Gelman Site 1,4-Dioxane Groundwater Contamination Plume Modeling and Forecasting

by

Yifan Luo

A practicum submitted in partial fulfillment of

the requirements for the degree of

Master of Science (Environment and Sustainability)

in the University of Michigan

April 2022

Practicum Advisors:

Dr. Andrew Gronewold

Dr. Derek Van Berkel

Abstract

Groundwater systems are intrinsically heterogeneous with dynamic spatio-temporal patterns, which cause significant challenges in quantifying and mapping their complex processes. However, accurate forecasting of regional groundwater contamination is commonly needed to identify its spatio-temporal dynamic that helps the public anticipate the timing and severity of potential groundwater quality issues and possibly serve as an early warning system. This study focuses on modeling a plume of 1,4-dioxane originating from the Gelman site beneath the city of Ann Arbor, Michigan. It proposed a novel methodology to consider the spatially and temporally irregular and uncertain nature of groundwater contamination data to analyze the historical trends of dioxane concentration and predict its transportation:

1. A random forest interpolation model was deployed to fill in or extend fragmented time series data gaps among all the monitoring wells;
2. Mann-Kendall test was applied to evaluate the trend of dioxane concentrations at various wells;
3. An automated time series machine learning (AutoTS) package was utilized to predict the best future values forecasts; and
4. An R-based Shiny web application was designed to allow visualization and quantification of dioxane contamination analytical data.

This research introduced a novel framework for filling spatial and temporal data sampling gaps in groundwater contamination to offer an effective and promising way to predict future plume concentration and spatial distribution.

Acknowledgment

First and foremost, this paper and the research would not have been possible without the exceptional support of my advisors, Dr. Drew Gronewold and Dr. Derek Van Berkel, for their patient guidance, enthusiastic encouragement, and valuable critiques. I would also like to acknowledge my colleague Dr. Yiwen Mei for his assistance and partnership in completing this project and all other SEAS Hydro research group members. And I wish to thank Jacqueline Courteau from Scio Township for reaching out to us with the golden opportunity to do this wonderful project and Roger Rayle from Coalition for Action on Remediation of Dioxane for his help in providing data, context, and support. I thank Ziliang Zhang and Lindsey Quint for their continued friendship and support during the course of this project. Lastly, and most importantly, I would like to thank my family for being a source of love, encouragement, and support throughout the work.

Table of contents

Introduction

Groundwater, which is in aquifers below the surface of the Earth, is one of the most important natural resources, serving as a major source for agricultural irrigation, industrial processes, mining, and drinking water. Of the world's total water supply of about 332.5 million cubic miles of water, 30.1% of the earth's freshwater consists of groundwater, while only 1.2% consists of surface water in lakes, rivers, and streams (Gleick 1993).

Unfortunately, groundwater is susceptible to pollutants. Groundwater contamination occurs when man-made products such as gasoline, oil, road salts, and chemicals get into the groundwater and cause it to become unsafe for human use. Nowadays, groundwater pollution threatens human and ecosystem health in many regions around the globe. The rapid flow of the groundwater through focused recharge is known to transmit short-lived pollutants into carbonate aquifers endangering the quality of groundwaters where one-quarter of the world’s population lives (Hartmann et al. 2021). Therefore, it is difficult and expensive to clean up when groundwater becomes contaminated. At the Gelman Sciences Inc. (now Pall Life Sciences) site in Ann Arbor, Michigan, wastewater containing 1,4-dioxane was discharged into unlined seepage lagoons and spray irrigated across a 15-acre field from 1966 to 1986. Efforts to remediate 1,4-dioxane emanating from the Gelman Site in Washtenaw County, Michigan, have been underway for the past 35 years. Although substantial quantities of dioxane have been removed from the aquifer system through pump-and-treat operations, numerous factors make complete aquifer restoration technically infeasible at the Gelman Site.

1,4-dioxane is classified as a Group B2 probable human carcinogen and causes kidney and liver damage and respiratory issues (USEPA 2006). When released into groundwater, its high miscibility and low retardation factor limit the ability of processes such as sorption to attenuate its concentrations as plumes migrate downward and outward from their source (USEPA 2017). 1,4-dioxane is readily soluble in water but resistant to microbial degradation and adsorption to soil particles (USEPA 2006). Therefore, its widespread use, high water solubility and mobility, and presence in numerous groundwater plumes have become an increasing concern for groundwater use in homes and businesses. Moreover, the glacial aquifer system affected by the plume contamination is highly heterogeneous. Consequently, contaminated plumes of groundwater have moved in various directions and at different depths, making it challenging to predict contaminant movement.

Nevertheless, groundwater aquifers are intrinsically complex and heterogeneous systems with dynamic spatio-temporal patterns, which cause great difficulty in accurately characterizing their changing status over time (Alley and Taylor 2001). Although fresh groundwater is often abundant and extensively used, it is difficult and expensive to accurately quantify and map its complex processes compared to surface water resources. However, reliable regional groundwater contamination concentration predictions are commonly needed to ensure proper groundwater resource management strategies are implemented within a region. Even when data on groundwater properties are available from monitoring wells, these irregularly spaced data are hence referred to as "data points." Point data from monitoring wells are typically sparse and give only a limited sampling of the aquifer's spatial distribution of water levels. It is difficult to piece these segments of data together into a complete picture of the region (Marchant and Bloomfield 2018, Oikonomou et al. 2018). Thus, they are generally not harnessed to their full potential to aid decision-making. Temporarily, these data collected from monitoring wells are often available only at irregular and infrequent intervals, leading to significant gaps in the time series data. Their locations and groundwater-surface levels may also vary significantly throughout an aquifer giving only a limited sampling of the spatial distribution.

Groundwater resources are always impacted by external factors such as pollution, climate change, agricultural irrigation, and other land-management practices. Long-term forecasting is a tool to demonstrate future sustainable water resources planning and management. However, a predictive analysis of spatial and temporal changes in groundwater contamination demands a deep understanding of past trends in contaminant concentration levels in aquifers.

