AgasedViz: visualizing groundwater availability of Ogallala Aquifer, USA

Article *in* Environmental Earth Sciences · February 2020

DOI: 10.1007/s12665-020-8851-6

4 authors:


Tommy Dang
Texas Tech University


Pham Van Vung
Texas Tech University


Huyen Nguyen
Texas Tech University


Ngan V. T. Nguyen
Texas Tech University

Environmental Earth Sciences (2020) 79:110
https://doi.org/10.1007/s12665-020-8851-6

THEMATIC ISSUE

AgasedViz: visualizing groundwater availability of Ogallala Aquifer, USA

Tommy Dang[1] · **Vung Pham**[2] · **Huyen N. Nguyen**[2] · **Ngan V. T. Nguyen**[2]

Received: 31 December 2018 / Accepted: 11 February 2020

Abstract

Water plays an immensely important role in human life, from household to agriculture and industrial development. Due to the heavy utilization and growing demands of water, the availability of water has become a high priority. Besides sources of surface water, such as streams, lakes, and rivers, aquifers are a significant source of water on Earth through the essential groundwater supply. In the United States, the Ogallala Aquifer, a primary geologic unit of the High Plains Aquifer System, is under massive exploitation, and the water level has decreased substantially. Analysis of the saturated thickness of this aquifer is urgently essential. Thus, an interactive visual analytics tool is necessary to enable users to visualize water availability and variation overtime at multiple locations within the region. Our interactive analytic tool carries out this by first retrieving and processing the data derived from the saturated thickness of groundwater using sensors integrated into the wells. Subsequently, a visualization consisting of a contour map and a time series heatmap is created based on the sensor data, to capture the trends and patterns, such as underground water distribution, spatial and temporal changes, and sudden decreases or increases of the water level. The visual components allow users to explore observational data, which organize the visual structure in supporting making an inference to gain insights. insights. This approach can be extended to apply for any geographic areas for water-level monitoring and controlling.

Keywords Saturated thickness · Ogallala Aquifer · Contour map · Heatmap · Spatio-temporal visualization · Visual analytics

Introduction

With the rapid population growth and social development, demands for water are rising at a fast rate. As surface sources of water, such as streams or rivers, become unavailable in arid areas, the viable alternative for these regions is groundwater contained in aquifers. These aquifers are saturated with moving water that can be extracted and brought up to the surface. Schloss (2000) define saturated thickness as the vertical thickness of the hydrogeologically defined aquifer, in which the pore spaces between the rock constructing the aquifer are filled with water. Many aquifers have a recharge zone, which is an area for new water to replenish the aquifer once the previous volume of water was relocated.

The Ogallala Aquifer, spreading from western Texas to South Dakota of the United States, has been used as the primary source of water for households and agriculture activities in the region. According to the Ogallala Aquifer Initiative 2011 Report (NRCS 2011) by the Natural Resources Conservation Service, United States Department

This article is part of a Topical Collection in Environmental Earth Sciences on "Visual Data Exploration", guest edited by Karsten Rink, Roxana Bujack, Stefan Jänicke, and Dirk Zeckzer.

✉ Tommy Dang
tommy.dang@ttu.edu

Vung Pham
vung.pham@ttu.edu

Huyen N. Nguyen
huyen.nguyen@ttu.edu

Ngan V. T. Nguyen
ngan.v.t.nguyen@ttu.edu

[1] Department of Computer Science, Texas Tech University, Box 43104, Lubbock, TX 79409-3104, USA

[2] Department of Computer Science, Texas Tech University, Lubbock, USA

Published online: 26 February 2020

of Agriculture (NRCS), approximately 27% of the irrigated area in the United States covers the Ogallala Aquifer and this system supplies drinking water to 82% of the population living inside the region. Besides, from the report of Closas and Molle (2016), the extensive exploitation of groundwater from the Ogallala Aquifer has gone beyond the natural recharge capability, leading to a significant reduction of the water level. The constant excessive use of water has put the aquifer under a severe threat of depletion. This problem could be alleviated by having a system that can help to monitor water level and, therefore, manage water use. In the United States, the United States Geological Survey[1] (USGS) has collaborated with data providers to monitor groundwater levels, which were authorized by the SECURE Water Act (2010) by US Department of the Interior in 2009. This groundwater network focuses on establishing a national network of wells at many levels, including Federal, State, and local. In 2018, a total of 1115 groundwater-level sites were added to the National Groundwater Monitoring Network (NGWMN), resulted in over 7000-well long-term water-level sites monitored in this system. More advanced methods, such as deploying drones to monitor the groundwater/surface-water interaction, are in the planning process, as presented in the USGS Groundwater monitoring report (United States Geological Survey 2018).

To evaluate the water availability of the aquifer system, we need to consider various hydro-geological parameters. An aquifer can have low porosity, thus a little amount of water available. An aquifer with low hydraulic conductivity may hammer the water extraction process. This study serves as a first proxy for water availability at a given location/area via saturated thickness. The data for saturated thickness is acquired from the sensors attached to wells in the Southern High Plains Aquifer of Texas. The visual prototype is transferable to other locations.

Our contributions in this paper are:

- This study proposes a new approach for visualizing water availability and allows users to convey the underground water distribution and information of water supply locations with ease. The overall visual component is integrated on a telescoping contour map, enabling direct interaction.
- An interactive visualization tool is implemented, called *AgasedViz*[2] (Analyzing Groundwater Availability from Social and Empirical Data) (Dang et al. 2019). Our spatial–temporal visualization supports a full range of interactions, such as filtering, brushing, and linking, and details on demand.
- *AgasedViz* is applied to the Ogallala Aquifer to explore and analyze the underlying characteristic of that region. The case studies are conducted with five participants to evaluate the usefulness and effectiveness of *AgasedViz*. Various droughts, such as the one in 2012, are highlighted and discussed.

The rest of this paper is structured as follows. We describe related work in Sect. 2. Section 3 provides a high-level overview of our visualization system. Next, we describe in detail the major stages and associated visual components in our system in Sect. 4. Through our interactions with subject-matter experts, we highlight some use cases in Sect. 5. Finally, we conclude the paper and discuss future work.

Related works

Monitoring groundwater resources has witnessed advancement in recent years, alongside with hydrological measurement development (Diersch 1998; Thakur 2017). Software platforms as monitoring systems provide a variety of features for visualization and analysis tasks, addressing the need for water resource management frameworks. Advanced tools can be in the form of a web-based analytical platform (Umwelt and GmbH 2015), which provides real-time reporting for urban groundwater monitoring. Analysis outcome can be exported in 2D and 3D geological environment (Serpescu et al. 2013), utilizing the sedimentary media analysis platform (Gogu et al. 2011). Such systems are built with connecting servers, which can be a drawback if the servers are down in case of traffic congestion. For the task of visualization and analysis, in this case, an interactive and lightweight application is required.

