

# Association rules and regression linear model of the groundwater population by the evaluation of uranium

Iing Lukman<sup>1,\*</sup> and Natalina<sup>2</sup>

<sup>1</sup>Department of Management, Faculty of Economics, Universitas Malahayati, Bandar Lampung, Indonesia

<sup>2</sup>Department of Environmental Engineering, Faculty of Engineering, Universitas Malahayati, Bandar Lampung, Indonesia

**Abstract.** The uranium available more on groundwater samples of certain types on the total alkalinity were relatively the same. But, the content of the uranium was higher in the samples. The multiple linear regression for pH as a dependent variable showed that the pH negatively correlated to the uranium, but the uranium was not significant for the linear regression model. The data of groundwater population from the samples of 127 with 12 variables of measurement of the Energy Department of the United States of America resulted in those association rules and linear regression models. The data has five factors of Producing horizon namely Ogallala Formation (TPO), Dockum Formation (TRD), Quartermaster Group (POQ), Whitehorse and Cloud Chief Group (PGWC), El Reno Group and Blaine Formation (PGEB). The step-wise linear regression for each of the five producing horizon codes was fitted to the data. Then, the regression models for each variable of producing horizon were obtained if pH was the dependent variable. If the Uranium was a dependent variable, then the regression models obtained were four only, with the model for PGEB was not able to be made. When pH as a dependent variable, it was depended upon Boron, Total alkalinity, and Bicarbonate.

## 1 Introduction

Uranium is a naturally occurring alpha-emitting radionuclide and plausible human carcinogen [1]. Uranium is a radioactive metal that occurs in low concentration in nature [2]. It is present in certain types of soils and rocks, especially granites [2]. The interim maximum acceptable concentration of uranium in drinking in Health Canada's Guidelines for Canadian Drinking Water Quality is 0.020 mg/L (milligrams per litre) [2]. Rural areas are depended heavily on groundwater sources approximately 40% of South Carolina (SC) residents regularly use ground-water as their primary drinking water sources [3]. Results from ecological investigation [4] suggested that colorectal, breath, kidney, and total cancer incidences may be elevated in areas with frequent groundwater use and elevated groundwater uranium. Uranium persists in the groundwater at former mining sites [5].

## 2 Methodology

The methodology of association rules and stepwise regression were applied for data analysis. Association rules are an important class of regularities that exist in databases. Since association rules were first introduced by [6], the problem of mining associations has received a

great deal of attention. The classic application is market basket analysis, it is analysis how the items purchased by customers are associated. An example of an association rule is as follows:

*Roticanai* ---> *tehtarik* [sup = 10%, conf = 80 %] This rule says that 10% of customers buy *roticanai* and *tehtarik* together, and those who buy *roticanai* also buy *tehtarik* 80% of the time [7].

Stepwise regression is a semi-automated process of building a model by successively adding or removing variables based solely on the t-statistics of their estimated coefficients. Properly used, the stepwise regression [8] option in Statgraphics (or other stat packages) puts more power and information at your fingertips than does the ordinary multiple regression option, and it is especially useful for sifting through large numbers of potential independent variables and/or fine-tuning a model by poking variables in or out. Improperly used, it may converge on a poor model while giving you a false sense of security. It's like doing carpentry with a chainsaw: you can get a lot of work done quickly, but it leaves rough edges and you may end up cutting off your own foot if you don't read the instructions, remain sober, engage your brain, and keep a firm grip on the controls. It is not a tool for beginners or a substitute for education and experience. The association rules could not uncover the relation

\* Corresponding author: [deiingofhelukman@gmail.com](mailto:deiingofhelukman@gmail.com)

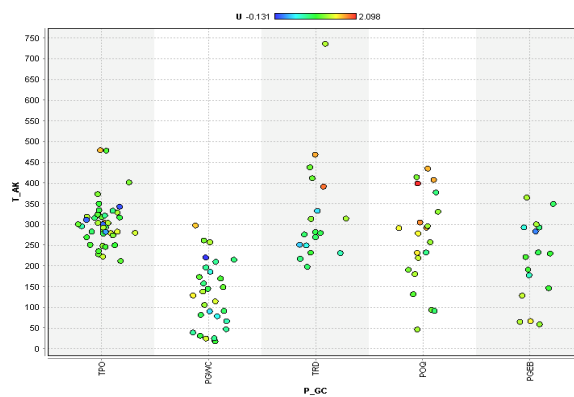
between uranium and other variables [9], and also it mentioned that pH had correlated negatively to uranium, and in the linear regression when pH is a dependent variable, the existence of uranium was not significant [9]. However, Uranium much more available on groundwater samples of PGEB > POQ > TRD > PGWC > TPO, where PGEB was El Reno Group and Blaine Formation, POQ was Quartermaster Group, TRD was Dockum Formation, the lay out of the data is as follows [10] In [11]:

