Groundwater Level Mapping Tool: An open source web application for assessing groundwater sustainability

Steven W. Evans
Norman L. Jones
Brigham Young University - Provo, njones@byu.edu
Gustavious P. Williams
Brigham Young University - Provo
Daniel P. Ames
Brigham Young University - Provo
E. James Nelson
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Other Civil and Environmental Engineering Commons

BYU ScholarsArchive Citation
Evans, Steven W.; Jones, Norman L.; Williams, Gustavious P.; Ames, Daniel P.; and Nelson, E. James, "Groundwater Level Mapping Tool: An open source web application for assessing groundwater sustainability" (2020). Faculty Publications. 4266.
https://scholarsarchive.byu.edu/facpub/4266

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Groundwater Level Mapping Tool: An open source web application for assessing groundwater sustainability

Steven W. Evans, Norman L. Jones, Gustavious P. Williams, Daniel P. Ames, E. James Nelson

A Kiewit Infrastructure Engineers, Englewood, CO, USA
b Civil and Environmental Engineering, Brigham Young University, Provo, UT, USA

ARTICLE INFO

Keywords:
Groundwater
Aquifer storage
Data imputation
Time series
Water resources management
Sustainability

ABSTRACT

Decision makers need an accurate understanding of aquifer storage trends to effectively manage groundwater resources. Groundwater is difficult to monitor and quantify since the data collected from monitoring wells are often available only at irregular and infrequent intervals. We present an open-source web application (app) to visualize groundwater data over time and automatically calculate changes in aquifer storage volume to help managers assess aquifer sustainability. This app uses a novel multi-linear regression (MLR) algorithm to impute missing data for infrequently sampled wells, using correlated data from other wells in the same aquifer. The app uses this MLR-imputed data to spatially interpolate water levels throughout an aquifer at user-specified time steps using GSLIB Kriging code. Based on our tests of unconfined aquifer systems, the imputed data increased the accuracy of the spatial interpolation over standard methods and resulted in estimates of aquifer storage change comparable to those of detailed USGS studies.

1. Introduction

Worldwide, groundwater is a major source for agricultural irrigation, industrial processes, mining, and drinking water. The United States Geological Survey (USGS) reports that 30.1% of the earth’s fresh water consists of groundwater, while 1.2% consists of surface water in lakes, rivers, and streams (Gleick, 1993). Although fresh groundwater is abundant, using this resource in a responsible and sustainable manner poses a significant challenge.

One of the great challenges for sustainable groundwater management is the ability to accurately characterize the changing state of an aquifer over time. Accurate characterization allows managers to implement sustainable practices, procedures, and regulations. Although fresh groundwater is abundant, and extensively used, it is difficult and expensive to accurately quantify, compared to surface water resources. The state of surface water resources is readily visible to the naked eye, viewable from satellites, can be measured easily, and is straightforward to quantify. This is not the case for groundwater, which generally requires drilling a series of monitoring wells in order to measure the location of the phreatic surface and characterize aquifer properties. While surface storage in a waterbody reaches approximately the same elevation throughout, groundwater surface elevations may vary significantly throughout an aquifer, depending on the overlying land use, pumping of irrigation wells, aquifer recharge, and other factors. Groundwater levels are heavily influenced by climatic, geographic, lithological, and human factors. For these reasons it is difficult to quantify and map aquifer water levels and storage volume changes.

Even when data on groundwater elevations are available from monitoring wells, these data are generally not harnessed to their full potential to aid in decision-making. These data are available at point locations scattered in time and space throughout an aquifer, and it is difficult to piece these segments of data together into a complete picture of aquifer-wide behavior (Marchant and Bloomfield, 2018). Point data from monitoring wells are typically sparse and give only a limited sampling of the spatial distribution of water levels in the aquifer, and the data observations from these monitoring wells are often temporally sporadic, including large gaps in the time series data (Oikonomou et al., 2018).

Because of these issues, quantifying groundwater resources requires both spatial and temporal interpolation and extrapolation to at least some extent. This presents two related, but separate problems, the first is temporal interpolation or imputation of temporal data at a well, and the
second is spatial interpolation of these well data at a specific time step to generate a groundwater surface elevation map.

One of the most widely used computer programs for spatial interpolation in this and other fields is the Geostatistical Software Library (GSLIB), developed at Stanford University (Deutsch and Journel, 1992). This library contains functionality for spatial interpolation using the Kriging technique pioneered by the South African statistician and mining engineer, Danie G. Krige. Kriging is used by many groundwater researchers when attempting to interpolate spatial data, as it provides both spatial estimation and also a map of the uncertainty associated with the interpolated data. Kumar (2006) used Kriging interpolation to estimate unknown depths to water table in an aquifer in Rajasthan, India and found that kriged groundwater levels satisfactorily matched the observed groundwater levels. Other researchers demonstrated that the accuracy of groundwater surface elevation maps could be improved in some cases by introducing topography to the interpolation using Kriging with an external drift (Boezio et al., 2006). Since groundwater levels change over time, sometime significantly, in most studies that use spatial interpolation of groundwater, the data are first lumped into temporal bins, and then interpolated spatially, assuming data within the bin is all from a specific time step (Ruybal et al., 2019).

