environmental Prediction (NCEP) [18]. The GLDAS model integrates satellite- and ground-based observational data to provide a database using four land surface models (LSMs) comprising CLM, Mosaic, Noah, and VIC [2]. The Management of such big data set can be achieved using the Land Information System (LIS) [19]. The main characteristic of this model has been illustrated in Table 2 [20,21]. According to Table 2, the hydrological data are in the spatial resolution of 0.25 to 1-degree and the temporal resolution of 3 hours or 1 month.

The format of GLDAS output data is in General Regularly-distributed Information in Binary form (GRIB) and Numerical Control (NC) accessible through some centers such as Hydrology Data and Information Services Center (HDISC). In this study, the surface soil moisture data at a depth of 40 cm with a spatial resolution of 0.25 degree and temporal resolution of 1 month were used for Khuzestan province. The data were extracted from NASA in NC format and executed using GIS software.

Table 2. Basic characteristics of the GLDAS data

Contents	Water and energy budget components, forcing data
Latitude extent	-60° to 90°N
Longitude extent	-180° to 180°E
Spatial resolution	0.25°, 1.0°
Temporal resolution	3-hourly or monthly
Temporal coverage	January 1, 1979 to present for the 1.0° data February 24, 2000 to present for the 0.25° data
Dimension	360 (lon) x 150 (lat) for the 1.0° data 1440 (lon) x 600 (lat) for the 0.25° data
Origin (1st grid center)	(179.5W, 59.5S) for the 1.0° data (179.875W, 59.875S) for the 0.25° data
Land surface models	CLM 2.0 (1.0°) MOSAIC (1.0°) NOAH 2.7.1 (1.0°) VIC water balance (1.0°) NOAH 2.7.1 (0.25°)

2.6. Mahalanobis distance-based method for detecting outliers

Data mining is a process of finding and extracting hidden information, patterns, and relationships in large data sets to suggest a prediction model that is understandable to human users. Nowadays, Data mining has a wide range of applications in various fields of applied sciences. According to Figure 4, clustering is one of the most widely used data mining branches which is significantly used in all sciences.

Detection of outlier data is a pre-processed step in data mining which sometimes can be independently considered as data mining operation. An outlier data is a data point that differs significantly from all of the members of a data set to which it belongs. As a simple definition, it can be stated that an outlier data has an abnormal distance from other data [22]. This may be a result of the inaccuracy in measuring or recording the observations and different resources for data collection and it increases the skewness of probability density function [23,24]. The outlier data in the analysis of the information are structurally and conceptually influential, and in some cases, it is not possible to reasonably deduce from the collected information, because they cause statistical errors. Hence, it can be useful for the researchers to be familiar with the methods of identifying and detecting outliers in the collected information.

In general, the outlier detection methods are divided into three categories including univariate, bivariate, and multivariate methods [25]. In this study, the Mahalanobis distance-based method, which is a branch of multivariate methods, has been applied to detect outliers [26]. The Mahalanobis distance for the i th variable is obtained from the following equation:

$$D_M(x_i) = \left((x_i - \mu)^T\, C^{-1}\, (x_i - \mu)\right)^{1/2} \tag{9}$$

Where, x_i is the vector of variables for the i th observation, μ is the vector of average variables, and C is the covariance matrix of the sample.

In this study, the Mahalanobis distance is calculated for all observations of a specific piezometric well; afterward, based on this distance, it can be labelled as an outlier data or not.