Several techniques have also been developed and used for the temporal interpolation of limited well time-series observations in groundwater studies. The Autoregressive Integrated Moving Average (ARIMA) family of models includes auto-regressive (AR), moving average (MA), auto-regressive moving-average (ARMA), auto-regressive integrated moving-average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), and periodic autoregressive

(PAR). And it has been extensively and repeatedly used for the analysis and forecasting in a wide range of groundwater-related studies (Box and Pierce 1970). For example, a SARIMA model was developed and tested to forecast the groundwater anomalies within the Colorado River Basin (Rahaman et al. 2019). Regression techniques such as linear or logistic regression have also been widely applied in groundwater modeling. For instance, a multiple linear regression (MLR) model was developed for the Osona, Spain, to determine aquifer vulnerability to nitrate pollution in groundwater (Boy-Roura et al. 2013). Rupert (2003) implemented a logistic regression technique to predict the probability of detecting atrazine and desethyl-atrazine in groundwater in Colorado.

A wide range of geostatistical modeling approaches has been implemented to analyze spatial variability and interpolation for groundwater data. At the same time, their purposes, emphases, and capabilities could also be very different. Kriging is a local estimation technique of the best linear unbiased estimator for the unknown values of both spatial and temporal variables. Rouhani and Wackernagel (1990) used intrinsic random functions for space-time kriging to perform temporal interpolation of monthly piezometric data in the Seine River basin in France. It reduces to a spatial estimation problem of interest for variables with observations rich in time but poor in space. Besides, process-based methods that either simulate or take into account physical processes of groundwater movement and pollutant transportation can be used to predict the concentrations of surface-derived solutes in groundwater (Barbash and Voss 2016).

This study aimed to provide an in-depth understanding of contamination level and its spatio-temporal distribution and took account of the spatially and temporally irregular and uncertain nature of groundwater contamination. Based on the historical trends analysis of dioxane concentration, we predicted its future transportation, then identified the areas that need additional attention and provided decision support for the administration of drinking water safety, water environment protection, and emergency response. First, a Mann-Kendall trend test was performed to evaluate the trend of dioxane concentrations at various wells. A random forest model was developed to fill in or extend fragmented time series data gaps among all the monitoring wells. Then, the automated time series machine learning (AutoTS) package was applied for predicting time series values based on comparing thousands of combinations of data preprocessing steps and algorithms to assess which best forecasts future values. Finally, an

R-based web application was designed to allow visualization and quantification of dioxane contamination analytical data. This methodology workflow (Figure 1) separately addresses spatial and temporal variation of groundwater contamination data and produces spatio-temporal forecasting of future plume concentration.

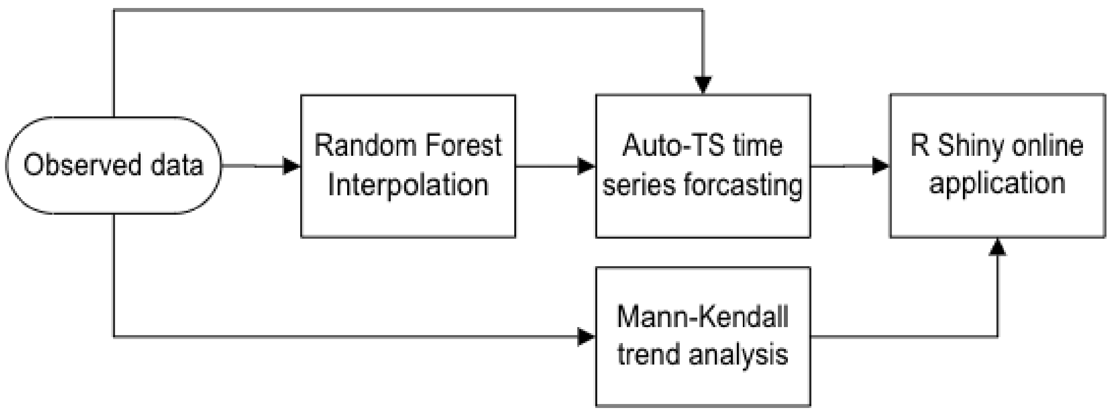
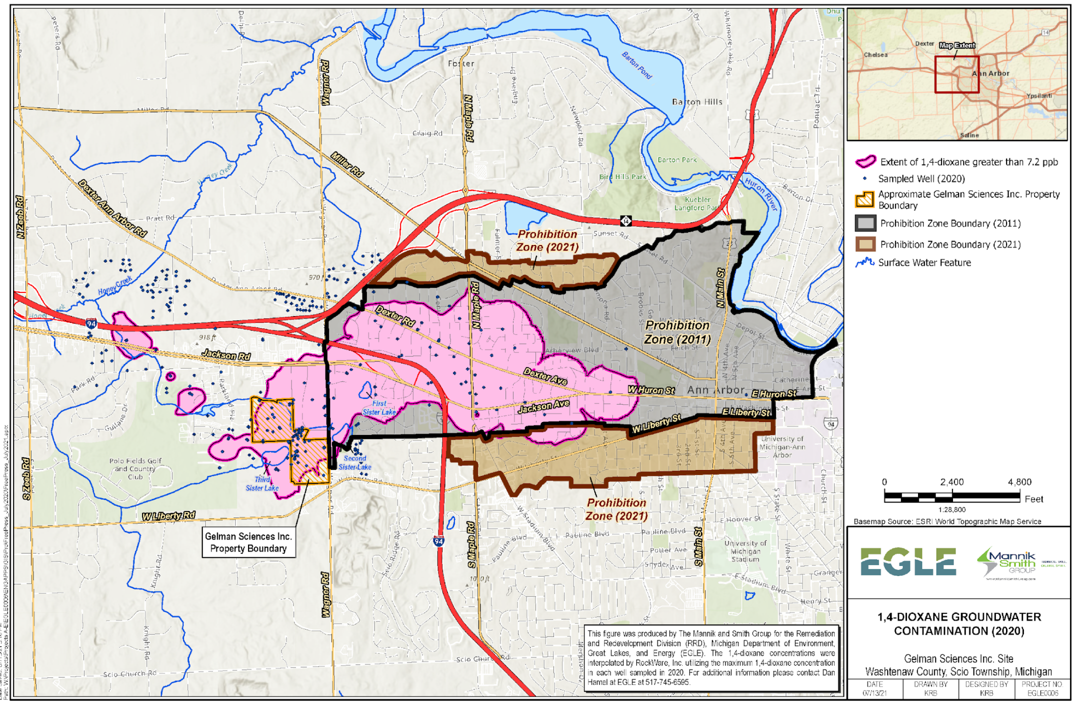