Several techniques have also been developed and used for the temporal interpolation and imputation of well time series observations. Rouhani and Wackernagel (1990) used Kriging to perform temporal interpolation of depth to water table time series measurements in a basin south of Paris, France. Bidwell (2005) forecasted groundwater levels one month ahead in Canterbury, New Zealand using an ARMAX model based on the eigenstructure of aquifer dynamics. Others have used classical time series models including auto-regressive (AR), moving-average (MA), auto-regressive moving-average (ARMA), auto-regressive integrated moving-average (ARIMA), seasonal auto-regressive integrated moving-average (SARIMA), and multiple linear regression (MLR) to predict groundwater levels (Khorasani et al., 2016; Mirzavand and Ghazavi, 2015; Sahoo and Jha, 2013). Oikonomou et al. (2018) employed an exogenous seasonal autoregressive integrated moving average (SARIMAX) stochastic model to describe the simulated groundwater level fluctuation process of a regional physical groundwater model and the Ensemble Smoother (ES) for predicting groundwater levels.

More recently, researchers have sought for increased accuracy in quantifying groundwater by using spatiotemporal interpolation. Ruybal et al. (2019) used the Arapahoe aquifer as a case study to demonstrate the benefits of spatiotemporal kriging over spatial kriging across a sparsely gauged and irregularly sampled aquifer. They found that spatiotemporal kriging allowed estimation of groundwater levels during times when data are not available, and avoided biases and anomalies caused by kriging with different data available in different time periods. Ahmadi and Sedghamiz (2007) conducted spatial and temporal analysis of an Iranian aquifer and concluded that “spatial structure was a little bit stronger than temporal structure.” Combining spatial and temporal interpolation yields more accurate estimates by leveraging both temporal and spatial relationships between observations (Ruybal et al., 2019).

We have developed an open source Python-based web app to allow visualization and quantification of groundwater resources, leveraging both temporal and spatial interpolation. We temporally interpolate time series data at each well using correlations with other wells in the aquifer, then spatially interpolate these data to generate groundwater level maps at selected time steps. This application is generalized to allow its use world-wide, and allows decision makers to accomplish the following:

1) View time series and other data for each well within an aquifer.
2) View maps and animations of aquifer-wide groundwater levels at different time periods.
3) Calculate and view estimates of total aquifer storage change.

This paper details the temporal and spatial interpolation methods used by the app to map aquifer drawdown and calculate aquifer storage change. Temporal interpolation is accomplished using MLR and exploiting correlated observation wells within an aquifer. Spatial interpolation is accomplished using simple kriging as implemented by GSLIB (Deutsch and Journel, 1992). We present the results of these interpolation methods and compare to other spatial and temporal interpolation methods. We compare the results of storage change calculated by the app to those detailed in USGS studies.

2. Methods

2.1. MLR exploiting correlated wells

To accurately impute data to address data gaps and to extend data beyond its sampled range, we developed a temporal interpolation method using MLR that exploits the correlation of wells within an aquifer. As wells within an aquifer are subject to similar forcing functions such as local recharge, water demand, and climate, water levels changes in these wells are often correlated. However, because of different well uses, the nearest well may not be the most correlated. We developed this method by assuming that wells within the same aquifer will likely exhibit similar characteristics in water table fluctuations and trends. Fig. 1 demonstrates visually these correlations between wells located in the same aquifer. Though there are some differences, there is a clear correlation between these wells as can be seen by the water levels rising and falling in the same general pattern throughout the aquifer. The three wells shown are located in the unconfined Cedar Valley Aquifer in southern Utah, USA. Well 37342113100801 (green) is located near the southern end of the aquifer, while the two other wells are located near the center (Fig. 2). We selected the Cedar Valley Aquifer as a test case because it contains several wells with a sufficiently long and detailed period of groundwater measurements and because the wells in this aquifer exhibit trends and patterns that pose significant difficulties for time series analysis. The time series in this aquifer are non-stationary, and neither increase nor decrease at constant rates.

The MLR method of time series interpolation uses this intra-aquifer correlation among the well data to compute estimates of the water surface elevation at a given well at a specific time. Having actual or estimated data at each well at each time step, allows a more accurate spatial interpolation to map the groundwater surface elevation. This approach can be used for extrapolation to extend the sampling range of a well, or for interpolation to impute data to fill large gaps within a time series. For estimating well data at a specific time, typically between two measured or estimated points, we use a curve fitting approach.

As the most accurate spatial maps are computed with data at every well, this method is particularly useful when the times series data for wells within an aquifer do not completely cover the time(s) of interest resulting from data gaps at a given well, only partial series at a given well, or well data that are sampled at dates different from the data at the other wells. Fig. 3 shows the time series data for Well 374210113044801 located in the Cedar Valley Aquifer. Observed data for the well exists from 1998 to 2008, however we used MLR to extend well data to the period of interest spanning from 1985 to 2015. While these extrapolated data may not be accurate for a given well, especially for long extrapolation periods, we will show that this method provides more accurate spatial interpolations.