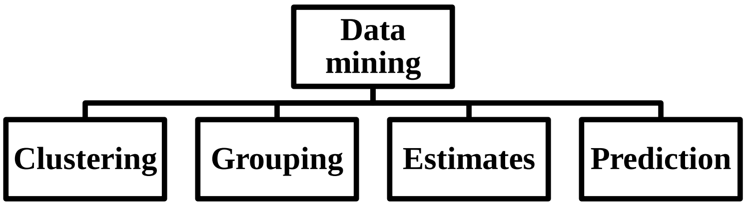


Figure 4. Data Mining Applications

3. Results

Before the analyzing of GRACE data, a pre-processing is performed on the piezometric data. In this step, the Mahalanobis Distance and a MATLAB code are employed to investigate the outliers in observational data of each piezometric well. In Figure 5, the outlier detection algorithm is implemented on a piezometric well with the coordinates λ=48.5570942°E and φ=31.1940226°N, located at the south of Ahvaz, is observed that on its data. As shown in Figure 5, the 36th-month data has been detected as an outlier data using the Mahalanobis algorithm.

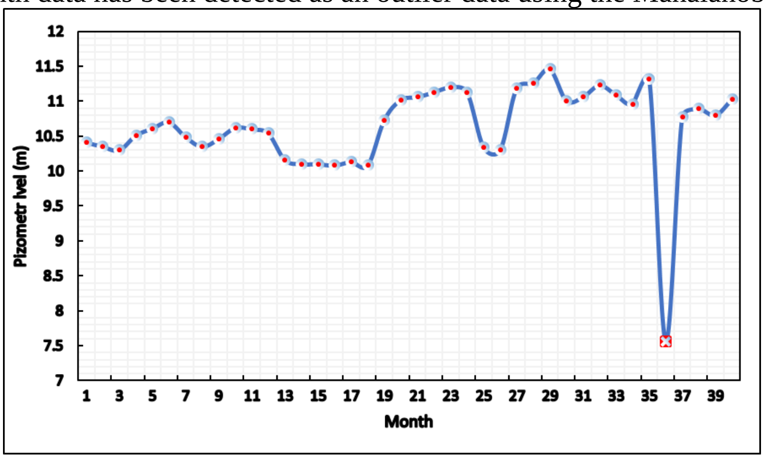


Figure 5. Outlier detection algorithm applied to piezometric data.

The 36th-month data has been detected as an outlier data Similarly, this algorithm has been applied to all piezometric data. The total number of observations were 19106 record and 2573 data were identified as the outlier data by the applied algorithm. The details of the outliers for each region have been presented in Table 3.

Table 3. Number and percentage of outliers per region

Row	Name of the area	Number / Percent Outdated data	Row	Name of the area	Number / Percent Outdated data	Row	Name of the area	Number / Percent Outdated data
1	Yashmineh zar	38 / 15.45	10	Chahar tang	32 / 15.61	19	Tembi	21 / 17.07
2	Dezful-Andimeshk	401 / 15.05	11	Andika	62 / 16.80	20	Masjed soleiman	18 / 21.95
3	khowis	325 / 13.04	12	Harkaleh	40 / 16.62	21	Shoeibieh	53 / 12.93
4	Abdolkhan	325 / 13.04	13	Yali	49 / 14.94	22	Kharan	118 / 13.70
5	Dehdez	17 / 13.82	14	Aghili	124 / 13.75	23	Mianab-Shoshtar	233 / 12.09
6	Iezeh	154 / 17.89	15	Gotvand	115 / 13.36	24	North Ahwaz	122 / 12.94
7	Ghaleeh tol	52 / 14.09	16	Batvand	62 / 13.75	25	South Ahwaz	80 / 12.20
8	Notreghi	41 / 20	17	Gholghir	54 / 16.46		Total	2573 / 13
9	Hendijan	18 / 14.6	18	Simoli	19 / 11.59			

After finding and removing the outlier data, a regression is fitted through the residuals using the third-order Spline function to obtain the substituted value. The Spline function is fitted through the well data, which was used in figure 5, and is demonstrated in Figure 6. As indicated in Figure 6, the 36th-month outlier data has been replaced by the value of 11.07 extracted from the regression.