Figure 1: Research methodology workflow

Methods

Description of the study site and data source

The site for this study is located in Washtenaw County, Michigan. The dioxane contamination originated from Gelman Sciences Inc., a former medical filter manufacturer located on Wagner Road across Second Sister Lake (Figure 2). Between 1966 and 1986, Gelman Sciences used 1,4-dioxane in their filter manufacturing process and disposed of wastewater onsite. It was first discovered in Third Sister Lake and nearby water wells in 1985, and then a comprehensive site investigation started later in 1986 (MDEQ 2004). Since then, the plume has continued to migrate west in Scio Township and northeast then east into Ann Arbor, moving towards the Huron River, contaminating local lakes and private drinking water wells. The topography of this area ranges from 940 feet above mean sea level in the vicinity of the Gelman property to the lower area of approximately 850 feet at the Honey Creek. There are currently over 250 monitoring and extraction wells located in and around the plume used to track water levels and dioxane concentrations, which are being used in this study.



The depth to 1,4-Dioxane contamination in groundwater is variable throughout the plume and the depth to contamination is not represented on this 2D plume map.

Figure 2: 2020 Gelman site groundwater contamination map (EGLE 2021)

Cleanup of dioxane-contaminated sites is difficult because of its resistance to natural biodegradation and high mobility (Adams et al. 1994). Remediation usually relies on ex-situ pump-and-treat approaches with advanced oxidation processes, the primary methods employed to treat dioxane-contaminated aquifers. Extracted water containing dioxane is mixed with hydrogen peroxide and ozone, or exposed to ultraviolet light, to break the carbon bonds (Zenker et al. 2003). Despite varying levels of cleanup efforts since the early 1990s, the groundwater contamination has continued to spread into a plume that is now roughly four miles long and one mile wide in a densely populated area and is in some cases venting to surface water. In addition, treated water containing low levels of dioxane and bromate produced in the treatment process is also being discharged to surface water in Scio Township.

Monitoring wells are used to sample the groundwater elevations to help determine groundwater flow direction and monitor the dioxane levels in the water. The monitoring wells at the Gelman site are tested monthly, quarterly, twice a year, or once a year, depending on their location. Gelman conducts the sampling and testing of over 250 monitoring wells following Michigan

Department of Environment, Great Lakes and Energy (EGLE) approved sampling plans. EGLE samples selected wells with Gelman quarterly and submitted the samples to the laboratory for analysis. All data used in this study are available in table format or for download at Scio Residents for Safe Water website: https://sites.google.com/site/srsworg/srsw-org/data or EGLE Gelman Sciences Recent Analytical Data Monitoring Well Results: https://www.michigan.gov/egle/0,9429,7-135-3311_4109_9846_30022-71616--,00.html. Please see EGLE Gelman Sciences, Inc. Site of Contamination Information Page for more information related to this project: https://www.michigan.gov/egle/0,9429,7-135-3311_4109_9846-71595--,00.html.

Mann-Kendall trend analysis

The Mann-Kendall test is a nonparametric test used to identify whether a time series has a monotonic upward or downward trend (Mann 1945, Kendall 1975). The Mann-Kendall test is applied to data for dioxane concentration at each monitoring well located in the plume area for this research. According to the test, H_0 (null hypothesis) assumes that observations do not show any monotonic trend or serial correlation over time, and H_1 (alternative hypothesis) assumes that there exists a consistent increasing or decreasing trend in time series observations. Here, a rejected null hypothesis reveals that "there is a trend in the time series of 1,4-dioxane concentrations in the groundwater". Since it is a nonparametric method, it does not require any assumptions regarding the underlying statistical distribution of the data. This test is also not sensitive to the sampling time interval over which the data are collected. For these reasons, the method is well suited for evaluating concentration trends over time.

Random forest interpolation

In order to fill in or extend fragmented spatial and temporal data gaps, a random forest model was developed to define a continuous function fitting the given values. Random forest regression is an ensemble machine learning method for classification and regression that operates by constructing a multitude of decision trees (Breiman 2001). It is nonparametric, and thus there are no underlying assumptions about the distribution of the errors or the data. Furthermore, it can identify linear and nonlinear statistical relationships between variables for classification and regression objectives (Breiman 2001). It works well with large numbers of predictors to distinguish observed data from synthetic data (Cutler et al. 2007). Random forest predictors can deal with model selection uncertainty, as predictions are based upon a consensus of many models and not simply a single model selected with some measure of goodness of fit. This study used predictors typically found in ecological classification applications, such as geographic variables (e.g., water table depth, elevation, latitude, and longitude), time-series data, and field measurements of 1,4-dioxane concentration.

AutoTS automated time series forecasting tool

In this study, we deployed a novel automated time series (AutoTS) machine learning package (Catlin 2022) to predict time series values based on comparing various time series forecasting techniques with open-source implementations. It is a “black box” which trains multiple time series forecasting models, including ARIMA, FB Prophet, regression, naive, and smoothing methods in just one line of code. Model selection is based on accuracy in model performance in forecasting future values for predictions with the least model error on out-of-sample data. It is to be noted that the groundwater contamination in different locations varies depending on the groundwater flows, water table levels, source spreading, soil properties, topographical characteristics, hydraulic conductivity, etc. And all these factors play an essential role in the performance of the time-series forecasting method. Since there is no best single method that can perform well for any given forecasting situation, a model built based on the historical dioxane concentration data for a particular well location may not provide similar accuracy for other sites. Due to this reason, selecting a single forecasting method as a proposed approach may not be realistic. Thus, a set of methods instead of one must be considered so that the best method can be selected based on their performance on location-specific data. While other statistical model forecasting studies usually use one model to fit all, the deployed AutoTS platform compared a multitude of models and allowed the selection of the most accurate one for each well.