**Table 1.** Lay out of groundwater data

<b>Sample Number</b>	<b>2201</b>	.....	<b>2210</b>
<b>Latitude</b>	33.21	.....	33.164
<b>Longitude</b>	101.445	.....	100.608
<b>P_GC</b>	TPO	.....	PGWC
<b>U</b>	7.99	.....	3.1
<b>AS</b>	17.6	.....	4.0
<b>B</b>	300	.....	625
<b>BA</b>	150	.....	750
<b>MO</b>	-4	.....	-4
<b>SE</b>	0.4	.....	-0.2
<b>V</b>	100	.....	25
<b>SO4</b>	35	.....	-30
<b>T_AK</b>	278	.....	175
<b>BC</b>	157	.....	101
<b>CT</b>	640	.....	540
<b>pH</b>	7.60	.....	7.56

PGWC was Whitehorse and Cloud Chief Group, and TPO was Ogallala Formation. In this paper, the step-wise linear regression was applied where the independent variables were pH and Uranium. The association rules were also applied.

Where P\_GC is Producing Horizon, U is Uranium, AS is Arsenic, B is Boron, BA is Barium, MO is Molybdenum, SE is Selenium, V is Vanadium, SO4 is Sulfate, T\_AK is Total Alkaninity, BC is Bicarbonate, CT is Conductivity, pH is an indicator for solution, whether acid or alkaline.



**Fig.1.** Uranium scatter diagram based on groundwater types sample by T-AK (Total Alkalinity) and P-GC (Producing Horizon Codes)

The association rules on Table 2 was not uncover the involvement of Uranium. As seen on Table 2, for example [SE, P\_GC = TPO] --> [T\_AK] (Confidence: 0.861), this indicated that: if Selenium was available in the groundwater and the Producing Horizon code was TPO, then it would have been available the Vanadium by confidence rates of 0.861 (86.1%). Then, if [SO4] --> [CT], 0.859 then it indicated that if available SO4, the it would be available CT by it confidence of .859 or 85.9 %.

**Table 2.** Association rules taken from the data [9]

Premises	Implies	Conclusion	Confidence
[BA, V]	→	[longi-tude, AS]	0.846
[AS, BC]	-->	[T_AR, Longi-tude, V]	0.846
[T_AK, BC, SE]	-->	[Longi-tude]	0.850
[T_AK, BC, BA]	-->	[Longi_tude]	0.850
[Longi_Tude, AS, V]	-->	[BC]	0.850
[Longi_Tude, AS]	-->	[P_GC = TPO]	0.854
[BC, SE]	-->	[Longi-tude]	0.857
[SO4]	-->	[CT]	0.859
[T_AK, BA]	-->	[Longi-tude]	0.860
[SE, P_GC = TPO]	-->	[T_AK]	0.861
[V, P_GC = TPO]	-->	[T_AK]	0.861
[SO4, Latitude]	-->	[CT]	0.861
[MO, SE]	-->	[Longi-tude]	0.861
[SE, P_GC = TPO]	-->	[V]	0.861

Premises	Implies	Conclusion	Confidence
[V, P_GC = TPO]	-->	[SE]	0.861
[SE, P_GC =TPO]	-->	[T_AK, Longi-tude]	0.861
[Longi-tude, SE, P_GC = TPO]	-->	[T_AK]	0.861

**Table 3.** R-square when pH is a dependent variable

Model	R	R Square
1	.175 <sup>a</sup>	.031
2	.247 <sup>b</sup>	.061
3	.311 <sup>c</sup>	.097
4	.672 <sup>d</sup>	.451
5	.669 <sup>e</sup>	.448