2.2. Temporal interpolation

The web app is based on the Python programing language and we use a number of Python tools and libraries to facilitate both programing and computation. The data, which consist of the time series information for every well in the aquifer, are stored in a single Pandas data frame constructed using the Pandas Python library ( McKinney, 2010). This data frame includes an index of regular 3-month intervals and we resample...
the time series observations of each well to these regular 3-month intervals using piecewise cubic Hermite interpolating polynomial (PCHIP) method (Fritsch and Carlson, 1980). PCHIP interpolation has a continuous first derivative, so the data are smooth, but honors the data limits or extents unlike cubic spline interpolation which is commonly used. We only use PCHIP interpolation for short intervals, if a well does not have data within a 3-month interval of the point in time to estimate, then the resulting data frame contains a null or missing value for that well at that time step. Next, those wells which contain data spanning the time period of interest are identified as “reference wells.” After identifying these reference wells, which contain full time series data for the period of interest, the Pearson Correlation Coefficient is calculated between the reference wells, which contain full time series data for the period of interest. The five reference wells are normalized using the minimum and maximum values of each well. These normalized data sets, bounded in red, are plotted in Fig. 5. The five reference wells are normalized as
\[
\frac{X_{i} - \min(X)}{\max(X) - \min(X)}
\]
and each of the reference wells. This correlation coefficient \( r_{XY} \) between the target well \( X \) and each reference well \( Y \) for \( n \) data points is calculated by Equation (1).
\[
r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}
\]

Data from the reference wells \( Y \) with the five highest correlation values \( r_{XY} \) are used as inputs for the process of MLR. Fig. 4 shows the time series data for the target well with missing time series data along with the 5 most correlated wells in the aquifer. The area within the red rectangle shows the data that will be used to train the MLR model to predict the target well’s missing time series values.

To effectively estimate the missing values of the target well, it is necessary to normalize the data. Notice from Fig. 4 the similar shape, but differing depths of each of the wells. This normalization transforms the data so that the data from the wells are in the same range, which allows prediction based on the similar shapes of the time series rather than the differing magnitudes. Data from the target and reference wells are normalized such that all values within the training data set are between 0 and 1. Let \( Y \) represent the target well, and \( R^1, R^2, R^3, R^4, \) and \( R^5 \) represent the five reference wells. The training data subsets will be represented as \( Y_{train} \) for the target well, and \( R_{train1}, R_{train2}, ..., R_{train5} \) for the five reference wells. The values of each well \( i \) are normalized as shown in Equation (2).
\[
Y_{norm} = \frac{Y - \min(Y_{train})}{\max(Y_{train}) - \min(Y_{train})}
\]

These normalized data sets, bounded in red, are plotted in Fig. 5. The reference wells are normalized using the minimum and maximum values of the training subset, rather than the entire data set, because this provides a more skilled prediction, especially when predicting time series values outside the scope of the training data.

The normalized time series data of the target well are assumed to be a linear combination of the normalized time series data of the five reference wells. At each time \( t \), the normalized value of depth to groundwater at the target well \( Y_t \) is approximated as a linear combination of the normalized depth to water table at time \( t \) of the reference wells \( R_t \). The equation for \( Y_t \) is shown in Equation (3), where \( e \) is an error residual term.
\[
Y_t = R^1\beta_1 + R^2\beta_2 + R^3\beta_3 + R^4\beta_4 + R^5\beta_5 + e
\]

Equation (3) can be rewritten in matrix form as Equation (4).
\[
\vec{Y}_t = \vec{R}\vec{\beta} + e
\]

The \( \vec{\beta} \) term represents weights determined by a regularized least-squares fit using the training data subset, such that the sum of the squared residuals \( e \) is minimized. These weights \( \vec{\beta} \) are obtained by solving Equation (5), where \( \lambda \) is a regularization term, and \( I \) is the identity matrix. This process is carried out in the app by using the regularized least-squares solver in the Statsmodels Python library (Seabold and Perktold, 2010).
When using a high complexity estimator, such as MLR, the bias error is generally small, and the variance very large. Bias error causes the estimator algorithm to miss relevant relations between features and target outputs, thereby under fitting the data. Variance is an error from sensitivity to small fluctuations in the training set, which can cause the model to train to random noise in the training data, rather than actual correlations, thereby overfitting the data. The introduction of $\lambda$ increases the bias of the estimate, but significantly decreases the variance, preventing overfitting and generally yielding a more accurate estimate. We used the Tikhonov or Ridge regularization method (Tikhonov and Arsenin, 1977).

Once $\beta$ has been obtained by means of the regularized least squares model using the training data subset, the unknown $Y$ terms may be estimated for each time step $t$ by solving Equation (3). Once the $Y$ terms have been obtained, they must be rescaled to their original extent, yielding $Y\text{estimate}$. This is accomplished by applying Equation (6).
Fig. 4. Well with missing data plotted with correlated reference wells.

Fig. 5. Normalized target well with normalized correlated reference wells.
The process is now complete for estimating missing time series values for the target well. The estimated time series $\hat{y}_{\text{estimate}}$ modelled using MLR that exploits correlated wells in the aquifer is shown in Fig. 6, together with the original recorded data.

### 2.3. Kriging interpolation

With the extended time series information from the MLR process, we have groundwater surface data at specific points scattered throughout an aquifer available for each 3-month time step within our period of interest. We can now use these point data to spatially map groundwater levels throughout the entire aquifer at any or all of the three-month time steps using Kriging interpolation.