The length of the time series can vary. Still, a statistical forecasting methodology is generally not valid for less than 20 sampling events, and some models even require at least 50 observations for accurate estimation (McCleary et al. 1980). More data is always preferable, but at the very least, a time series should be long enough to capture the phenomena of interest. Twenty observations per well are required as the minimum sample size for time-series forecasting to ensure better performance. Annual forecasting results from 2021 to 2025 were generated on the well-level time series analysis plots and the data summary table.

R Shiny web application

The Gelman Site 1,4-Dioxane Groundwater Contamination Plume Modeling and Forecasting tool has been developed as an R Shiny web application with a simple user interface to facilitate easy access for water experts, decision-makers, and the public. This application can be developed entirely with R code and turns R code into interactive and dynamic displays. In this web environment, our R Shiny browser-based application can be hosted on a remote server that multiple users can access simultaneously via a web interface. It eliminates the need for users to obtain and maintain the high-performance hardware required by the models and deals with software installation and operating system incompatibility issues. All needed to use the web app are an internet connection and a web browser. And its interactive nature makes it a well-suited medium for conveying complex scientific concepts to the general public and creating decision support tools that harness cutting-edge modeling techniques.

This application depends on functions contained within DWQ wqTools R-package: https://github.com/utah-dwq/wqTools. Code was adapted from Utah DWQ Lake Profile Dashboard:
https://gallery.shinyapps.io/lake-profile-dashboard/?_ga=2.214639319.1989981629.1650218987-410083735.1644383161 developed by Jake Vander Laan from Utah Division of Water Quality.

Application usage

The application can be accessed here: http://50.17.61.223:3838/gelman_shiny/. It consists of two main inputs: a map and a table. To build plots for any individual sampling well, click on the desired well in either the map or the table. The map shows all wells with concentration data available, while the table shows both wells and their associated geographic information. The map and table inputs are responsive to each other. When users click a well on the map, the table will automatically associate that well with its geographic attributes. When users click on a row in the table, the map will spontaneously zoom to its location. Outputs include time-series analysis plots, plume geospatial movement animation, and a data summary table. The time-series analysis plots are responsive to both the map and table input. They will automatically render the plot any time the user updates one of the input widgets. Several data plotting and review options are available. The plume geospatial movement animation shows how the 1,4-dioxane plume had been transported from its source Gelman site since 1986 and projects our forecasting movement till 2025. And the summarized data table aggregates all the trend analysis, historical high value, and 2021-2025 five-year forecasting data into one table. Thus we can see if there is any well that will meet any screening levels or health criteria within the next five years.

Inputs

Map element input

Individual monitoring locations are displayed as circle markers on the map (Figure 3). In addition to displaying well sites, well labels are also available. Topographic and satellite base map layers are both available. To turn on any of these layers, hover the mouse over the layers control box (top left of the map, under zoom buttons) and select one or more layers.

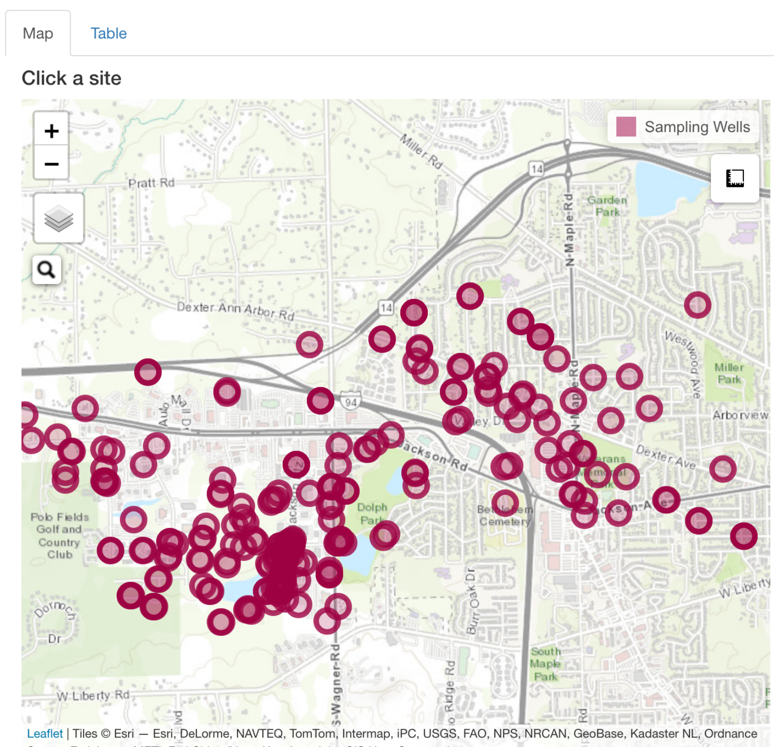


Figure 3: Map element input

- The Search feature button (represented by a magnifier) lets users search for a street address or well name. To search for a well, click the magnifying glass (top left of the map, under the layers control box) and start typing in a street address or well name. Locations matching the search will appear as they type.
- The Map zoom +/- buttons (represented by a plus sign and a minus sign with a gray background) at the top left corner can be used to zoom in or out of the map. Alternately, the mouse wheel also lets users zoom in and out, while left clicking and dragging lets users pan around the map.
- The Layers control button (represented by a stack of three squares) displays a key of the symbols used on the map. It also lets users choose which information they want to see. Click the checkbox for a layer that will turn that layer on or off on the map. A topographic or satellite base map lets users choose what their background map looks like.
- The Measure button (represented by two horizontally overlapping rulers) at the top right corner lets users measure distances or areas between features on the map. Click the map where users want to start the measurement and click once for every vertex of the line/area they want to measure. Double-click to complete the line/area.

Table input

The Search bar (represented by a white box) at the top lets users search for well name, well type, or any range of depth, elevation, latitude, and longitude (Figure 4).