Changes in groundwater levels and saturated thickness in aquifers have been studied in prior literature. Water-level monitoring programs based on observation wells can enhance the interpretation of trends observed in water levels (Taylor and Alley 2002), in which a groundwater monitoring network may provide insights for strategic planning and decision-making (Krishna Thakur 2012). In the central High Plains aquifer, the saturated thickness declines of about 100 feet in several areas between the Arkansas and Cimarron Rivers (McMahon 2001). In parts of Kansas, the Ogallala Aquifer has lost up to 60% of the saturated thickness (Cross); therefore, raising awareness on this matter is crucially important.

Uddameri et al. (2017) implemented visualization and analysis, including both locations of the well and its water depth, based on the combination of ArcGIS 10.5 (Environmental Systems Research Institute 2016)[9] with R integration (R Core Team 2013), with terrestrial water storage data obtained from Gravity Recovery and Climate Experiment

[1] https://www.usgs.gov/.

[2] Online demonstration of *AgasedViz*: https://AgasedViz.github.io.

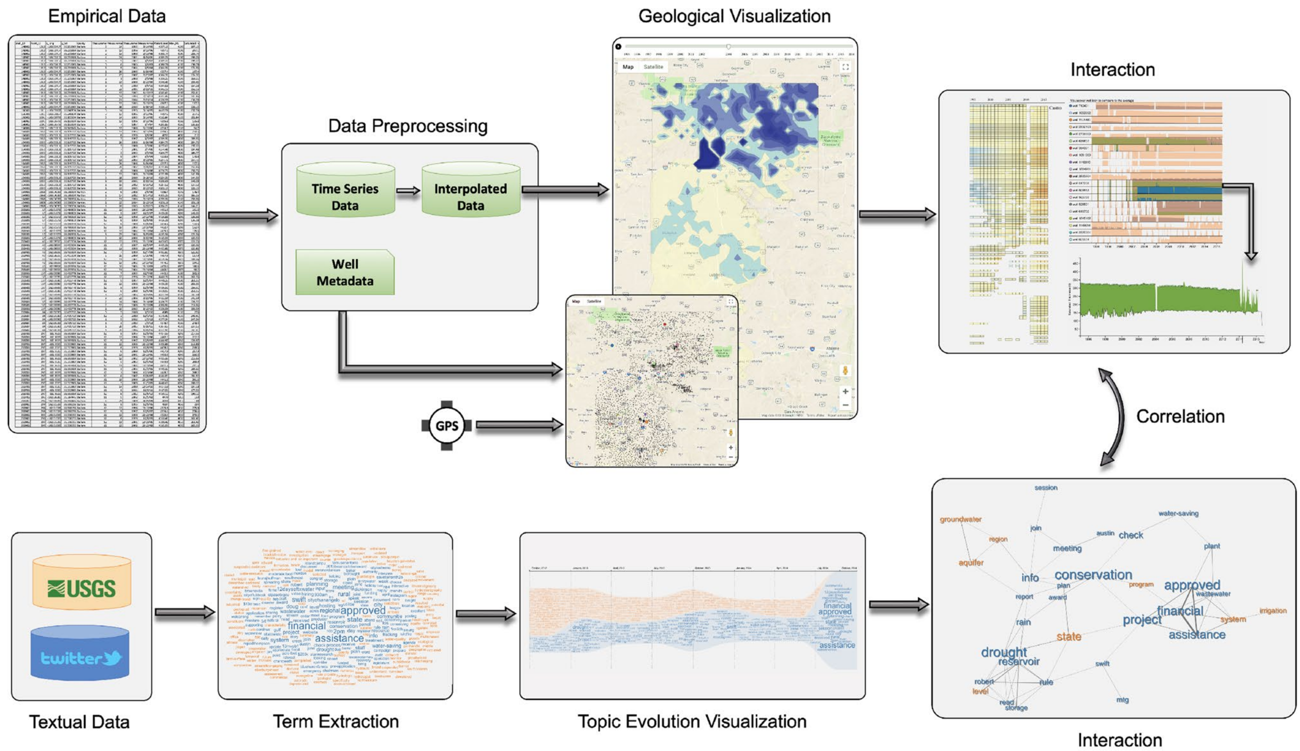

Fig. 1 Major stages in *AgasedViz* visualization: data preprocessing, data visualization, and interactions

satellite systems. Previously, we have presented a web prototype for analyzing groundwater at the Ogallala Aquifer (Dang et al. 2017). The application allows users to select the time range and geographic area of interest to visualize the saturated thickness. However, it is difficult to track the changes in a larger geographic area. Here, in this application, we plot the saturated thickness difference, which makes the comparison task more manageable. Besides empirical data, social data are incorporated to provide comprehensive insight. Moreover, animated features offer a quick overview of changes over longer time intervals.

System overview

This section describes stages of *AgasedViz* in detail. *AgasedViz* is developed using JavaScript and, in particular, the D3.js library developed by Bostock et al. (2011). The primary goal of *AgasedViz* is to create an interactive visual analytics tool that demonstrates water availability distribution and saturated thickness of groundwater in the region of the Ogallala Aquifer. The water availability is presented with spatial and temporal perspectives. The tool enables users to interact and investigate the trends and water variation via historical data and track changes spatially and chronologically. The changes in water supply, such as sudden decreases or increases, are presented and can be applied for further inference.

AgasedViz implements the analysis tasks below to meet the primary goal, based on representative analysis task types (Amar et al. 2005):

- **T1**: Provide an overview of water availability spatially and chronologically from well data sources.
- **T2**: Retrieve and display details on demand. In particular, users can select a well on the map to view its historical data or search for a term (i.e., “drought”) to pull and display all related social media discussion.
- **T3**: Characterize data distribution, applied for both quantitative and qualitative data.
- **T4**: Detect anomalous occurrences.
- **T5**: Present development of topics from concern in groundwater availability over time.
- **T6**: Display relationships of topics dynamically.
- **T7**: Correlate empirical and social data.

Figure 1 depicts an overview of *AgasedViz*. The *AgasedViz* architecture consists of two components: empirical data analytics and social data analytics.

In each workflow, input data are preprocessed and employed for visualization. The visualization and interactions aid users to gain deep insights into the groundwater availability (visualization task **T1**). After conducting a user study, the final stage involves taking the results from both workflows for correlation and confirmation of the correctness of the information provided (visualization task **T7**). The workflows and analysis of corresponding features developed to satisfy the tasks are discussed in more detail in Sect. 4.

The workflow for empirical data includes raw data acquisition and its preprocessing stage, hydro-geological visualization, and interaction. Raw data are preprocessed and interpolated for a continuous visualization. The details for data processing are demonstrated in Sect. 4.1.1. After that, the hydro-geological visualization is constructed, which is the integration of spatio-temporal data into a web-based service map. The visualization showcases empirical data presentation with levels of granularity and anomaly points (visualization task **T3, T4**). Further interactions are provided, such as sliding along the timeline for time tracking, customizing visual arrangement, and mouseover for details on demand (visualization task **T2**) support users to have multiple perspectives to the empirical data. Data collection, visualization, and interaction of the workflow for empirical data are described in great details in Sect. 4.1.