For each time step to be mapped, we develop a variogram for use in the Kriging interpolation. A model variogram is developed by first creating and plotting the experimental variogram from the available point data. The experimental variogram is a function of the Euclidean distance $d(i,j)$, and the semivariance $V(i,j)$ between each pair of points $(i,j)$ in the dataset. The Euclidean distance $d(i,j)$ between a pair of points $(i,j)$ with coordinates $(x_i, y_i)$ and $(x_j, y_j)$ is calculated for each pair of points in the dataset using Equation (7).

$$d(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{7}$$

The semivariance $V(i,j)$ between a pair of points $(i,j)$ with data values $z_i$ and $z_j$ is calculated for each pair of points in the dataset using Equation (8).

$$V(i,j) = \frac{(z_i - z_j)^2}{2} \tag{8}$$

An ordered pair $(d(i,j), V(i,j))$ of distance and semivariance is created for each pair of points in the dataset. The ordered pairs of distance and semi-variance $(d(i,j), V(i,j))$ are sorted into ten bins of equal intervals based on distance $d(i,j)$. The range of these bins extends from the minimum distance between two observation points in the aquifer to the sum of the minimum distance and the half the maximum distance between observation points in the aquifer. This approach of eliminating extremely distant points from the bins produces a more accurate estimation because the smallest distances are the most important for developing a proper variogram for accurate estimation (Kitanidis, 1997). The use of the distant observations skews the automatic variogram fitting routine towards the large variances of the distant observations, which negatively impacts the estimation. We do not expect data to be correlated at large distances, so eliminating these points fits our conceptual model of the data.

Once the ordered pairs are sorted into the ten bins spanning equal intervals, the mean distance $d$ and semivariance $V$ are computed within each bin. These ordered pairs compose the experimental variogram, and are used to fit a spherical variogram model, which has been found to produce the best results for groundwater level data by several researchers, as well as from initial inspection of preliminary test case results (Gundogdu and Guney, 2007; Nikroo et al., 2010). The spherical variogram model $V(d)$ is a function of the distance $d$ of an observation point from the interpolation point and is given by Equation (9), defined by the nugget $(n)$, range $(r)$, and sill $(s)$, where the partial sill $(p)$ is the difference between the sill and the nugget.

$$V(d) = \begin{cases} 
 n + p \left( 1 - \frac{d}{r} \right) & d \leq r \\
 p & d > r 
\end{cases} \tag{9}$$

For a given nugget value, $n$, the optimal parameters $r$ and $p$ are determined using least squares optimization. We fit the model by minimizing the residuals of the spherical variogram model compared to the ordered pairs of the experimental variogram. In this optimization, the residuals are weighted using a logistic function so that the weights vary from a value approaching one at distance zero to a value approaching zero as the distance increases. This weighting allows the least squares optimization to fit the data better at closer distances, producing a better variogram for estimation (Kitanidis, 1997). Fig. 7 shows an example of a spherical variogram model fitted to an experimental variogram using this method.

With the variogram parameters defined, the web application carries out the interpolation throughout the aquifer at a user-defined resolution by employing a Python-wrapping of the GSLIB Fortran code (Deutsch and Journel, 1992). In cases where the data yields a singular matrix, the app is written to revert to a simple inverse-distance weighted interpolation (Franke and Nielson, 1980; Shepard, 1968) for these singular grid points.

### 2.4. Calculation of aquifer storage change

Using the results of the temporal and spatial interpolation, it is possible to calculate changes in total aquifer storage volume. This is accomplished by performing mathematic operations on $n$ series of raster datasets, $R$, of groundwater levels produced at specific times by the spatial interpolation phase. The first dataset (corresponding to the earliest time step in the series), known as $R_0$, serves as the baseline from which changes in aquifer storage are measured.

These changes are calculated by first calculating the drawdown $D_i$ by subtracting, $R_i$ from the baseline $R_0$, for each time step, $n$, in the raster series. The drawdown $D_i$ is calculated on a cell-by-cell basis by applying Equation (10) for each of the $n$ timesteps, resulting in a new set of $n - 1$ raster datasets of drawdown.

$$\forall i \in \{1, \ldots, n\} \quad D_i = R_i - R_0 \tag{10}$$

**Fig. 6.** Well time series modelled by MLR
Aquifer-wide storage changes \( C_i \) are then calculated for each time step by multiplying the drawdown \( D_j \) at each grid cell \( j \) by the average aquifer storage coefficient \( p \) and the grid cell area \( A_j \), and summing over all grid cells in the aquifer, as shown in Equation (11).

\[
\forall i \in \{1, \ldots, n\} \quad C_i = \sum_{j=1}^{\text{Aquifer}} D_j p A_j
\]

(11)

The aquifer storage coefficient \( p \) will be either the storativity for a confined aquifer or the specific yield for an unconfined aquifer. Both coefficients are dimensionless and represent the volume of water lost from a unit area of an aquifer due to a unit decline in head. The specific yield is close to, but typically smaller than the porosity. Storativity is equal to the specific storage multiplied by the aquifer thickness. These values may vary throughout an aquifer, and an aquifer may even be partially confined and partially unconfined. This tool is meant to produce a rough estimate of storage volume change and currently uses an average storage coefficient for the entire aquifer. If available, storage coefficients could be assigned to each grid cell and the tool could be modified to determine a more accurate estimate, and the \( p \) parameter in Equations (2)–(11), would become \( p_j \).

The area \( A_j \) is not constant for each grid cell over the dataset, since the grids are defined at a specified latitude and longitude resolution. Each cell has constant height, but the cell width is dependent on the cell latitude. Cells closer to the equator will have larger widths than those nearer the poles. The area of each grid cell \( A_j \) in the aquifer is calculated based on the resolution of the grid \( g \), the mean radius of the Earth \( R \), and the latitude \( l_j \) of the center of each grid cell \( j \) as shown in Equation (12).

\[
\forall j \in \{\text{Aquifer}\} A_j = R^2 \sin g \left( \sin \left( \frac{l_j}{2} \right) - \sin \left( \frac{l_j - g}{2} \right) \right)
\]

(12)

When the Groundwater Level Mapping Tool is in metric mode, the mean radius of the Earth \( R \) is 6,371,000 m \((\text{Moritz, 1980})\). The aquifer storage is calculated in cubic meters, since the depth to groundwater measurements and therefore drawdown \( D_j \) is also in meters. In imperial units, the Groundwater Level Mapping Tool reports changes in aquifer storage volume in Acre-ft. The app uses a mean Radius of the Earth \( R \) of 3959 miles and is converted to acre-ft for reporting and display.