Map | Table

Click a site

Search:

Well_Name	Depth	Elevation	Latitude	Longtitude	Well_Types
110 Parkland Plaza	91	936.1181	42.2831878991157	-83.8142040421868	Residential Wells
170 Aprill	50	907.2557	42.2823095310774	-83.8079959541735	Residential Wells
175 Jackson Plaza	103	929.4347	42.2815607379006	-83.8012221620632	Monitoring Wells
2575 Valley	2	941.8386	42.2856315010442	-83.7826484201087	Residential Wells
2819 Dexter Rd	-9999	939.9844	42.2875484913914	-83.7872791612204	Monitoring Wells
3365 Jackson Rd	104	936.0822	42.2844648500619	-83.7961025975852	Monitoring Wells
3563 ELIZABETH RD	-9999	954.3443	42.2901487280406	-83.8012447612354	
373 Pinewood Deep	233	934.5783	42.2874114974318	-83.790018257077	Monitoring Wells
373 Pinewood Shallow	96	934.5783	42.2874114974318	-83.790018257077	Monitoring Wells
4141 Jackson Rd	98	936.7078	42.2843064836086	-83.8131476290099	Residential Wells
4401 Park West	-9999	933.5312	42.2841248755096	-83.8172779068422	Residential Wells
4470 Jackson Rd	-9999	918.6182	42.2864536739999	-83.8187727000281	Residential Wells
456 Clarendon	102	942.355	42.2865810912895	-83.7857566736153	Monitoring Wells
4601 Park 4 inch	52	898.2162	42.282339756526	-83.8202569871109	Residential Wells
4601 Park 6 inch	42	895.2655	42.282745538356	-83.8202858503823	Residential Wells
465 Dupont	181	928.4702	42.289177688323	-83.7929150910502	Monitoring Wells
4742 Park Rd	50	889.5266	42.2846366687432	-83.8229720957591	Residential Wells
5005 Jackson Rd	-9999	859.9006	42.2864875443625	-83.8258275996101	Residential Wells
5115 Jackson Rd	-9999	866.8431	42.2873257799033	-83.8271184245712	Residential Wells
697 South Wagner Rd	-9999	938.8018	42.2750421349773	-83.7990112958945	Residential Wells

Figure 4: Table input

Description of each variable:

- Well name: Name of wells used for monitoring groundwater contamination.
- Depth: Boring depth measured in feet height below ground level. It means the distance from the ground surface to the bottom of the well screen or the bottom of the open hole when a well screen is not used.
- Elevation: Top of Casing elevation measured in feet height above mean sea level. The casing is a tubular structure placed in the drilled well to maintain the well opening and determine the groundwater flow direction.
- Latitude/Longitude: geographic Cartesian coordinates for a sampling well location.
- Well type: Includes Monitoring, Residential, Extraction, and Miscellaneous Wells.

Outputs

Time series plot output

The "well data analysis" tab shows plots summarizing historical patterns and five-year forecasting in dioxane concentration at the selected well (Figure 5). We labeled the x-axis the time-axis, and the y-axis is for the 1,4-dioxane concentration in ppb. So, we can see any past trends and predict future contamination level changes in five years. There are four horizontal lines of 1, 4, 7.2, and 85 ppb with different shades of gray, which separately represent: the detectable level; a new response trigger if detected in sentinel wells; the current EGLE generic residential cleanup criterion, and the Fourth Amended Consent Judgment drinking water standard; previous Third Amended Consent Judgment drinking water standard.

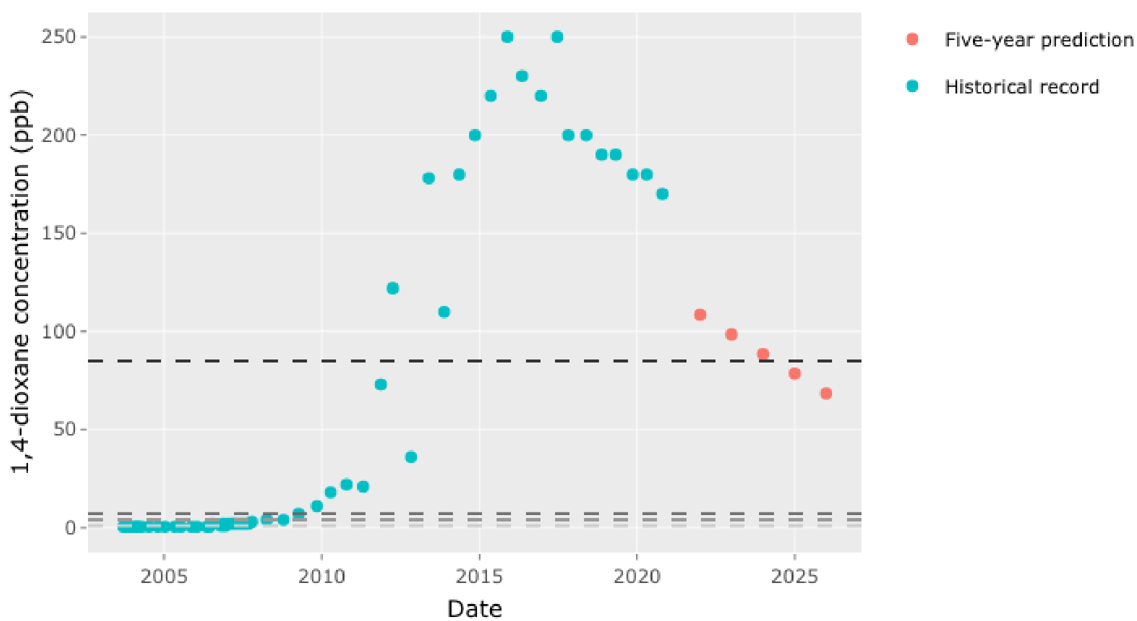


Figure 5: MW-91 time series plot output

Plume projection animation output

This plume projection animation shows historical and forecasting groundwater pollutant transportation over time (Figure 6). The 30×30 meter grids were interpolated to calculate the contaminated groundwater projection and transportation inside the plumes. This grid requires 49 rows and 93 columns for 4,557 total pixels in the active study area where plumes are displayed and characterized.

- A time slider is set up to move through each month/year. Once the user hits the Play button on the bottom right of the time slider (represented by a triangle pointing to the right), the geospatial distribution will automatically move through time. While playing, the Pause button (represented by two vertical lines) appears in its place. The value control dot (represented by a white dot in between) can be dragged along the slider bar to set the map time interactively.
- The Map zoom +/- buttons (represented by a plus sign and a minus sign with a gray background) at the top left corner can be used to zoom in or out of the map. Or the mouse wheel also lets users zoom in and out, while left-clicking and dragging lets users pan around the map.
- The dark red color represents areas of high plume concentration, and the bright yellow indicates regions where the dioxane level is relatively low.

