Application of Time Series Analysis on Temporal Variation of Fluoride in Groundwater around Southern Taiwan Science Park

Ting-Nien Wu\textsuperscript{1,2}, Jan-Yee Lee\textsuperscript{1}, Chen-Hsiang Huang\textsuperscript{1,2}

1. Department of Environmental Engineering, Kun Shan University, Tainan 71003, Taiwan, R.O.C.
2. Free Air Engineering Consultants Company, Taichung 40766, Taiwan, R.O.C.

\texttt{wutn@mail.ksu.edu.tw, jylee@mail.ksu.edu.tw, xianging@gmail.com}

Abstract

This paper demonstrated a case study on how to utilize time series analysis as mining tool to track the transience of fluoride release and to predict the fluoride concentration in groundwater. Southern Taiwan Science Park was selected as study area, and seven groundwater monitoring wells located in the domain of fluoride release were subjected to time series analysis. The measured fluoride levels in groundwater are between 0.4 and 3.6 mg/L, and the series data is stationary in mean and variance during the period of 2005 and 2009. Time series analysis is a useful tool for extracting interesting pattern from ordered sequence of observations. Based on extracting information from ACF and PACF, the trend, interesting patterns, rules, or models within the original data set were filtered out. The common time series models, ARMA and ARIMA, were employed to interpret the information beneath the monitoring data of groundwater quality. Through verification by Akaike’s information criterion (AIC) and Schwartz’s Bayesian Information Criterion (SBC), the ARMA(1,1) model was identified as the best fitted model for data interpretation and estimation. Accordingly, this developed numerical model can effectively interpret and forecast the fluoride level in groundwater as referring the prior information.

Keywords: time series analysis, fluoride, groundwater contamination, monitoring well, data mining.

1. Introduction

Many factors including climate, topography, aquifer lithology, surface water recharge, saline water intrusion, and human activities may have a significant impact on groundwater quality and further limit its use as water supply. For the purpose of conserving underground water resources, hundreds of monitoring wells were established and monitored quarterly or biannually in Taiwan. Monitoring data from each well directly reflects groundwater quality through comparing with the regulated standards, but the origins of contaminant sources and their transience may not be easily observed from monitoring data. By using statistical tools, it is achievable to identify the sources of potential groundwater contamination [1], to allocate the spatial domains of contaminant sources [2-3], and to track the transience of a certain contaminant or potential sources [4-5].

Southern Taiwan Science Park (STSP) is one of the major integrated circuits and optoelectronics industry clusters in Taiwan. There are 45 monitoring wells and 6 groundwater table recorders established for monitoring groundwater quality so far. Our previous study of principal component analysis pointed out that the potential sources of groundwater contamination around STSP were identified as salinization, arsenic dissolution, industrial leakage, mineralization, and agricultural activities [4]. Hydrogen fluoride (HF) is widely used in integrated circuits industry, green energy technology and energy saving industry, and thus the industrial leakage might cause fluoride release to groundwater. It is crucial to track the transience of fluoride release and to predict the fluoride concentration in groundwater.

This paper employed time series analysis to anatomize the temporal variations of fluoride level in groundwater. Several approaches of time series analysis were investigated to ascertain the trend of temporal variations on fluoride level in groundwater; including autoregressive process (AR), moving average process (MA), autoregressive moving average process (ARMA), and autoregressive integrated moving average process (ARIMA) models. This study provides a showcase to interpret the fate of fluoride release from mining groundwater monitoring data. Furthermore, knowledge discovery regarding fluoride release in groundwater is especially valuable for STSIP Administration Bureau to improve groundwater management and prevent groundwater contamination.
2. Study Area

Southern Taiwan Science Park located in Tainan County where the major production came from agricultural activities. The study area belongs to Chianan Plain groundwater subregion that is a shallow aquifer consisting of 3 to 5 mixed layers of clay, silt and silty sand. This type of geological structures may hamper percolation and limit groundwater recharge to the deeper layers. According to hydrogeologic survey within STSIP, groundwater flows in the direction of northeast to southwest and finally discharges to Taiwan Strait. In this region, the transmissivity and storativity of aquifer are 0.014m²/min and 3×10⁻⁴ respectively [6].

Southern Taiwan Science Park accommodates 7 industry clusters, including integrated circuits, optoelectronics, biotechnology, telecommunications, precision machinery, computer & peripherals, green energy and energy saving. The manufacture processes are complicated and varied with different industry clusters. Lots of kinds of chemicals were used in factory production, and it may turn to be the potential source of groundwater contamination once occurring leakage.