On the other hand, the workflow for social data is initiated with the term extraction process of raw input. An interactive visualization is created to demonstrate the development and evolution of topics during the years (visualization task **T5**). Besides, we propose a force-directed network to show the relationships among the keywords, allowing users to explore the underlying connections between topics (visualization task **T6**). The details of the components included in this workflow are presented in Sect. 4.2.

Visual components

AgasedViz consists of two main sections, as illustrated in Fig. 2, in which Box A contains the workflow for social data, and Box B presents the workflow for empirical data.

Workflow for empirical data

Data preprocessing

Raw data of saturated thickness are collected from the Southern High Plains Aquifer of Texas, from 5241 unique wells, in the period of 22 years from 1995 to 2016 (Uddameri et al. 2017). The sensors, which collect the raw data, are integrated into the wells at multiple locations. The saturated thickness t_s in an unconfined aquifer is calculated by:

$$t_{\mathrm{s}} = d_{\mathrm{ls}} - w - e_{\mathrm{b}},$$

where d_{ls} is land surface datum, w is water level, and e_{b} is elevation at the bottom of the aquifer. d_{ls} and e_{b} are measured with respect to mean sea level, while w is below ground surface.

After this acquisition step, monthly data are then preprocessed (Uddameri et al. 2017). This process results in a set of time series data and well metadata. Time series data are then employed to acquire interpolated data, both of which are then utilized as input for visualization. At every single time step of the time series data, *AgasedViz* interpolates to generate the data in a grid format for the contour visualization of the saturated thickness levels over the monitoring area. The interpolation step is required, because the contour visualization is continuous, while the distribution of the wells is discrete. Furthermore, wells are not equally distributed in the geographical surface of the region. Some grid cells contain one or more wells, while others do not have any. The interpolation gives average saturated thickness values for the cells that contain one or more wells and *null* otherwise. Specifically, the followings are the steps of the interpolation process:

1. Calculate the whole coverage area (the boundary of the longitudes and latitudes of all the wells).
2. Divide this whole area into a grid of size $N \times N$ rectangle cells (N is set to *40* in this application via experiment to balance between visual quality and performance).
3. Calculate the saturated thickness of a cell as the average saturated thickness values of all the wells in that cell or *null* in case there is no well in the cell.
4. Generate the contour layers from the grid data using the D3 library contour algorithm.[3] This library computes contour polygons by applying marching squares (Maple 2003) to a rectangular array of numeric values.

Hydro-geological visualization

The hydro-geological visualization contains the contour map, heatmap, and control panel, from left to right. The distribution of water availability is presented with various options based on the derived information types selected.

The control panel

The control panel at the top right corner of Box B in Fig. 2 allows users to adjust the main view settings:

Color scale: This is used to update the color encodings of both the contour map and heatmap. The two options are *Absolute value* and *Difference*, describing different

[3] https://github.com/d3/d3-contour.

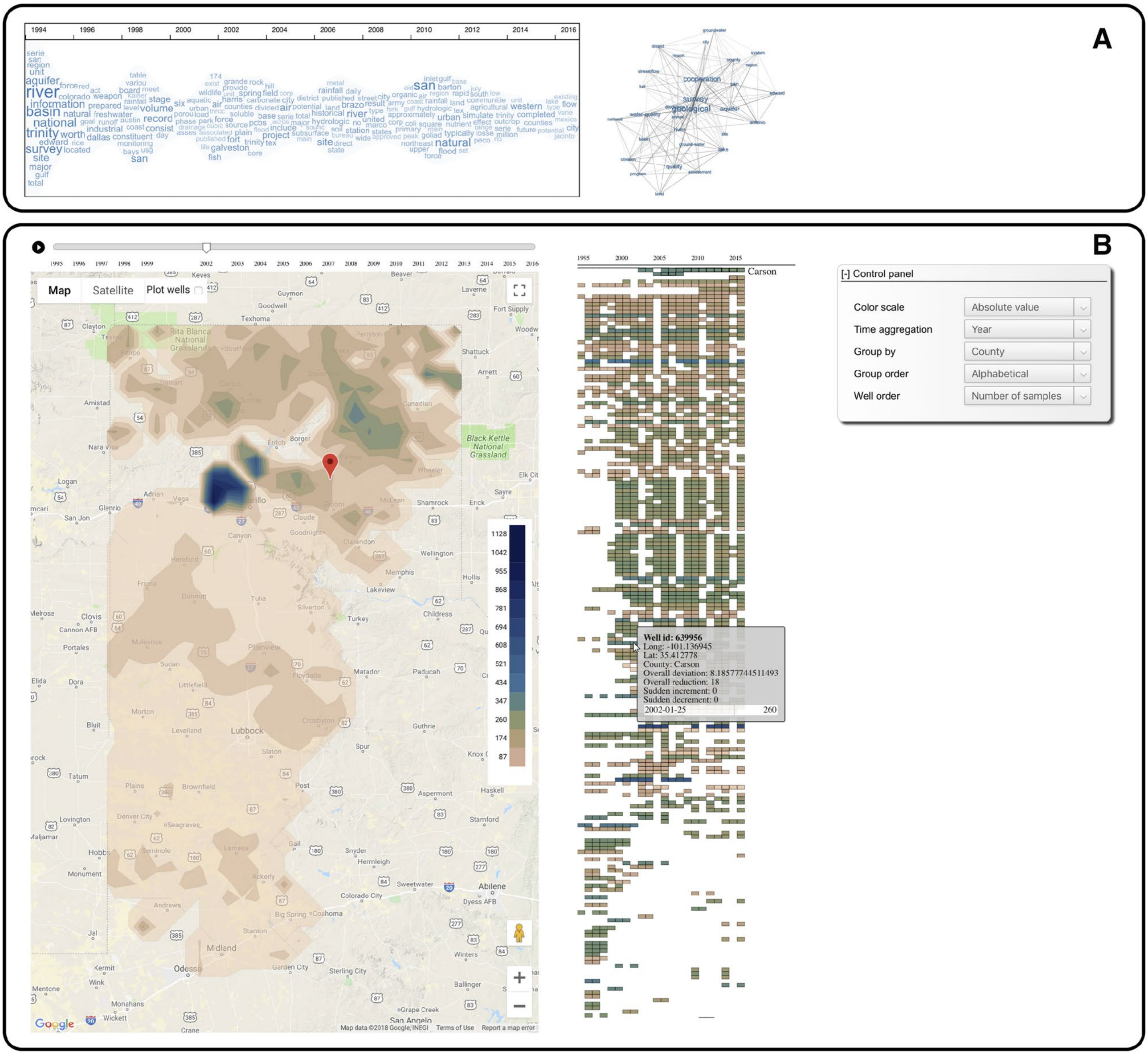

Fig. 2 User interface of *AgasedViz*, **a** workflow for social data and **b** workflow for empirical data

perspectives on data distribution (visualization task **T3**); in both cases, the values are measured by feet and can be converted to meters.

Time aggregation: This settings provide selection for grouping data by *Month* or *Year*, by which the data cells are illustrated in a timely manner (visualization task **T1**). The value for each data cell is calculated from the average measurements in a month or a year, respectively.

Group by: The current selection is by *County*. This means that each well, in the form of a data cell, is classified by the county it locates in (visualization task **T1**).