After the pre-processing of the observational data, the output from the GRACE data processing was calculated without applying filters and soil moisture (Figure 7). As shown in Figure 7, not applying filters to the GRACE data can cause adverse effects, which challenges the feasibility of the GRACE data. For this reason, the Gaussian filter was applied to the GRACE data (Figure 8). According to the previous studies, the best radius of the Gaussian filter was selected 350 km (Joodaki 2014); then, the GRACE model was implemented via applying the Gaussian filter without applying soil moisture obtained from GLDAS model (Figure 9).

The Root Mean Square Error (RMSE) and the Normalized Root Mean Squared Error (NRMSE) were estimated to 0.8 and 1.3, respectively. It should be noted that in order to achieve groundwater level changes, the data of 446 piezometers gathered from throughout the Khuzestan province have been used.

To increase the accuracy, the amount of soil moisture obtained from the GLDAS model is also added to the observational data. As shown in Figure 10, considering the soil moisture led to the RMSE and NRMSE, 0.8 and 0.2, respectively representing the higher accuracy.

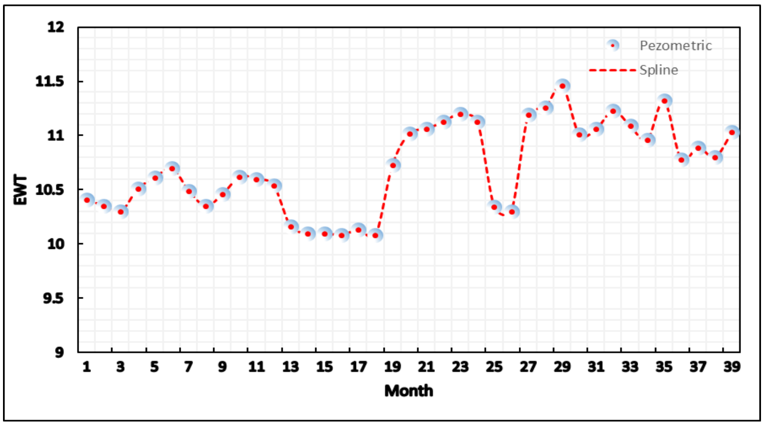