3. Groundwater Monitoring Data

For the purpose of groundwater protection, there are 45 monitoring wells were established for routine groundwater sampling and analysis within STSIP. The established monitoring wells were quarterly sampled and analyzed since 2000. The analyzed items of groundwater quality include pH, electrical conductivity (EC), temperature, total dissolved solid (TDS), total organic carbon (TOC), fluoride, ammonia, nitrite, nitrate, Cu, Cr, Ni, Cd, Pb, As, Zn, Hg, Fe, Mn, and some volatile organic compounds (VOCs). Based on the results of principal component analysis on groundwater monitoring data from 2005 to 2007, fluoride release was identified as the third principal component contributing 10.7% of the total variance in the original data set [4]. Also, fluoride release has been affecting groundwater quality since 2006. By using 2-step cluster analysis, EC, arsenic, fluoride, iron, and nitrite were appointed as the characterized variables to correspond with each PC. The domain of the corresponding PC was determined through connecting the locations of the classified monitoring wells in each cluster. As shown in Figure 1, there are 7 monitoring wells classified in the domain of fluoride release.

4. Method

Time series analysis is known as a powerful tool for summarizing ordered sequences of observations with various degrees of correlations over time by employing the concepts of stochastic process for dynamic systems. It is well-known that time series analysis is a common approach of economic forecasting, and its practical application also covers from neurophysiology to astrophysics. Typically, time series analysis is based on the assumption that the data consist of systemic pattern and random error. Through extracting information from autocorrelation functions (ACF) and partial autocorrelation functions (PACF), it can filter out noise in order to make the pattern more salient. The first step in time series analysis is to perform some preliminary processing to make the raw data suitable for further analysis. The following steps for analyzing and modelling time series include constructing the graph, data transformation, examining the ACF and PACF, identifying proper orders of differencing and moving average, and model diagnosis. In this study, all the analysis was computed by the SPSS-12.0 software.

Two common models of time series analysis including ARMA and ARIMA models were employed to explore the linkage between fluoride concentration and groundwater sampling time. The algorithm of ARMA and ARIMA models were summarized as below [7]:

1. Autoregressive Moving Average Process Model
(ARMA)  
The process \( \{Z_t\} \) is said to be an ARMA model if it can be expressed as  
\[
\Phi_p(B)Z_t=\theta_q(B)a_t ,
\]  
where,  
\[
\Phi_p(B)=1-\Phi_1B-\cdots-\Phi_pB^p ,
\]  
\[
\theta_q(B)=1-\theta_1B-\cdots-\theta_qB^q
\]  
The model is usually referred to ARMA(p,q) where p and q are integers greater than or equal to zero and refer to the order of the autoregressive and moving average parts of the model.

2. Autoregressive Integrated Moving Average Process Model (ARIMA)  
On the other hand, if the process \( \{Z_t\} \) can be expressed as the following form, then the model is said to be an ARIMA model.  
\[
\Phi_p(B)(1-B)^dZ_t=\theta_0+\theta_q(B)a_t ,
\]  
where,  
\[
\Phi_p(B)=1-\Phi_1B-\cdots-\Phi_pB^p
\]  
\[
\theta_q(B)=1-\theta_1B-\cdots-\theta_qB^q
\]  
It is a generalization of ARMA model. The model is generally referred to as an ARIMA(p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

5. Result and Discussion  
Based on 2-step cluster analysis, 7 monitoring wells were allotted to the domain of fluoride release. The time series data of fluoride concentration in each monitoring well was given in Figure 2. The measured fluoride levels in groundwater are between 0.4 and 3.6 mg/L, and the series data is stationary in mean and variance during the period of 2005 and 2009. The peak fluoride level observed in MW13 and MW25 may be resulted from an accidental leakage in May 2006. The average fluoride level in MW26s was found 3 to 4 times higher than other monitoring wells. The location of MW26s is close to a great LCD manufacturer, and the potential source of fluoride release is pointed to this manufacturer. Also, the contamination of fluoride is more serious in the deep well.

Because the series data meet the requirement of variance stabilization, there is no need for data transformation. As mentioned above, time analysis is based on extracting information from ACF and PACF that filter out the trend, interesting patterns, rules, or models for the data set. In Figure 3, the ACF and PACF exceeding the confidence interval occurs only in the first 2 lags. Accordingly, ARMA(1,1), ARMA(1,2), ARMA(2,1), and ARIMA(1,1,1) models were employed for estimation in time series analysis.

The applied time series models might describe the data adequately, while model comparison is usually examined by Akaike’s information criterion (AIC) and Schwartz’s Bayesian Information Criterion (SBC). AIC is served as a criterion for assessing the goodness of model fitting, and the smaller AIC value normally corresponds to the better model fitting for a given data set. As a general rule, SBC has the same trend with AIC that the smaller value explains the better model fitting.