Group order: This option allows users to choose a method to arrange the groups. From the *County* selection of *Group by*, the viable options for this preference are *Alphabetical*, which sorts the groups by the name of county alphabetically, and *Number of wells*, in which the groups are arranged by the number of wells within each group.

Well order: This preference enables users to rearrange the display of well order. The sorting criteria can be chosen from:

- *Alphabetical*: The wells are sorted by well-identity numbers (Well ID).

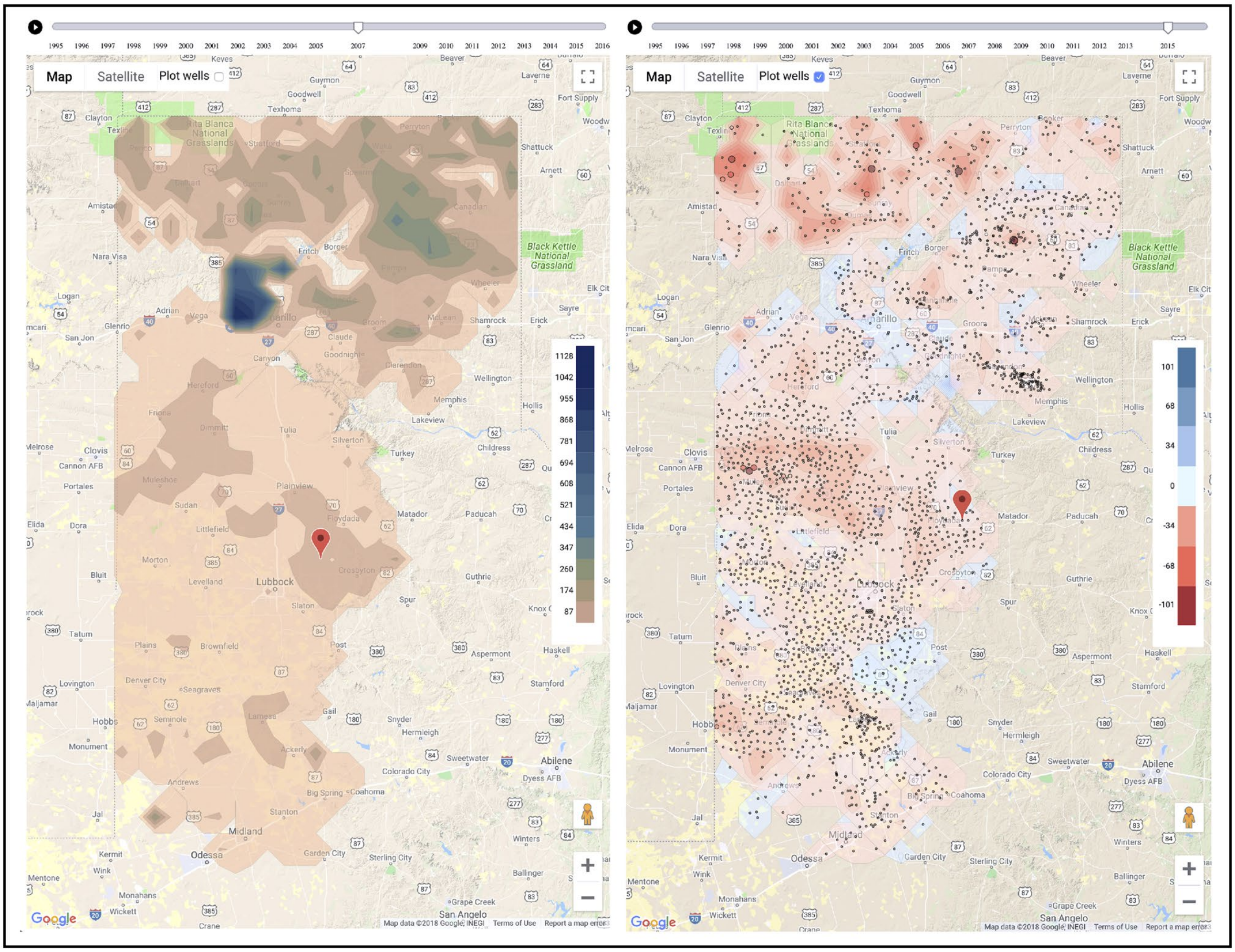

Fig. 3 The contour map: (left) absolute saturated thickness and (right) difference from the mean value

- *Number of samples*: Number of valid sensor readings over the entire period of time.
- *Sudden increment*: Sudden increment yields the changes in groundwater thickness are substantial following an upward trend. This option allows the well data to be presented in descending order, start with the well, which has the maximum sudden increment throughout the entire period (visualization task **T4**).
- *Sudden decrement*: This is similar to *Sudden increment*, but with a downward trend (visualization task **T4**).
- *Standard deviation*: Standard deviation is a metric that expresses how elements of a group are spread out from the mean value. The well that has the groundwater thickness that varies the most is placed at the top, followed by the ones that have less variance.
- *Overall reduction*: The reduction is calculated by the difference of saturated thickness at the beginning until at the end of the monitored period. The topmost well has the biggest reduction.

The contour map

The contour map is on the left side of Box B in Fig. 2, which provides a high-level overview of groundwater availability for the selected area. The input data for building the contour map are wells' measurements and the time series data; when these measurements are unavailable, then the interpolated data are applied. Interpolation of wells in the area forms the bounding of each region. The Google map platform[4] serves as a base for plotting geographical areas with a telescoping functionality (Fig. 3).

The contour map is color encoded by the measured and interpolated saturated thickness to give a high-level overview of the area (visualization task **T1**). The *Absolute value* option presents the absolute thickness of groundwater, ranging from the sand color to darker blue. The *Difference* option

[4] https://cloud.google.com/maps-platform/.