Figure 6. Fitting the spline function to the observed data

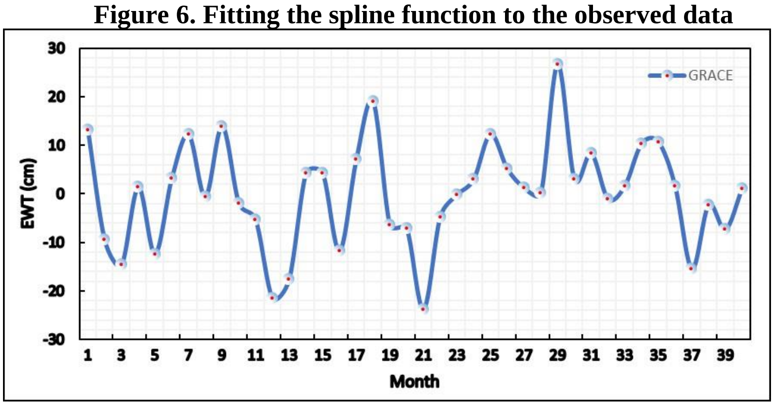


Figure 7: Output of GRACE data without applying Gaussian filter

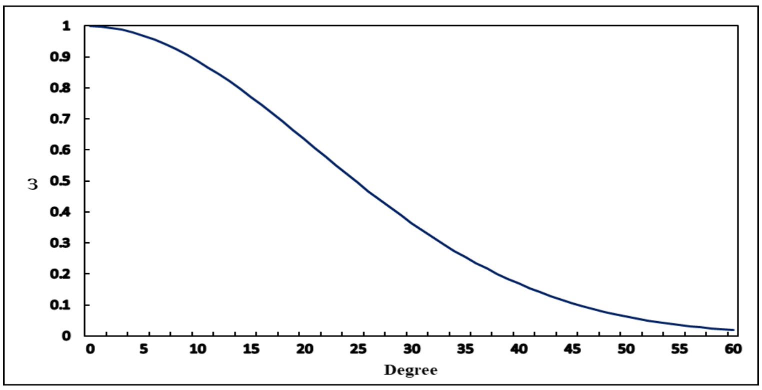


Figure 8: Gaussian filter

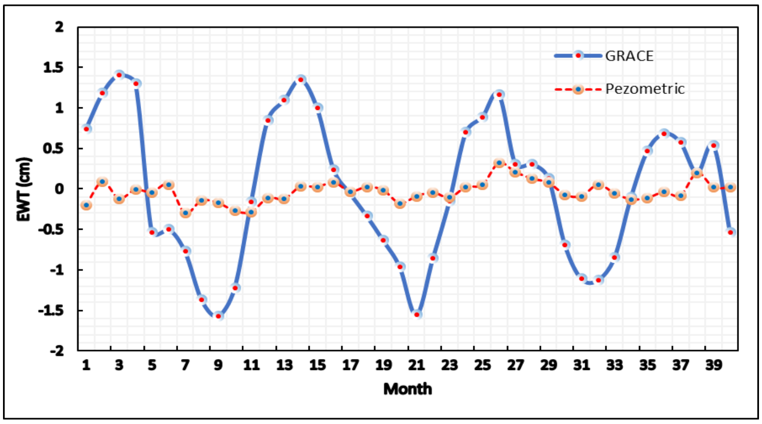


Figure 9: GRACE results and piezometric wells after applying the Gaussian filter

As mentioned earlier, the outlier data, which were detected using the Mahalanobis algorithm, have been removed from the data set, and consequently, the substituted values have been estimated using the Spline function. Figure 11 illustrates the model output for a situation without applying the outlier detection algorithm. In this case, the RMSE and NRMSE were estimated 0.7 and 0.3, respectively. Based on the NRMSE calculated for both situations, it was concluded that outlier detection algorithm improved the results.

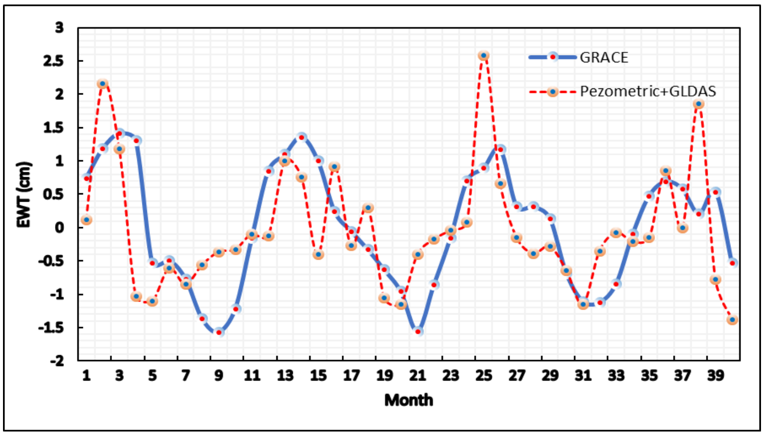


Figure 10.Comparison of the results of GRACE with observations of piezometric wells after applying Gaussian filter and soil moisture content of GLDAS model (using outlier detection algorithm)

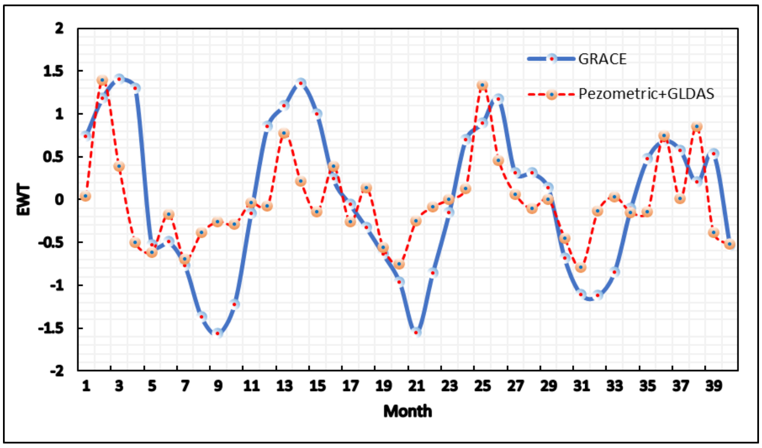


Figure 11. GRACE results, piezometric wells and GLDAS without applying outlier detection algorithm