The results of the ARMA and ARIMA models showed that time series analysis can briefly reduce the complexity of the original data set. Table 1 illustrated the parameter estimation as well as AIC and SBC for 4 fitted ARMA and ARIMA models. Comparing both AIC and SBC of 4 fitted models, the ARMA(1,1) model gained the minimum values that was accordingly selected as the best fitted model for data interpretation.

![Figure 2. Fluoride concentrations of each monitoring well within the domain of fluoride release during year 2005 and 2009.](image-url)
Figure 3. ACF and PACF of fluoride concentration for the ARMA model.

Table 1. Estimation Results of ARMA and ARIMA models with the AIC and SBC values.

<table>
<thead>
<tr>
<th>Models</th>
<th>parameter</th>
<th>coefficient</th>
<th>standard error</th>
<th>T-Ratio</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(1,1)</td>
<td>AR1</td>
<td>0.41</td>
<td>0.21</td>
<td>1.92</td>
<td>326.67</td>
<td>335.34</td>
</tr>
<tr>
<td></td>
<td>MA1</td>
<td>0.68</td>
<td>0.17</td>
<td>4.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>constant</td>
<td>1.02</td>
<td>0.04</td>
<td>26.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>AR1</td>
<td>-0.49</td>
<td>0.30</td>
<td>-1.65</td>
<td>326.17</td>
<td>337.73</td>
</tr>
<tr>
<td></td>
<td>MA1</td>
<td>-0.25</td>
<td>0.29</td>
<td>-0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA2</td>
<td>0.33</td>
<td>0.09</td>
<td>3.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>constant</td>
<td>1.02</td>
<td>0.04</td>
<td>23.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>AR1</td>
<td>-0.7</td>
<td>0.29</td>
<td>-2.43</td>
<td>327.49</td>
<td>339.06</td>
</tr>
<tr>
<td></td>
<td>AR2</td>
<td>-0.3</td>
<td>0.09</td>
<td>-3.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MA1</td>
<td>-0.49</td>
<td>0.30</td>
<td>-1.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>constant</td>
<td>1.02</td>
<td>0.05</td>
<td>19.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA(1,1,1)</td>
<td>AR1</td>
<td>-0.11</td>
<td>0.09</td>
<td>-1.20</td>
<td>340.32</td>
<td>348.97</td>
</tr>
<tr>
<td></td>
<td>MA1</td>
<td>0.94</td>
<td>0.07</td>
<td>13.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>constant</td>
<td>0.002</td>
<td>0.005</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Residual ACF and PACF of fluoride concentration for ARMA(1,1) model.
Thus, the obtained ARMA(1,1) model can be expressed as the following:

\[ F_t - 1.02 = 0.41(F_{t-1} - 1.02) + a_t - 0.68a_{t-1} \] (1)

where, \( F_t \) represents fluoride concentration at time \( t \), 
\( a_t \) represents the white noise at time \( t \), 
\( F_{t-1} \) represents fluoride concentration at time \( t-1 \), and 
\( a_{t-1} \) represents the white noise at time \( t-1 \).

In other words, the fluoride level at time \( t \) is subjected to stochastic influence by the previous fluoride level at time \( t-1 \). Because the parameter coefficient of AR1 is less than 1, it denotes that the predicted fluoride level won’t be increasing with time.

If the residual estimation error for a particular model, the so-called white noise, is within a specified tolerance level, then the applied model is in goodness of model fitting. The residual ACF, PACF, and Box-Ljung tests are tools for model diagnostic checking. As illustrated in Figure 4, the residual ACF and PACF plots indicated no significant spikes. Also, the Box-Ljung test for the residual estimation error is not significant. It implicated that ARMA(1,1) is a suitable model for the prediction of fluoride concentration in groundwater within the study area.

6. Conclusions

The source of fluoride in groundwater might be most likely resulted from industrial leakage, especially around the industrial park of high electronic technology. By using time series analysis, the numerical models were developed for interpreting and forecasting the fluoride level in groundwater as referring the prior information. The fitted model tentatively discovered the temporal trend of the fluoride level in groundwater, and the predicted fluoride level revealed that the underlying processes such as dispersion and diffusion might effectively attenuate the fluoride level in groundwater. This study has successfully demonstrated a showcase of applying time series analysis to track the transience of fluoride release and to predict the fluoride concentration in groundwater. With the aid of time series analysis, the improvement of groundwater management as well as contaminant control is under progress in Southern Taiwan Science Park.

References