An example of the output of this storage volume change calculation procedure for the Cedar Valley, UT Aquifer is shown in Fig. 8.

2.5. Web application interface

To facilitate ease of access for groundwater managers to the algorithms described above, the Groundwater Level Mapping Tool has been developed as a web application with a simple user interface. The tool is built on Tethys Platform, an open source platform that lowers the barrier for environmental web app development \((\text{Swain et al., 2016; Tethys, 2020})\). The Groundwater Level Mapping Tool consists of two main components, the user interface and administrator interface. The administrator interface allows the uploading of data to the app, and the processing of the data to create maps and animations of interpolated

Fig. 7. Semivariance function fitted to experimental variogram.

Fig. 8. Example output of aquifer storage volume change.
groundwater levels. The user interface allows the user to view the time series data and the groundwater level maps and animations created by the administrator. This allows trained users to generate data sets.

The first step in using the app is to use the administrator interface to import groundwater data into the app. Groundwater data are organized by region, and then by aquifer within a region. Region and aquifer boundaries are uploaded as shapefiles and are then converted to GeoJSON objects for display. Groundwater data are imported in a two-step process: first a file of well location data is uploaded. This is a CSV file containing well id, aquifer id, lat/lon coordinates, and ground surface elevation for each file. A second CSV file is then uploaded containing the water level measurements. Each measurement consists of well id, date of measurement, and depth to groundwater. For locations in the USA, the well data can be directly imported from the United States Geological Survey’s NWIS web service (USGS, 2020).

Once the data are imported, the administrator interface can be used to perform spatial and temporal interpolation using the algorithms described above. The user selects a time range and a time interval—for example, 1950 to 2015 with a 5-year interval. The user also selects a set of interpolation options. For each interval, the app first interpolates the time series for each well using the MLR algorithm and then interpolates spatially using kriging. This process results in a set of groundwater elevation rasters, one for each time interval. These rasters are stored in a netCDF file. Raster algebra is used to create two additional sets of rasters: depth to groundwater and drawdown. The drawdown rasters represent the deviation from the groundwater levels at the beginning of the selected time interval. Finally, the app calculates the volume between the sequences of water level rasters and multiplies this volume by a storage coefficient to compute a cumulative change in water storage change vs time curve as shown in Fig. 8. This entire process can then be repeated for other aquifers or for the same aquifer but with different time intervals, interpolation options, etc.

Once the data are imported and the interpolation is complete, the user interface for the web app can be used to explore and visualize the results (Fig. 9). On the left, the user selects the region, aquifer, data type, and raster animation. The wells and the selected rasters are shown in the map window on the right. The maps are supported using the Leaflet JavaScript library (Agafonkin, 2019). As the user selects a well, metadata about the well is selected and the time series for the selected well is displayed below the map. A set of controls can also be used to animate the rasters in the selected netCDF file using a Leaflet animation plugin in combination with a THREDDS server (UCAR, 2020). Aquifer storage change time series are displayed below the individual well time series (not shown in figure).

3. Results and discussion

3.1. Testing of MLR in Cedar Valley, UT

We tested the method of time series extension and imputation using...
MLR detailed in Section 2.1 using a set of ten wells from the Cedar Valley Aquifer. Each well contained data from 1980 to 2015. The locations of these 10 wells was shown earlier in Fig. 2. We used the measurements taken from 1980 to 1995 to train the model, and then compared the model predictions to the actual data taken after 1995 to test the model. Both the training and testing datasets include significant peaks and valleys in the data and cannot be predicted easily using classical time series analysis methods. The testing dataset also includes values far outside the scope of the training dataset. We used the MLR method of time series extension to make predictions at each well, and then compared against a naïve prediction, where the groundwater was assumed to remain constant after 1995, and a linear least squares prediction (Jackson et al., 2019; Roberts et al., 2018). Figs. 10, 11, & 12, show results typical of these model predictions, with the training data (green), the measured data (solid red), the MLR prediction (dashed red), the naïve prediction (dashed blue), and the least squares prediction (dashed yellow) for each well. The figures show the model predictions both for the training period and for the prediction period. The figures also include data available prior to the training period for context, this is labeled “testing data” with a solid red line, but these data were only used for comparison in the prediction period.

Table 1 shows that the MLR method outperformed the naïve and least squares estimation methods for each well (with the exception of Well 374744113055001). Table 1 shows the root-mean-square error (RMSE) value for each of the ten wells in the Cedar Valley study area for the MLR, naïve, and least squares prediction methods. The MLR method decreased the RMSE value by an average of 57% from the naïve method, and 68% from the least squares method.

We also tested the method of time series extension using MLR with correlated wells against Kriging spatial interpolation, using a jackknifing approach (Quenouille, 1949). We used the same ten wells in the Cedar Valley Aquifer for this comparison. We estimated the depth to groundwater on December 31, 2014 at each testing well by implementing the PCHIP time series interpolation and then Kriging spatial interpolation, omitting the measured depth at the testing well from the interpolation. The depth to groundwater on December 31, 2014 was then estimated at each testing well by implementing the MLR technique, using data from 1985 to 1995 as the training set, and then estimating twenty years of groundwater depths from 1995 to 2015. Table 2 presented the results of both of these estimates, the actual measurement, and the % error for each estimate for each of the ten tested wells.