4. Discussion and Conclusion

Wells are one of the main resources of freshwater supply. Nowadays, the severe loss of groundwater level causes adverse impacts like the huge pits that have occurred in many parts of Iran. Hence, the monitoring of groundwater can be considered one of the most essential actions for managing the available water resources. For along years, the wells have been monitored through the observational wells that were excavated in the specific regions. This method is pointwise and not providing any useful information for other areas. The increasing growth of the remote sensing techniques like GRACE satellite makes it possible to efficiently monitoring the groundwater resources. Using this tool enables water resource managers to have a comprehensive insight into the groundwater changes over a large area.

In this study, the GRACE satellite and observational data that were used for estimating and comparing the groundwater level changes in Khuzestan province are from November 2007 to April 2011. A full description of the observation wells has been given in Table 1. In this study, the monthly changes in water level for all observation wells have been averaged throughout the study area. As well, due to the missing data on December 2010 and January 2011 in the GRACE database, these two months were not included in the calculations.

The existence of noise in the high degrees and orders is one of the challenges that researchers are facing with the GRACE satellite. To overcome this challenge, a Gaussian filter with a radius of 350 km has been used. Comparison of Figures 7 and 9 results that using filters in the GRACE equations made more accurate responses. For using this filter, it is important to consider a reasonable trade-off between the error value and the applied resolution of the GRACE. Because of reducing the degree and order, the noise in the GRACE output and the resolution of the area are both decreased.

The existence of the outlier data in the observed data can be considered as a source of error. This occurs due to the traditional approach of data gathering from the piezometric wells. Indeed, one of the main reasons for the weakness of traditional methods is the existence of such errors in the observational data. One of the applications of data mining is the detection of outlier data. In this study, the outlier data were detected using the Mahalanobis Distance method; then the outlier data were replaced by the modified values obtained from the Spline function through the remaining observational data. In this study, the total number of observational data is 19106 data, from which 2573 data were identified as the outlier data using the outlier detection algorithm which comprise 13% of the total data. Among the regions in Table 1, the maximum and the minimum number of outlier data were assigned to the Masjed Soleiman with 21.95% outlier data and the Simile with 11.59% outlier data, respectively.

In Figure 9, the modified GRACE outputs via the Gaussian filter were compared with the pre-processed observational data. Since the soil moisture is also considered as the GRACE output, the global model GLDAS was employed to compensate for this deficit in observational data. The GLDAS data at the soil depth of 40 cm were used. The GLDAS data were implemented with the GIS software and interpolated monthly for the Khuzestan province. A comparison between the GRACE model output and the observational wells added into the GLDAS model indicated the RMSE and NRMSE, 0.8 and 0.2, respectively, while without applying the outlier detection algorithm, these values were obtained 0.7 and 0.3, respectively. Based on the NRMSE, the outlier detection algorithm enhanced the results.