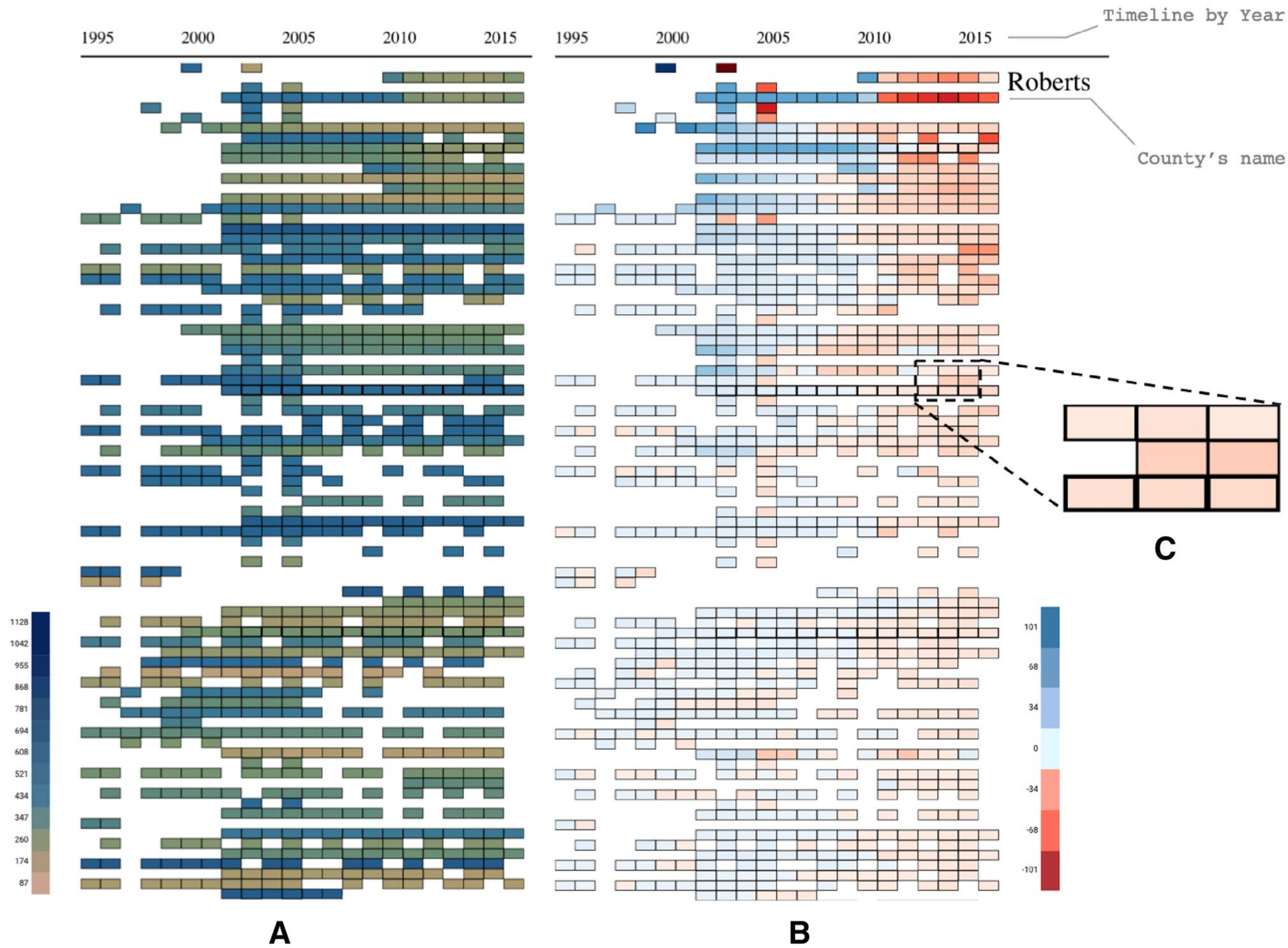

Fig. 4 The time series heatmaps for Roberts county in Texas: **a** the heatmap in the scale of *Absolute value*, with corresponding color scale on the left, **b** the heatmap in the scale of *Difference to average*, with corresponding color scale on the right, and **c** magnified presentation of consecutive cells, with the difference in thickness of cells indicating on number of samples

compares the current saturated thickness of a location to its average saturated thickness over the entire observed period 1995–2016: blue for above the average and red for below the average. The wells can be plotted on the map, and their details (well metadata and historical readings) can be displayed via a pop-up window on user request (visualization task **T2**).

On top of the contour map is a time slider to set the year for the contours. An autoplay function (the “Play” button) animates how these saturated thickness contours change over the years.

The heatmap

A close-up heatmap is presented in Fig. 4. Each row in the heatmap presents monthly measurements for an individual well by a series of cells along the horizontal timeline. Based on the color scale, whether it is for absolute value or difference from average, the color of each cell represents the corresponding value. The cell’s border thickness indicates the number of samples for each cell per month: the more samples a data cell represents, the thicker its black boundary is. Empty cells, shown by blank spaces, represent missing data—no measurements are available. From these visual components, users can quickly identify over-sampled and under-sampled wells. Moreover, via the heatmap, users can view the groundwater value changes over time and captivate anomalies via the unusual representation inside the heatmap (visualization task **T4**). In panel (b) of Fig. 4, most of the rows are transforming from blue to red as we move from left to right. This indicates that wells in Roberts county are drying up.

The display time preferences are *Month* and *Year*. For the *Month/Year* option, each cell represents the mean value of all the samples in that month or year. The *Year* preference gives a concise summary of the data, while the *Month* preference illustrates data with a finer granularity.

The visual components inside the heatmap can be rearranged by aggregation with several hierarchy criteria. The default **Group by** preference is *County*, which means that the highest level of presentation for grouping is under the name for each county. The groups are then ordered under **Group order** setting, in *Alphabetical* or *Number of wells*. Inside each group, the *Well order* customization provides various options for data exploration: with each criterion, the well that best satisfies the criteria is put at the top order in each group. The criteria details are described in Sect. ***Control Panel***.

User interaction

AgasedViz enables users to interact with the visual components through mouse manipulation: mouse over, mouse click, and mouse drag.

Contour map interaction

Users can track changes spatially and chronologically on the contour map via the timeline by clicking on the *Autoplay* button at the end of the time axis. This interaction allows users to discover and track the changes in groundwater availability throughout the entire time period (visualization task **T1**).

The display mode of the map (Map or Satellite) can be selected by a mouse click. This feature is supported by the Google map platform, allowing the user to switch back and forth between map and satellite views. The map view showcases a clear presentation of the viewing area, while the satellite view, which is made of images (satellite or aerial images) of the actual surface, offers natural features of the region such as water, plants, or bare ground. While the former provides a pleasant and clear presentation, the latter offers the hydro-geological visualization more relatable to users. When users click the option *Plot wells*, the locations of all the wells are shown within the map. Providing that *Plot wells* are enabled when users mouse over a location of a well on the contour map, the detailed information of the well is presented by a pop-up tooltip. This detailed information contains ID of the well, longitude, latitude, county, overall deviation, overall reduction, sudden increment, and sudden decrement (visualization task **T2**). The dates of the sampling process are also indicated for every well, with the corresponding measurement for each day.

Heatmap interaction

Detailed information is provided when a user mouses over an individual cell in the heatmap, with similar content as in the interaction of contour map. Therefore, with such information, when users mouse over a row, the changes in groundwater supply can be captured. When users select a cell on the heatmap, its corresponding location on the contour map is indicated (visualization task **T2**).

Control panel interaction

Mouse clicks are performed within the control panel to select the criteria. Dropdown lists are provided for the selection.

Workflow for social data

Data preprocessing

The input data are collected from the publications of the USGS and the Twitter social network.[5] The input data for saturated thickness are from the period 1995 to 2016. The USGS scientific reports for Texas Water data are filtered and extracted from the same period for the purpose of correlation. The Twitter data are acquired from the Texas Water Development Board's official account in the period 2012–2016. After the data acquisition stage, stop words, which are commonly encountered in texts without dependence on a particular topic (Uysal and Gunal 2014), are removed. In our case, stop words are from general context such as pronouns, articles, prepositions, etc., and outside general context, but are irrelevant in water context, such as webpage links. On the other hand, examples for words in the groundwater context are "drought", "aquifer", "irrigation", hashtags such as "txwater" (Texas Water), or an abbreviation such as "twdb" (Texas Water Development Board). Meanwhile, terms regarding non-groundwater contexts can be "projects", "rural", "financial", describing other aspects in community's concern.