The MLR produced reasonably accurate results, with seven of the ten tested wells exhibiting less than 5% error, and nine of the wells exhibiting less than 10% error. The errors are significantly smaller for the MLR estimate than for the Kriging estimate with the exception of W-2, where the error is practically the same, and W-7, where the MLR error is greater than Kriging. Kriging produces a better estimate at this well for two reasons: W-7 is quite close to W-6 and W-8 (see Fig. 2), which decreases the variance of the Kriging interpolation; and the time series for W-7 contains only four points in its training dataset from 1985 to 1995 as shown in Fig. 14, which decreases the accuracy of the MLR method. Considering these factors, it is unsurprising that Kriging outperformed MLR in this instance. These results demonstrate that in most cases where
Environmental Modelling and Software 131 (2020) 104782

S.W. Evans et al.

Environmental Modelling and Software 131 (2020) 104782

10

data are available from a different time period than desired, it is more accurate to interpolate temporally using MLR to reference other wells rather than to interpolate spatially. This is significant, since some researchers using different temporal interpolation methods have previously concluded that “spatial structure was a little bit stronger than temporal structure” (Ahmadi and Sedghamiz, 2007). In reality, the MLR method uses both as it builds the model using data from other wells (which brings in spatial information) rather than just the time series from the target well. However, one of the limitations of our method relative to Kriging is that it does not produce an uncertainty estimate.

3.2. Testing of MLR in Ogallala Aquifer, Texas

We performed an additional test of the method of time series extension by MLR with correlated wells using 467 wells located in the Ogallala Aquifer in the Texas Panhandle, USA, each well with data from 1960 to 2010 (Fig. 13). This aquifer is unconfined and recharge to the aquifer is via snowmelt and rainwater. We selected this area for testing because of the large amount of data available for hundreds of wells in the aquifer.

We divided the time series observations from each of these wells into a training set from 1960 to 1995, and a testing set from 1995 to 2010. In turn, we used the MLR method to predict the values of depth to water table at each well from 1995 to 2010. We compared these predictions to the actual values of the testing dataset. Figs. 14 and 15, and Fig. 16 show some of the results of this prediction with the training data (green), the measured data (solid red), the MLR prediction (dashed red), the naïve prediction (dashed blue), and the least squares prediction (dashed yellow) for each well. In some cases, where data were not available (e.g., Figs. 15 and 16) data were predicted both prior to and post of the training period. There were no measured data for comparison for predictions in the periods prior to the training data.

As shown in the preceding figures, the MLR method generally outperformed the naïve and least squares methods. Fig. 14 shows that the

<table>
<thead>
<tr>
<th>Map ID</th>
<th>Well ID</th>
<th>Measured Value</th>
<th>MLR Estimate</th>
<th>Kriging Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Depth to GW (ft)</td>
<td>Depth to GW (ft)</td>
<td>% error</td>
</tr>
<tr>
<td>W-1</td>
<td>Well 373236113111401</td>
<td>−47.6</td>
<td>−49.9</td>
<td>(4.8%)</td>
</tr>
<tr>
<td>W-2</td>
<td>Well 373509113101101</td>
<td>−90.6</td>
<td>−92.1</td>
<td>(1.7%)</td>
</tr>
<tr>
<td>W-3</td>
<td>Well 373542113122401</td>
<td>−103.7</td>
<td>−101.4</td>
<td>(2.2%)</td>
</tr>
<tr>
<td>W-4</td>
<td>Well 374105113085001</td>
<td>−74</td>
<td>−79.1</td>
<td>(6.9%)</td>
</tr>
<tr>
<td>W-5</td>
<td>Well 374132113063601</td>
<td>−93.5</td>
<td>−96.2</td>
<td>(2.9%)</td>
</tr>
<tr>
<td>W-6</td>
<td>Well 374304113052901</td>
<td>−109.7</td>
<td>−113.8</td>
<td>(3.7%)</td>
</tr>
<tr>
<td>W-7</td>
<td>Well 374423113053301</td>
<td>−75.6</td>
<td>−60</td>
<td>(20.6%)</td>
</tr>
<tr>
<td>W-8</td>
<td>Well 374423113053401</td>
<td>−67</td>
<td>−67.2</td>
<td>(0.3%)</td>
</tr>
<tr>
<td>W-9</td>
<td>Well 374744113055001</td>
<td>−18.6</td>
<td>−20.2</td>
<td>(8.6%)</td>
</tr>
<tr>
<td>W-10</td>
<td>Well 374745113022901</td>
<td>−68.9</td>
<td>−69.4</td>
<td>(0.7%)</td>
</tr>
</tbody>
</table>

Fig. 13. Wells in the Ogallala aquifer in the Texas Panhandle.
MLR method was able to capture a period of decrease in water table elevation following a period of increase, this trend was not predicted using either the naïve or least squares methods. Fig. 15 shows that the MLR model was able to correctly predict a variation from a basically constant decrease in water levels from 1950 to 1990, while the least-squares method simply continued along the same linear trend line.

Fig. 16 also demonstrates the MLR method’s ability to correctly model changes in aquifer depletion rate.