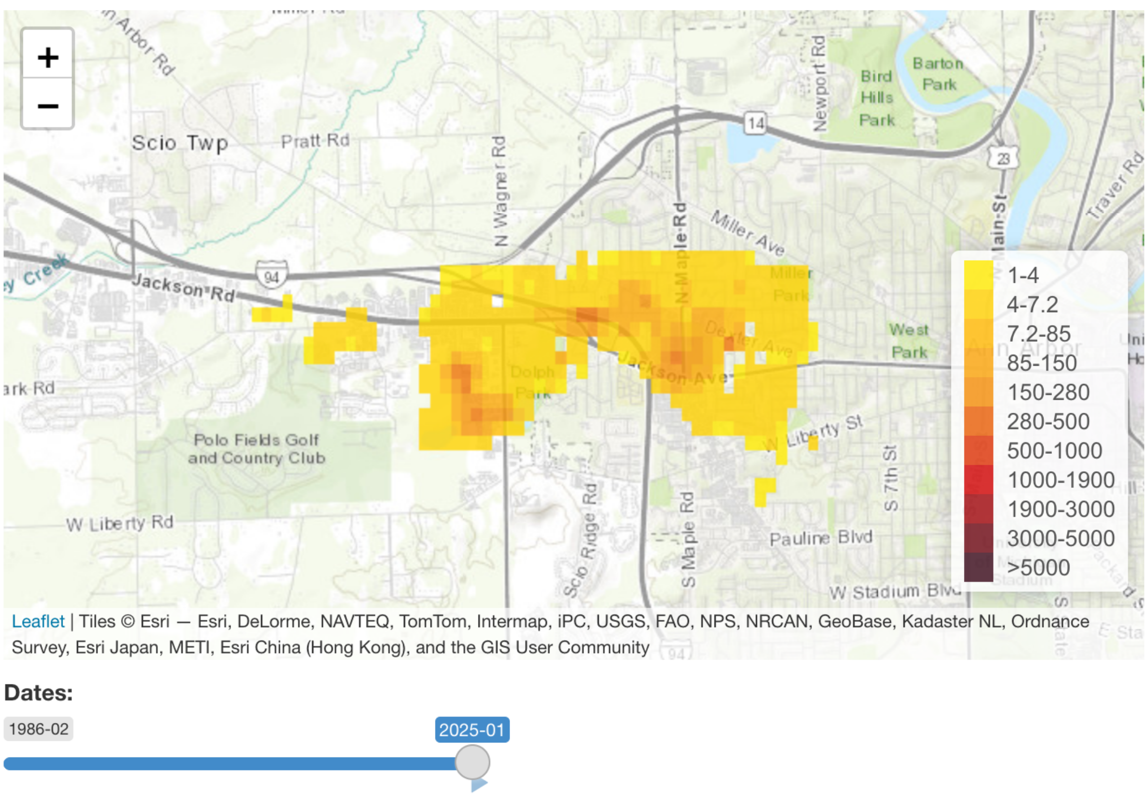


Figure 6: Plume projection animation output

Data summary table output

The summary table visualizes all well-level information in table form (Figure 7), including the trend analysis, historical high value, and five-year forecasting data between 2021 and 2025. Besides, four columns determine whether the forecasting values of the well will meet any of the screening levels or health criteria in five years. The Search bar (represented by a white box) lets users search for a specific well by name.

Search:

Well.name	Trend	Historical.High	X2021.Predict	X2022.Predict	X2023.Predict	X2024.Predict	X2025.Predict	Predict.High
170 Aprill	no trend	26	2.30091909	0.903776599	0	0	0	2.30091909
175 Jackson Plaza	increasing	1324	893.8101637	899.332249	993.6219514	1105.759578	1020.256106	1105.759578
2575 Valley	increasing	110	81.65572344	86.4442456	91.23276776	96.03557447	100.8240966	100.8240966
2643 Dexter Rd	no trend	12	31.2847344	4.196700165	4.446853344	4.697006522	4.947159701	31.2847344
2652 Dexter Rd	decreasing	190	8.676364752	3.316567004	3.881633931	1.673521092	3.29924695	8.676364752
2690 Dexter Rd	increasing	490	986.9208333	1774.089369	2571.005607	3371.169546	4102.914522	4102.914522
2805 Dexter Rd	increasing	1599	1246.573571	1129.964222	1165.463665	1142.963665	1299.015585	1299.015585
2819 Dexter Rd	no trend	1085	96.714957	39.94493068	0	0	0	96.714957
305 Pinewood	no trend	3	1.038979573	1.073988992	1.10899841	1.144007829	1.179017247	1.179017247
3245 Kingwood	decreasing	65	0	0	0	0	0	0
333 Jackson Plaza	decreasing	60000	19	19	19	19	19	19
3365 Jackson Rd	decreasing	4600	180	180	180	180	180	180
354 Pinewood	no trend	4	0.5	0.5	3	1.75	1.75	3
3563 Elizabeth Rd	no trend	4	0.775534669	1.775558304	1.775558304	1.775534669	1.775534669	1.775558304
373 Pinewood Shallow	decreasing	4285	179.0685417	147.5426234	115.4222453	82.70577865	49.39485209	179.0685417

Showing 1 to 245 of 245 entries

Figure 7: Data summary table output

Disclaimer

The information contained in this application is for general information purposes only and should not be used for navigation, regulatory, permitting, or other legal purposes. We have used our reasonable efforts to ensure that the data analysis we release is complete, accurate, and useful. However, because we do not create the data and the processing required to make the data useful is complex, we cannot be liable for omissions or inaccuracies. Both the time series forecasting and spatial movement projection are provided ‘as is.’ Furthermore, some uncertainty and limitations associated with this data analysis will be described in the limitations section.

Results

Historical trend analysis summary

Out of 245 sampling wells, almost half (48.6%) showed decreasing trends from our Mann-Kendall test (Figure 8). More than a quarter (25.7%) exhibited an increasing trend, despite over three decades of clean-up and treating contaminated groundwater occurring since the early 1990s. The other 63 wells (25.7%) showed no significant trend.