The list reports have been use can be find in usgs website under publication section. We attract data by downloading and copy report its abstract content manually from provided filter in usgs website.[6] The Twitter posts collect from Texas Water Development Board account from January 1 2012 to December 31 2016 base on query input in Twitter website.[7]

Due to the complex of content, we do not decide the collection of words relate to groundwater. However, by calculating the sudden attention and the frequency of the term, the word stream visualization display only the keywords which have high probability relating to social attention at that specific time range. Additionally, the account which we chose

[5] https://twitter.com/.

[6] https://www.usgs.gov/centers/tx-water/publications?logstash-usgs-pw%3Apalladium_root_publication_type=Report&logstash-usgs-pw%3Apalladium_root_topics=&logstash-usgs-pw%3Apalladium_root_publication_year_date=&sort=&page=0.

[7] https://twitter.com/search?f=tweets&q=from%3Atwdb%20since%3A2012-01-01%20until%3A2016-12-31&src=typd.

is the official account of government; meanwhile, almost all posts relating to water development such as investment in water treatment, water development project, meeting, or water-related news.

USGS reports contain geographical framework, water-quality assessments, and other technical evaluation. Twitter data provide more general information with a broader scope, including investment and financial support for water-related projects, with corresponding personnel and events, or the socioeconomic aspects of such projects. In terms of keyword extraction, we retrieve the title and abstract of each report for the USGS publications and from the entire content of tweet for the Twitter data. The frequency of a term's occurrences in a text report is then calculated. Terms that have a large number of occurrences are seemed to have more significant meaning than the ones that have fewer appearances.

Visualization

The topic evolution visualization

The visualization focuses on the presentation of the topic evolution chronologically, which is the left component of Box A in Fig. 2. The overall shape is a stream of words, spreading over the timeline (visualization task **T5**), utilizing the technique by Dang et al. (2019), called WordStream. The technique can be expanded to include context for the textual content (Nguyen and Dang 2019). The thickness of the stream depicts the amount of information at a particular time point. Each term is associated with the timestamp of the report which it belongs to and put in the corresponding location related to the time point (visualization task **T3**). The more frequency a term accumulated, the bigger its font size is.

Inside the stream, *AgasedViz* provides two options for word presentation: by frequency and by the sudden attention. With frequency option: words that have higher frequency throughout the time span are more likely to be essential terms. Hence, they are emphasized by bigger font size, compared to less frequent words. On the other hand, with sudden attention option, the significance of a word is defined its sudden appearance along the timeline: if a word is absent in the previous time step but the current time step, the attention will be drawn on this recently arise term (visualization task **T4**). Regarding this criterion, the terms which have sudden appearances are then sorted by frequency in descending order. From the selected preference, the top words are chosen to put into the stream first. Collision checking is implemented to make sure that there are no overlaps between any two words; therefore, terms are properly arranged inside the stream. Detail information of every word is provided on interaction, which is covered in Sect. 4.2.3.

The network

The network describes the relationships among the essential terms in the corpus, which is shown in the right of Box A in Fig. 2. These essential terms are extracted by taking the top frequency terms for the visualization. The network is based on a force-directed graph, in which the vertices demonstrate the terms, while the edges present the connection between two terms (visualization task **T6**). For a pair of two words, the more co-occurrences they have in all the reports, the stronger their relationship is, and therefore thicker edges shown in the network. The font size of each word is proportional to its accumulated frequency in the corpus.

User interaction

When a user mouses over an individual term in the visualization, the specific details of each word are provided by the tooltip, including the content of the term, frequency in the corresponding report, and the year of the report (visualization task **T2**). The tooltip also lists the sources and titles of the reports containing the term in that year.

The network allows users to zoom in or zoom out for a closer or broader look at the relationships. Users also can rearrange the network components' position by dragging any vertices to examine further the connections, regardless of the possible clutter view due to a large number of nodes and links. The absolute location of a vertex may vary as users drag it around for exploring the relationships, but the associated vertices to it and the corresponding edges are fully preserved. When users click on the link (edge) between any two terms, a tooltip window pops up with the corresponding content from the social data, providing context for the relationship of these terms.

Stories and discussion

We gathered qualitative responses and feedback about *AgasedViz* from five experts, one post doc researcher and one associate professor (both have at least 10 years of experience in earth sciences), one is field technician supervisor of the High Plains Underground Water Conservation District, and the other two are soil scientists at the Natural Resources Conservation Service (NRCS). All of them have vast experience and enthusiasm in protecting and preserving groundwater supply. The study begins with the introduction of *AgasedViz*, and then, the experts are free to use *AgasedViz* before providing feedback and comments.

AgasedViz provides dual perspectives on the situational awareness of saturated thickness. Empirical data reflect precise measurements with statistics from the field. Social data from publications, scientific reports, and social media data