We compared the results of the time series models using the Range Normalized Root-Mean-Square Error (NRMSE) method (Jackson et al., 2019; Roberts et al., 2018). Table 3 shows the mean and median NRMSE values for the 467 tested wells.

Fig. 17 shows a box and whisker plot of the NRMSE values for the 467 wells, with the MLR Model in blue, the Naïve Prediction in orange, and the Least Squares Prediction in grey. Overall, the MLR Model exhibited the best results.

The MLR method outperformed both the naïve and least squares method in 314 of the 467 tested wells (67%). In those cases where the naïve prediction outperformed the MLR method, it was generally by a small margin, as shown in Fig. 18. In this case, the NRMSE value of the MLR model was 0.033, while that of the Naïve prediction was 0.026, both of which are acceptably small errors. For the majority of these wells, the measured data followed a linear, constant trend, well suited to the naïve prediction method, though even in these cases the MLR method performed well. Fig. 18 shows an example of this situation, where the data in the testing period are essentially constant.

We tested the MLR method with correlated wells for accuracy against Kriging spatial interpolation, using a jackknifing approach with the 407 wells in the Ogallala Aquifer. We selected these wells because they each contained time series data from 1960 to 2010, which enabled comparison of the estimates to actual measured values. We estimated the depth to groundwater on December 31, 2009 at each testing well by implementing the PCHIP and then Kriging interpolation, omitting the measured depth at the testing well from the interpolation. We then estimated the depth to groundwater on December 31, 2009 at each testing well by implementing the MLR technique, using data from 1960 to 1995 as the training set, and then estimating fifteen years of groundwater depths from 1995 to 2010. We compared the estimated depth to water table obtained from both of these methods to the measured value for the testing well and computed the percent absolute error. Fig. 19 shows a box and whisker plot of the percent absolute error for both the MLR (blue) and Kriging (orange) methods for these 407 testing wells.

Table 4 shows the mean and median absolute percent error, and Table 5 shows the mean and median absolute error (ft) for the Kriging and MLR estimates for the 407 wells in the Ogallala Aquifer. The MLR method outperformed the Kriging method significantly. Out of the 407 tested wells, 340 estimates (80%) yielded lower error using the MLR method, while 67 (20%) yielded lower error using Kriging. Over 80% of the MLR estimated depths were within ten feet of the actual measurements, which is quite accurate, considering the average depth to groundwater in this area is approximately 200 feet.

3.3. Testing of aquifer storage change volume estimation

We tested the method of aquifer storage change calculation (Section 2.4) by comparing calculated results with the results from various studies for the Cedar Valley Aquifer in Southern Utah, USA. This aquifer has recently experienced land subsidence, the opening of fissures, and some damage to infrastructure because of over pumping of the aquifer (Inkenbrandt et al., 2014). One difficulty these studies faced was the development of an accurate water budget to estimate the storage change of the aquifer. For example, in the USGS conceptual water budget for the year 2000, aquifer recharge was estimated as 42,000 acre-ft/yr, while discharge was estimated at 38,000 acre-ft/yr, a 4000 acre-ft/yr surplus. This estimated surplus is in direct conflict with observed drawdown of wells in the aquifer, as noted in the USGS report (Brooks and Mason, 2005). The USGS also estimated storage change in the aquifer using a groundwater model, which estimated annual recharge at 27,100 acre-ft/yr and discharge at 34,800 acre-ft/yr, a 7700 acre-ft/yr deficit.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>MLR Model</th>
<th>Naïve Prediction</th>
<th>Least Squares Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.108</td>
<td>0.181</td>
<td>0.250</td>
</tr>
<tr>
<td>Median</td>
<td>0.076</td>
<td>0.145</td>
<td>0.211</td>
</tr>
</tbody>
</table>
This estimate seems more logical, as it matches the observed trends in lowering groundwater levels in the area. The USGS also estimated recharge in the aquifer using a Chloride mass-balance method, which yielded an estimated recharge of 20,800 acre-ft/yr. Finally, the USGS employed a Basin Characterization Model (BCM) to estimate precipitation recharge throughout the basin. This model estimated recharge at 20,900 acre-ft/yr. The results of these studies, all carried out by the USGS, demonstrate the difficult nature of aquifer storage quantification (Heilweil and Brooks, 2010; Inkenbrandt et al., 2014; Thomas and Taylor, 1946). Based upon these reports and other studies, the Utah Division of Water Rights concluded that “the average annual groundwater deficit is probably about 7600 acre-feet” over the last fifteen years for the Cedar Valley aquifer (D. Jones, 2016). Inkenbrandt et al. (2014) concluded that the deficit in 2000 was 10,700 acre-ft.

We used the Groundwater Level Mapping Tool to calculate aquifer storage change in the Cedar Valley aquifer for the period from 2000 to 2015. This corresponds to the same 15-year period studied by Jones (2016). In this calculation, we used a specific yield value of 0.1, which was used by Bjorklund et al. (1978) and by Inkenbrandt et al. (2014). We

![Fig. 17. Box and whisker plot of NRMSE values in the Ogallala aquifer.](image)

![Fig. 18. Time series model for well 1003903.](image)

![Fig. 19. Box and whisker plot of percent estimation error in Ogallala aquifer.](image)