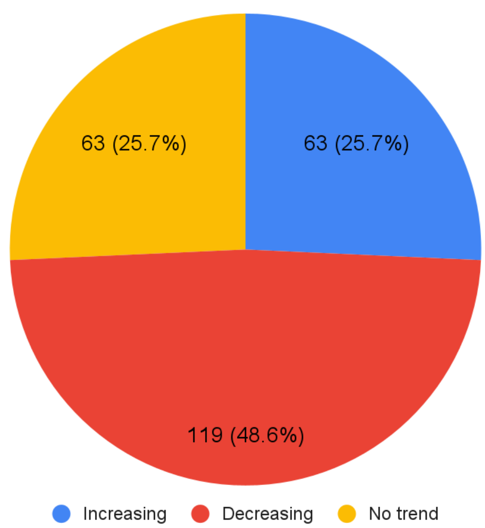


Figure 8: historical trend analysis pie chart

Five-year forecasting data summary

Our AutoTS forecasting model showed 53 wells (21.6%) forecasting values below the detection limit of 1 ppb among 245 sampling wells (Figure 9). More than a third of wells (34.3%) with predicting concentrations lower than 4 ppb of a new response trigger if detected in sentinel wells. The EPA’s Integrated Risk Information System assessed the 4 ppb water quality criterion for excess cancer risk of one in 100,000 (USEPA 2013). And only 92 wells (37.6%) will meet the current MDEQ residential drinking water cleanup criteria in the next five years. 60.8% of the wells will meet the previous third amended Consent Judgment drinking water standard before 2025.

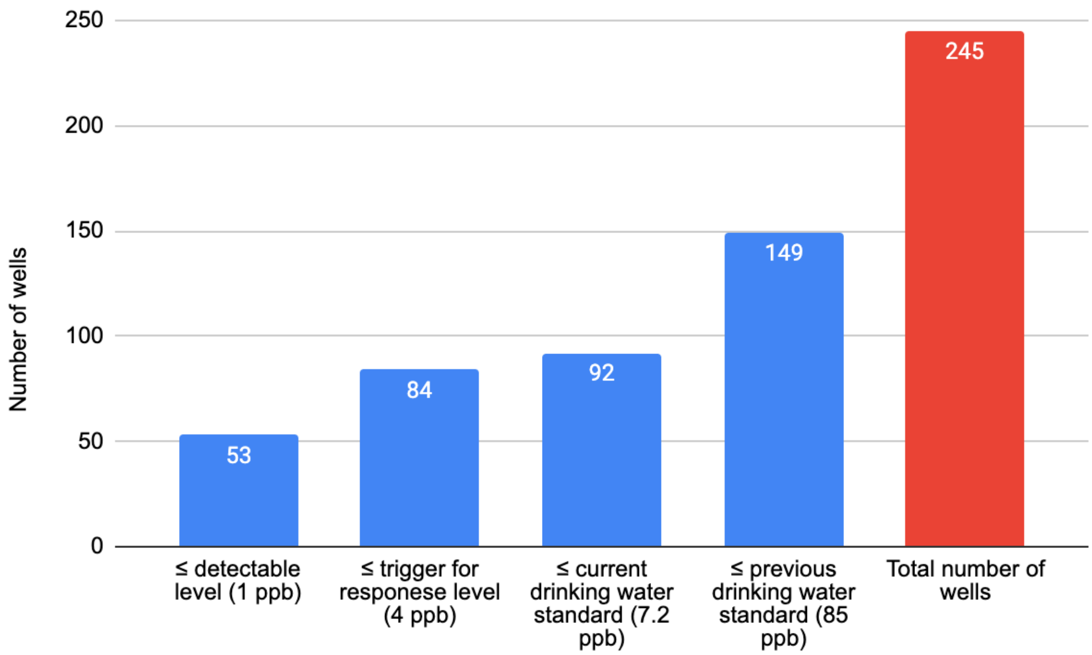


Figure 9: Number of wells that meets screening levels or health criteria bar chart

Limitations

There are several limitations in the current work; hence our findings need to be interpreted carefully. Below are some possible methodological limitations that were too complex to take into consideration in this study.

A plume is a three-dimensional distribution of contaminants in groundwater, and its shape and movement are affected by source spreading, geological complexities, hydrological characteristics, and biological activities. In this study, the estimated plume boundary on two-dimensional maps cannot show the depth to contamination or thickness of the contaminated zone.

Moreover, not all wells with available data could be analyzed and forecasted in our research. Studies found that estimation performance increases with increasing time points as the bias of the standard error of the autoregressive parameter decreases (Krone et al. 2017). Wells with less than 20 observations, which does not reach the minimum number of observations for time-series analysis, cannot be used for trend analysis and forecasting.

At last, not all wells shown within the plume boundaries have detectable levels of 1,4-dioxane. All groundwater and treated water samples were analyzed by gas chromatography-mass spectrometry (GC–MS) with selected ion monitoring (USEPA 2008). This method allows for a target detection limit of approximately 1 ppb. And all samples not detected 1,4-dioxane were all labeled as a fixed value of 0.5 ppb. EPA risk assessments indicate that the drinking water concentration representing a 1×10^{-6} cancer risk level for 1,4-dioxane is 0.35 ppb (USEPA 2013). Therefore, the detection and quantitation limits should be as low as possible so that the carcinogenicity risk of 1,4-dioxane can be prevented earlier.

Conclusion

The methodology adopted here permits us to analyze the data without masking the important temporal periodicities and spatial non-stationarity. This framework could serve as a valuable tool for enhancing groundwater time series for intermittent missing data and continuous data gaps, as demonstrated in this effort. Overcoming existing data limitations with methodologies that augment the available groundwater contamination sampling data could improve the quality of input data for data-driven analysis and groundwater contamination plume modeling and forecasting. It could provide further insight into groundwater quality restoration and improve groundwater resources planning and management through efficient conjunctive use of surface water and groundwater resources.

Bibliography

Adams, Craig, C. P. Grady Jr, and R. Cowan. 1994. "Elimination of 1,4-Dioxane from Polyester Manufacturing Wastewater through Point-of-Generation Pretreatment." *The Third South Carolina Environmental Symposium*, January. https://digitalcommons.usu.edu/cee_facpub/560/.