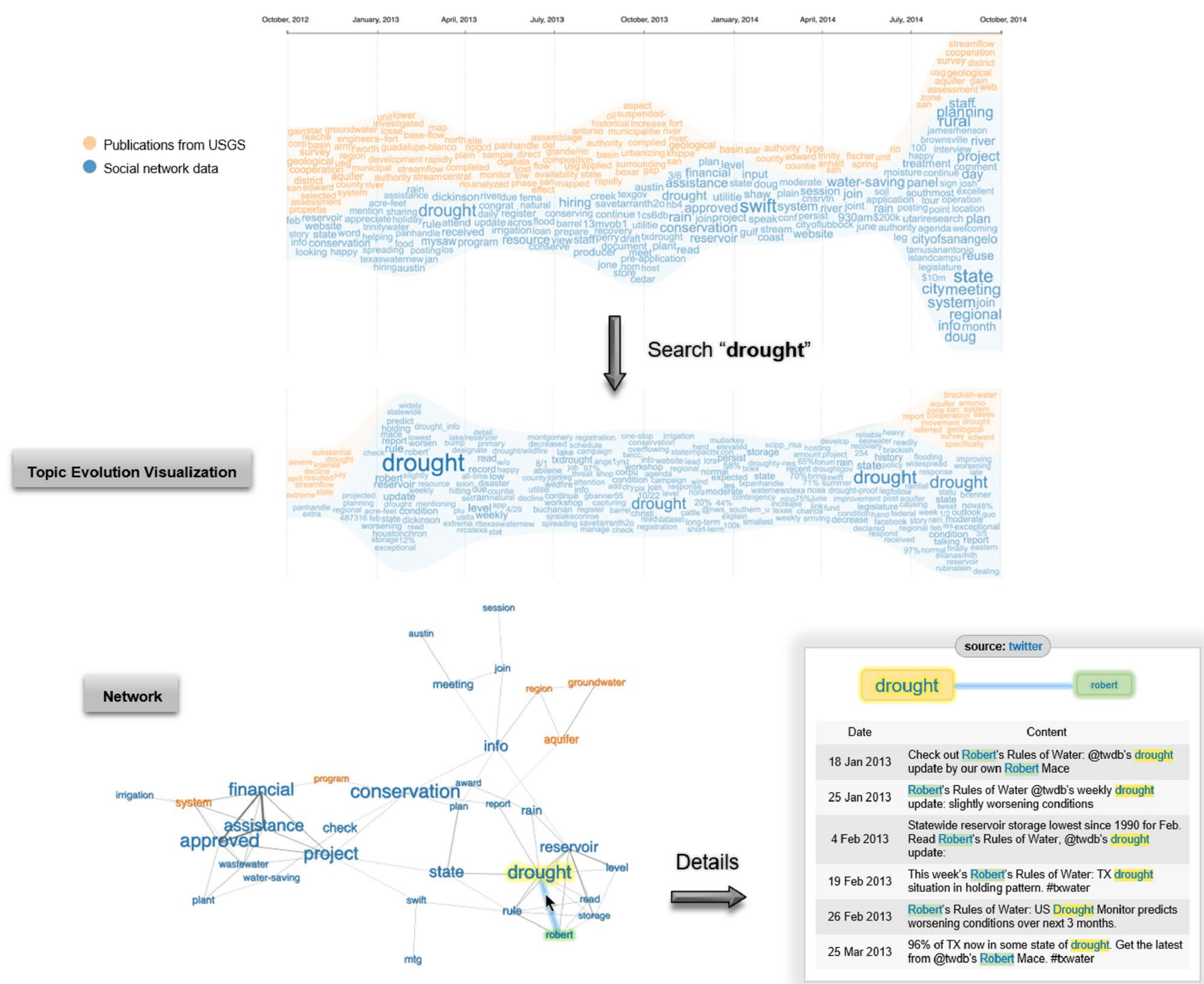

Fig. 5 Empirical data visualization: workflow and interactions (filtering and details on demand)

bring another point of view to the matter. While the publications and scientific reports discuss geographical frameworks in the context of technical evaluation and assessment, social media data convey "the wisdom of crowd". Since limited availability and data deficiency can lead to reliability issues, gathering multiple data sources can help to alleviate this problem by complementary information to one another source, bringing a more comprehensive view to users.

In terms of empirical data, the well readings are collected regularly in January annually. At some months, there might be multiple measurements for one well in a single month. The visualization begins with the hydro-geological visualization, whose components are linked graphically through consistent color encoding to each display option: *Absolute value*, *Difference from average*, and *Difference from the previous step*, applying on both the contour map and the heatmap. Besides the telescoping feature on the map, the interactivity also includes the time slider for comparison between timesteps. From the animated contour map, saturated thickness corresponds a decreasing trend over the past 20 years. This can be achieved from the "Overall reduction" option, which shows that, in more recent years, the color dark red is dominating, meaning that the water level is far below average. The considerable decline can result from an imbalance between discharge and recharge (as discussed by McGuire et al. 2012). The recharge can come from precipitation, but with a little amount, especially in the wintertime, the NRCS soil scientists stated. Meanwhile, the discharge in the region consists of irrigation and intensive farming activities in the North and middle areas, and petroleum production in the South. The massive amount of feedlots in the region contains corn production to feed animals, besides crops of cotton or dairies. On the other hand, the oil and gas industry makes heavy consumption of water in the petroleum refinery and other production processes.

Regarding social data, Fig. 5 presents the visualization from the social point of view. The topic evolution visualization is divided into two layers, according to two sources: publications from USGS and social media data from Twitter. The Twitter account in focus is the Texas Water Development Board, reflecting the latest update from water news for the region. As one can see from the visualization, one of the most concerned terms is "drought", in January 2013. Users can input this term into the search box to narrow down the analysis. The result is described in a similar structure as topic evolution, where the words inside each stream are now the ones that have co-occurrence with "drought" in the textual data. The relationship of highly important words is depicted in the force-directed network, which also supports interactive operations. When a user hovers the mouse over the link between any two terms (nodes), the link and associated terms are highlighted. A tooltip window pops up with details of these two terms in the context of twitter posts. In this case, we choose the pair "drought" and "robert". As shown at the bottom of Fig. 5, the pop-up window shows all tweets containing these two terms. The corresponding phrases are "Robert's Rules of Water" and "Robert Mace". Robert's Rules of Water is the weekly update of drought conditions in Texas, which is now known as Water Weekly. The earlier title of the update is named after Dr. Robert Mace, former Deputy Executive Administrator for Water Science and Conservation at Texas Water Development Board. Robert's Rules of Water is posted via Twitter about the drought condition, hence the co-occurrence of the two terms. In this way, the topic evolution and network visualization make up a continuation of the story, emphasizing the perspective from the social data.

The two sources of data can be utilized to build a broad view of the matter. The contour map illustrates a significant water reduction in the year 2012—2013, indicating the severe droughts in the previous year. On the other hand, in Fig. 5, the term "drought" is strongly emphasized, especially in the period of early 2013, which represents an important topic at this time. This visualization from social data was found to correlate positively with the empirical data. According to the scientists, due to the excessive water consumption, the Ogallala is depleting and may run dry in the near future.

Overlapping between multiple aquifers over certain areas can be an obstacle for characterizing an aquifer's condition. Under a part of Ogallala Aquifer, there is Dockum Aquifer, which is a minor aquifer located in the northwest part of Texas. For this study, we need to get the data from the wells that receive only water from Ogallala Aquifer. In terms of deep pumping depths, the Dockum Aquifer produces poor water quality and low yields (Bradley 2003). According to the post doc researcher that we interviewed, Dockum is deeper than Ogallala, and the sandstone layer can be considered as a barrier for separation between them.

The Ogallala Aquifer underlies the Great Plains region in the United States, particularly in the High Plains states Texas, New Mexico, Oklahoma, Kansas, Colorado, and Nebraska (Global Change Course, I.S.U. 2003). Generally, the flow of Ogallala Aquifer has the direction from the North to the South of the region. Therefore, if the Northern areas have droughts or low levels of water availability, the condition of the Southern areas will be negatively affected. However, as one expert has pointed out that the inspecting part of the Ogallala Aquifer in South Plains, Texas has localized characteristics, which means that the water supply of a specific local area is not affected by the nearby regions.

That said, the data from empirical measurements can be affected by various factors, whether it is the geographical characteristic of that particular region or subjective human impacts on measuring. Lack of personnel or human resources can also contribute to the data availability issue. By utilizing the social data from reports, publications, and especially social media data, more perspectives are put on the analysis, and, hence, provide a comprehensive view of the water availability situation.

This study, however, is subject to several limitations that could be addressed in future research. As the visualization comprises of both empirical data and social data, the defects originate from a variety of possible causes. First, the data source achieved from field measurement for empirical data has missing values, which we were unable to control or track systematically. The lack of data leads to a discrete characteristic of the visualization at several points, which, therefore, affect the analysis and reasoning process. We have applied the interpolation method where applicable to construct new data points, in an attempt to fill in these gaps. Second, regarding the social media data from Twitter, we retrieved the data from the account of the Texas Water Development Board, which can be subject to the availability of the news reporting policy of the Board. This issue can be alleviated in future research with an extended range of accounts on social media, so that the data availability will not depend on solely one institution.