References

1. Esmaeili, Mand Motagh, M (2016) Improved persistent scatterer analysis using amplitude dispersion index optimization of dual polarimetry dataISPRS Journal of Photogrammetry and Remote Sensing, 117, pp.108-114.
2. Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., & Bosilovich, M (2004) The global land data assimilation systemBulletin of the American Meteorological Society, 85, pp.381-394
3. Abart, C (2005) Assessment of solution strategies for GRACE gravity field processing.
4. Wahr, J., Swenson, S., Zlotnicki, V., & Velicogna, I (2004) Time-variable gravity from GRACE: First resultsGeophysical Research Letters, 31.
5. Ramillien, G., Famiglietti, J.S., & Wahr, J (2008) Detection of continental hydrology and glaciology signals from GRACE: a reviewSurveys in geophysics, 29, pp.361-374
6. Awange, J., Sharifi, M., Keller, W., & Kuhn, M (2009) GRACE application to the receding Lake Victoria water level and Australian droughtObserving our changing earth: Springer, pp387-396
7. Longuevergne, L., Scanlon, B.R., & Wilson, C.R (2010) GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USAWater Resources Research, 46.
8. Voss, K.A., Famiglietti, J.S., Lo, M., De Linage, C., Rodell, M., & Swenson, S.C (2013) Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran regionWater Resources Research, 49, pp.904-914
9. Ferreira, V.G., Gong, Z., & Andam-Akorful, S.A (2012) Monitoring mass changes in the Volta River basin using GRACE satellite gravity and TRMM precipitationBoletim de Ciências Geodésicas, 18, pp549-563
10. Ahmed, M., Sultan, M., Wahr, J., Yan, E., Milewski, A., Sauck, W., Becker, R., & Welton, B (2011)Integration of GRACE (Gravity Recovery and Climate Experiment) data with traditional data sets for a better understanding of the time-dependent water partitioning in African watershedsGeology, 39 , pp479-482
11. Ahmed, M., Sultan, M., Wahr, J., & Yan, E (2014) The use of GRACE data to monitor natural and anthropogenic induced variations in water availability across AfricaEarth-Science Reviews, 136, pp.289-300

12. Rodell, M., Famiglietti, J., Wiese, D., Reager, J., Beaudoing, H., Landerer, F., & Lo, M (2019) Emerging trends in global freshwater availability Nature, 565, E7-E7, vol 557, pg 651.
13. Tapley, B.D., Bettadpur, S., Watkins, M., & Reigber, C (2004) The gravity recovery and climate experiment: Mission overview and early resultsGeophysical Research Letters, 31.
14. Buis, A (2012) At 10, GRACE Continues Defying, and Defining, GravityNASA.
15. Rummel, R., Balmino, G., Johannessen, J., & Visser, P.a.W., P (2002) Dedicated gravity field missions-principles and aimsJGeodyn, 33, pp3-20
16. Wahr, J., Molenaar, M., & Bryan, F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACEJournal of Geophysical Research: Solid Earth, 103, pp30205-30229
17. Joodaki, G (2014) Earth mass change tracking using GRACE satellite gravity data.
18. Houser, P., & Rodell, M (2002) GLDAS: an important contribution to CEOPGEWEX Newsletter, May, 2, 9.
19. Fang, H., Hrubiak, P., Kato, H., Rodell, M., Teng, W.L., & Vollmer, B.E (2008) Global land data assimilation system (GLDAS) products from NASA hydrology data and information services center (HDISC).
20. Fang, H., Beaudoing, H.K., Teng, W.L., & Vollmer, B.E (2009) Global Land data assimilation system (GLDAS) products, services and application from NASA hydrology data and information services center (HDISC).
21. Rui, H., & Beaudoing, H (2015) Global Land Data Assimilation System Version 2 (GLDAS-2) Products, Last revisedNational Aeronautices and space administration.
22. Grubbs, F.E (1969) Procedures for detecting outlying observations in samplesTechnometrics, 11, pp.1-21
23. Barnett, V.L., T (1994) Outliers in statistical dataSchool of Mathematical & Physical Sciences, National Technical University of Athens, Greece.
24. Das, P., & Haimes, Y.Y (1979) Multiobjective optimization in water quality and land managementWater Resources Research, 15, pp1313-1322
25. Hair, J.F., & R. E (1998) Andersen, R.LMultivariate data analysisRentice Hall, Upper Saddle River, New Jersey,.
26. Filzmoser, P., Garrett, R.G., & Reimann, C (2005) Multivariate outlier detection in exploration geochemistryComputers & geosciences, 31, pp.579-587