Alley, William M., and Charles J. Taylor. 2001. “The Value of Long-Term Ground Water Level Monitoring.” *Groundwater* 39 (6): 801–1. https://doi.org/10.1111/j.1745-6584.2001.tb02466.x.

Barbash, Jack E, and Frank D. Voss. 2016. “Design and Testing of a Process-Based Groundwater Vulnerability Assessment (P-GWAVA) System for Predicting Concentrations of Agrichemicals in Groundwater across the United States.” *Scientific Investigations Report*. https://doi.org/10.3133/sir20145189.

Box, G. E. P., and David A. Pierce. 1970. “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models.” *Journal of the American Statistical Association* 65 (332): 1509–26. https://doi.org/10.1080/01621459.1970.10481180.

Boy-Roura, Mercè, Bernard T. Nolan, Anna Menció, and Josep Mas-Pla. 2013. “Regression Model for Aquifer Vulnerability Assessment of Nitrate Pollution in the Osona Region (NE Spain).” *Journal of Hydrology* 505 (November): 150–62. https://doi.org/10.1016/j.jhydrol.2013.09.048.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/a:1010933404324.

Catlin, Colin. 2022. “AutoTS: Automated Time Series Forecasting.” PyPI. February 6, 2022. https://pypi.org/project/AutoTS/.

Cutler, D. Richard, Thomas C. Edwards, Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. “Random Forests for Classification in Ecology.” *Ecology* 88 (11): 2783–92. https://doi.org/10.1890/07-0539.1.

EGLE. 2021. “Gelman Site Groundwater Contamination Plume Map – 7.2 Ppb (2020).” https://www.michigan.gov/documents/egle/Dioxane-Plume_Map-7-2-ppb-Gelman-Sciences-Inc-Remediation-2020-2021-07-13_731040_7.pdf.

Gleick, Peter H. 1993. “Water and Conflict: Fresh Water Resources and International Security.” *International Security* 18 (1): 79. https://doi.org/10.2307/2539033.

Hartmann, Andreas, Scott Jasechko, Tom Gleeson, Yoshihide Wada, Bartolomé Andreo, Juan Antonio Barberá, Heike Brielmann, et al. 2021. “Risk of Groundwater Contamination Widely Underestimated because of Fast Flow into Aquifers.” *Proceedings of the National Academy of Sciences* 118 (20). https://doi.org/10.1073/pnas.2024492118.

Jake, Vander Laan. 2022. “Utah DWQ’s WqTools R-Package.” GitHub. January 7, 2022. https://github.com/utah-dwq/wqTools.

Kendall, Maurice G. 1975. *Rank Correlation Methods.* Bristol: C. Griffin.

Krone, Tanja, Casper J. Albers, and Marieke E. Timmerman. 2017. “A Comparative Simulation Study of AR (1) Estimators in Short Time Series.” *Quality & Quantity* 51 (1): 1–21. https://doi.org/10.1007/s11135-015-0290-1.

Mann, Henry B. 1945. “Nonparametric Tests against Trend.” *Econometrica* 13 (3): 245. https://doi.org/10.2307/1907187.

Marchant, B.P., and J.P. Bloomfield. 2018. “Spatio-Temporal Modelling of the Status of Groundwater Droughts.” *Journal of Hydrology* 564 (September): 397–413. https://doi.org/10.1016/j.jhydrol.2018.07.009.

Mccleary, Richard, and Richard Hay. 1980. *Applied Time Series Analysis for the Social Sciences*. Beverly Hills, Calif.: Sage Publications.

MDEQ. 2004. “Information Bulletin: Gelman Sciences, Inc. Site, Scio Township, Washtenaw County, Michigan, March. Lansing, MI: Michigan Department of Environmental Quality.” https://www.michigan.gov/documents/deq/deq-rrd-GS-GelmanFactSheet_286728_7.pdf.

Oikonomou, Panagiotis D., Ayman H. Alzraiee, Christos A. Karavitis, and Reagan M. Waskom. 2018. “A Novel Framework for Filling Data Gaps in Groundwater Level Observations.” *Advances in Water Resources* 119 (September): 111–24. https://doi.org/10.1016/j.advwatres.2018.06.008.

Rahaman, Md, Balbhadra Thakur, Ajay Kalra, and Sajjad Ahmad. 2019. “Modeling of GRACE-Derived Groundwater Information in the Colorado River Basin.” *Hydrology* 6 (1): 19. https://doi.org/10.3390/hydrology6010019.

Rouhani, Shahrokh, and Hans Wackernagel. 1990. “Multivariate Geostatistical Approach to Space-Time Data Analysis.” *Water Resources Research* 26 (4): 585–91. https://doi.org/10.1029/wr026i004p00585.

Rupert, Michael. 2003. “Probability of Detecting Atrazine/ Desethyl-Atrazine and Elevated Concentrations of Nitrate in Ground Water in Colorado.” https://pubs.usgs.gov/wri/wri02-4269/pdf/WRIR02-4269.pdf.

USEPA. 2006. “Treatment Technologies for 1,4-Dioxane: Fundamentals and Field Applications.” *Remediation Journal* 17 (3): 81–88. https://doi.org/10.1002/rem.20135.

USEPA. 2008. “Method 522 - Determination of 1,4-Dioxane in Drinking Water by Solid Phase Extraction (SPE) and Gas Chromatography-Mass Spectrometry (GC/MS) with Selected Ion Monitoring (SIM).” Cfpub.epa.gov. November 12, 2008. https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NERL&dirEntryId=199229.

USEPA. 2013. “Toxicological Review of 1,4-Dioxane (with Inhalation Update).” https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0326tr.pdf.

USEPA. 2017. “Technical Fact Sheet - 1,4-Dioxane.” https://www.epa.gov/sites/default/files/2014-03/documents/ffrro_factsheet_contaminant_14-dioxane_january2014_final.pdf.

Zenker, Matthew J., Robert C. Borden, and Morton A. Barlaz. 2003. “Occurrence and Treatment of 1,4-Dioxane in Aqueous Environments.” *Environmental Engineering Science* 20 (5): 423–32. https://doi.org/10.1089/109287503768335913.