Conclusion

In this paper, we present *AgasedViz*, an interactive, visual analytic tool for demonstrating saturated thickness from two different sources (with different characteristics): empirical data and social data. The empirical data present the real numerical statistics collected from the field, while the social data convey other related aspects among the community, such as projects to be implemented, and financial resources for groundwater plans. Data availability issues in both empirical and social data contribute to the current limitations, which will be addressed in future research with extended scope of data sources. The visualization employs interactive geological maps and textual content presentation, showing a strong correlation between them and, therefore, helping users to raise awareness for protecting and preserving water. The use cases are conducted with High Plains Underground Water Conservation District and Southern Great Plains Soil Survey Region 9 at Lubbock, Texas. For future work, we will extend the application with a prediction feature, which will estimate the groundwater availability at a particular time in the future, on different granularity levels: from seasons or months to discrete dates. This feature will bring more insights into the drought status and its consequences later on. A more extensive study to validate the usefulness and usability of our tool is another direction for our future work. The user study can be extended for general users to gather feedback on visual representation and ease of use, how *AgasedViz* can be improved to deliver more specific and accurate information on saturated thickness and groundwater availability to a broad user community.

Acknowledgements The authors gratefully thank Mr. Keith Whitworth, the field technician supervisor of the High Plains Underground Water Conservation District, for his valuable comments on the groundwater characteristic. We also want to thank Mr. Thomas (Craig) Byrd, Mr. Todd Carr, and Mr. Alain Basurco, soil scientists at Southern Great Plains Soil Survey Region 9 at Lubbock, Texas, for their supports. The authors would like to thank the journal of Environmental Earth Sciences reviewers for their insightful comments that significantly improved the article and acknowledged the inclusion of some of their ideas.

References

Amar R, Eagan J, Stasko J (2005) Low-level components of analytic activity in information visualization. In: Information visualization, 2005. INFOVIS 2005. IEEE symposium on, pp 111–117

Bostock M, Ogievetsky V, Heer J (2011) D3 data-driven documents. IEEE Trans Vis Comput Graph 17(12):2301–2309

Bradley RG, Kalaswad S (2003) The groundwater resources of the Dockum aquifer in Texas. http://www.twdb.texas.gov/publications/reports/numbered_reports/doc/R359/Report%20359%20Dockum%20Final.pdf

Closas A, Molle F (2016) Groundwater governance in America

Cross J A path forward for Oklahoma's groundwater monitoring and assessment. http://water.okstate.edu/news-events/news/water-center-news/a-path-forward-for-oklahoma2019s-groundwater-monitoring-and-assessment

Dang T, Nguyen LH, Karim A, Uddameri V (2017) STOAViz: visualizing saturated thickness of Ogallala aquifer. In: Rink K, Middel A, Zeckzer D, Bujack R (eds) Workshop on visualisation in environmental sciences (EnvirVis). The Eurographics Association. https://doi.org/10.2312/envirvis.20171102

Dang T, Nguyen HN, Pham V (2019a) WordStream: interactive visualization for topic evolution. In: Johansson J, Sadlo F, Marai GE (eds) EuroVis 2019—short papers. The Eurographics Association. https://doi.org/10.2312/evs.20191178

Dang T, Pham V, Nguyen HN, Nguyen NV (2019b) AgasedViz: visualizing groundwater availability of Ogallala Aquifer. Environmental Earth Sciences, USA

Diersch H (1998) Treatment of free surfaces in 2D and 3D groundwater modeling. Math Geol 2:17–43

Environmental Systems Research Institute (2016) ArcGIS Desktop: Release 10.5. ESRI, Redlands, California, CA. http://www.esri.com/

Global Change Course I.S.U. (2003) The Ogallala aquifer depletion. http://www.meteor.iastate.edu/gccourse/issues/society/ogallala/ogallala.html

Gogu R, Velasco V, Vázquez-Suñe E, Gaitanaru D, Chitu Z, Bica I (2011) Sedimentary media analysis platform for groundwater modeling in urban areas. Advances in the research of aquatic environment. Springer, Berlin, pp 489–496

Krishna Thakur J, Gossel W, Weiss H, Wycisk P (2012) Optimizing a long-term groundwater monitoring network using geostatistical methods. IAHS-AISH Publication, UK, pp 133–138

Maple C (2003) Geometric design and space planning using the marching squares and marching cube algorithms. In: 2003 international conference on geometric modeling and graphics, 2003. Proceedings, pp 90–95. https://doi.org/10.1109/GMAG.2003.1219671

McGuire VL, Lund KD, Densmore BK (2012) Saturated thickness and water in storage in the high plains aquifer, 2009, and water-level changes and changes in water in storage in the high plains aquifer, 1980 to 1995, 1995 to 2000, 2000 to 2005, and 2005 to 2009. US Geol Surv Sci Investig Rep 5177:28

McMahon PB (2001) Vertical gradients in water chemistry in the central high plains aquifer, southwestern Kansas and Oklahoma panhandle, 1999. Water Resour Investig Rep 1:4028

Nguyen H, Dang T (2019) Eqsa: earthquake situational analytics from social media. arXiv:1910.08881 (arXiv preprint)

NRCS (2011) Ogallala aquifer initiative 2011 report. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1048827.pdf

Office of Congressional and Legislative Affairs (2010) SECURE Water Act|US Department of the interior. https://www.doi.gov/ocl/hearings/111/SECUREWaterAct_031610

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/ **(ISBN 3-900051-07-0)**

Schloss JARWB (2000) Current saturated thickness. http://www.kgs.ku.edu/HighPlains/atlas/atcst.htm

Serpescu I, Radu E, Gogu CR, Boukhemacha MA, Gaitanaru D, Bica I (2013) 3D geological model of Bucharest city quaternary deposits. In: Proceedings of 13th SGEM geoconference on science and technology in geology exploration and mining, vol 2, pp 1–8

Taylor CJ, Alley WM (2002) Ground-water-level monitoring and the importance of long-term water-level data, 1217-2002. US Geological Survey

Thakur JK (2017) Hydrogeological modeling for improving groundwater monitoring network and strategies. Appl Water Sci 7(6):3223–3240. https://doi.org/10.1007/s13201-016-0469-1

Umwelt U, GmbH IZ (2015) Environmental monitoring, analysis and interactive visualization—EMAIV. http://uizentrum.de/en/environmental-monitoring-analysis-and-interactive-visualization-emaiv/

United States Geological Survey (2018) Groundwater monitoring. https://www.usgs.gov/water-resources/groundwater-and-streamflow-information/groundwater-monitoring

Uysal AK, Gunal S (2014) The impact of preprocessing on text classification. Inf Process Manag 50(1):104–112

Venki Uddameri, Karim Abdullah EU, Srivastava P (2017) Sensitivity of wells in a large groundwater monitoring network and its evaluation using grace satellite derived information. Sensitivity analysis in earth observation. Elsevier, Oxford, pp 235–256

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