Table 4

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>MLR Method</th>
<th>Kriging Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Absolute Percent Error</td>
<td>4.4%</td>
<td>29.4%</td>
</tr>
<tr>
<td>Median Absolute Percent Error</td>
<td>2.1%</td>
<td>12.8%</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>MLR Method</th>
<th>Kriging Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Absolute Error</td>
<td>6.2 ft</td>
<td>26.4 ft</td>
</tr>
<tr>
<td>Median Absolute Error</td>
<td>3.5 ft</td>
<td>19.1 ft</td>
</tr>
</tbody>
</table>
followed the procedures outlined in Section 2.1 (MLR) for temporal interpolation, Section 2.3 (Kriging) for spatial interpolation, and Section 2.4 for the final calculation. Using these methods, we calculated the change in aquifer storage volume in the Cedar Valley aquifer between March 2000 and March 2015. Fig. 20 presents a time series plot of aquifer depletion as calculated by the Groundwater Level Mapping Tool. Over the 15-year period, the aquifer was depleted by 125,000 acre-feet, an average of 8300 acre-feet per year. The value calculated by the Groundwater Level Mapping Tool is comparable to the USGS/Utah Division of Water Rights, which calculated a value of 7600 acre-feet per year. The Groundwater Level Mapping Tool estimated a water budget deficit of 11,500 acre-feet per year for the year 2000, comparable to the Inkenbrandt et al. (2014) estimate of 10,700 acre-feet per year.

Using a similar approach, we also used the Groundwater Level Mapping Tool to calculate changes in water storage in the Beryl Enterprise Aquifer in southern Utah. This is an unconfined aquifer that has been subjected to substantial storage depletion in recent decades. We compared the results against a study prepared by the United States Geological Survey in cooperation with the Utah Department of Natural Resources, which concluded that between 1937 and 1978, the aquifer lost between 1.3 and 1.5 million acre-feet of storage (Mower and Sandberg, 1982). Using the USGS estimate of 0.2 as the specific yield, the Groundwater Level Mapping Tool calculated a storage loss of 1.45 million acre-feet between 1937 and 1978. The Utah Division of Water Rights estimated that the annual depletion rate of this aquifer around the year 2012 was approximately 65,000 acre-feet per year (K. L. Jones, 2012). The Groundwater Level Mapping Tool estimates this rate as 66,000 acre-feet per year, again using the USGS specific yield estimate of 0.2.

4. Conclusions

The objective of this work was to develop a technique to improve the accuracy of groundwater level mapping using a simple, low-cost approach. Our technique makes it possible to generate historical and current water level maps where multi-linear regression exploit correlations between wells in an aquifer to impute water levels during periods where measurements are sparse. We have shown that the technique is superior to simpler forms of imputation and is generally superior to spatial Kriging.

We have encapsulated this algorithm in a web-based tool that enables water managers to make informed decisions and implement wise management plans and regulations regarding the sustainable use of aquifers. The tool is ideally suited for developing countries where the data imputation method can be particularly helpful in dealing with sparse groundwater data and the open source nature of the tool lowers the barrier for implementation.

The novel contributions of this work are as follows:

1) **Open source web-based assessment tool.** The open source Groundwater Level Mapping Tool allows water managers and other decision makers to quickly and easily compute and view trends in aquifer storage level changes. We developed the application to generate maps and animations of groundwater levels and drawdown which can be used to inform decision makers, enabling them to identify areas of concern and develop groundwater management plans to ensure the long-term sustainability of aquifers. The application is useful for calculating an estimate of aquifer storage change over time, typically a painstaking, laborious task.

2) **Groundwater data imputation method.** To improve the accuracy of this application in mapping and quantifying groundwater, we developed a method of data imputation, MLR with correlated wells. This method of temporally extrapolating recorded data to unsampled time periods, using correlated data from other wells, was used in conjunction with Kriging spatial interpolation to create maps of groundwater levels at specified time steps. This method outperformed the typical Kriging spatial interpolation method using only wells with measured data at the specified time steps, yielding more accurate maps of groundwater levels.

3) **Aquifer storage change estimate.** This tool and the accompanying methods provide a simple approach for estimating historical changes in aquifer storage. We found that the Groundwater Level Mapping Tool’s automated method yielded results comparable to several detailed USGS studies in Utah’s Cedar Valley and Beryl-Enterprise area. Once computed these aquifer storage change estimates can be used by managers to validate water budget estimates and determine if an aquifer is being managed in a sustainable fashion.

One limitation of the tool is that it currently uses a single average storage coefficient for each aquifer. While this is a reasonable approach for most aquifers, there are cases where sufficient data exists to characterize the spatial variation of the storage coefficient within an aquifer. In such cases, the tool could be easily modified to include a grid of storage coefficients as input. At this point in time, we have only validated the aquifer storage results using unconfined aquifers. More studies are needed to test the method on confined aquifer systems. Another limitation of the tool is that does not currently produce an uncertainty estimate.

5. Software availability

The “Groundwater Level Mapping Tool”, which implements the methods described here, can be found at https://tethys2.byu.edu/apps/gw/ (username: guest, password: guest). The tools included in this paper are all open source. The web app code can be found at: http://github.com/stevenwevans2/tethysapp-gw/tree/timeseries. The web app was implemented on Tethys Platform (home page: }
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Aeronautics and Space Administration, USA Grant 80NSSC18K00440. Author contributions to this article are as follows: Steven Evans conducted the bulk of the work and is the primary contributor; Norman L. Jones provided the funding, and coordinated and assisted with the research; Daniel P. Ames and E. James Nelson provided computational and technical support; and Gus P. Williams provided statistical support for this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2020.104782.

References

Kumar, V., 2006. Kriging of groundwater levels—a case study. J. Spatial Hydrol. 6 (1).