**Table 4.** Model for each producing horizon code

No.	Model Parameter	F-value	Sig.
1	B	3.963	0.049
2	B, CT	4.015	0.020
3	B, CT, T_AK	4.402	0.006
4	B, CT, T_AK, BC	25.094	0.000
5	B, T_AK, BC	33.231	0.000

- a. Predictors: (Constant), B.
- b. Predictors: (Constant), B, CT.
- c. Predictors: (Constant), B, CT, T\_AK.
- d. Predictors: (Constant), B, CT, T\_AK, BC.
- e. Predictors: (Constant), B, T\_AK, BC
- f. Dependent Variable: pH

The five regression models on Table 3 and 4 are as follows: All the models have its pH as dependent variable, the first model is for producing code of TPO as follows:

$$pH = 7.444 + 0.175 B \tag{1}$$

The regression model indicated that the pH in the groundwater samples of the Producing Horizon code of TPO (Ogallala Formation) is depended upon the element of Boron. The second for producing code of TRD as follows:

$$pH = 7.513 + 0.273 B - 0.261 CT \tag{2}$$

indicated that the pH in the groundwater samples of Producing horizon code of TRD is depended upon Boron and Conductivity (CT). However, pH and CT are negatively correlated. It can be seen in the equation (2)

that if the CT increased in one unit, then the pH will decrease by 0.261.

The third is for POQ as follows:

$$pH = 7.720 + 0.277 B - 0.261 CT - 0.199 T_{AK} \tag{3}$$

The equation (3) indicated that the pH are negatively correlated either to CT or to T\_Ak (Total Alkalinity).

The forth is for PGWC as follows:

$$pH = 7.614 + 0.155 B - 0.075 CT - 3.920 T_{AK} + 3.811 BC \tag{4}$$

The equation (4) indicated that pH is negatively correlated either to CT or to T\_Ak. However, pH is positively correlated to either Boron or Bicarbonate (BC) as it can be seen on the coefficients of B (Boron) and BC.

The fifth is for PGEb as follows:

$$pH = 7.571 + 0.12 B - 3.99 T_{AK} + 3.910 BC \tag{5}$$

The equation (5) indicated that the pH is negatively correlated to the T\_Ak (total alkanity). The equation (1) until (5) can be applied or the five models can be recommended for application in the future. The every equation of pH for every Producing Horizon code is different and thus all the models are the recommended model. It meant that the pH rate of the groundwater is depended upon Boron (The first model); it depended upon B and CT (The second model); it depended upon B, CT, and T\_Ak (The third model); it depended upon B, CT, T\_Ak, BC (The fourth model); it depended upon B, T\_Ak, BC (The fifth model).

**Table 5.** Regression models where pH is a dependent variable

Model	Standardized Coefficients	Sig.	
			Beta
1.	Constant	7.444	0.000
	B	0.175	0.049
	CT	-0.261	0.012
2.	Constant	7.513	0.000
	B	0.273	0.007
	CT	-0.261	0.012
	T_Ak	-0.199	0.028
3.	Constant	7.720	0.000
	B	0.277	0.006
	CT	-0.261	0.012
	T_Ak	-0.199	0.028
	BC	3.811	0.000
4.	Constant	7.614	0.000
	B	0.155	0.049
	CT	-0.075	0.366
	T_Ak	-3.920	0.000
	BC	3.811	0.000
5.	Constant	7.571	0.000

Model		Standardized Coefficients	Sig.
		Beta	
	B	0.120	0.079
	T_Ak	-3.999	0.000
	BC	3.910	0.000

The coefficient of T\_AK is -3.95, it meant the it has a negative correlation to the pH. Moreover, if we take a look at the Model which also stratified by it producing horizon codes when the Uranium as an independent variable, and After operating the step-wise regression onto the data, then the results is as seen on the Table 6.

**Table 6.** Model where Uranium is a dependent variable

Model		Standardized Coefficients	Sig.
		Beta	
1.	Constant	14.003	0.000
	MO	0.348	0.000
2.	Constant	9.441	0.000
	MO	.377	.000
	SO4	.212	.012
3.	Constant	-13.353	.013
	MO	.258	.002
	SO4	.451	.000
	T_AK	.459	.000
4.	Constant	-12.301	.020
	MO	.362	.000
	SO4	.530	.000
	T_AK	.441	.000
	B	-.206	.029

- a. Dependent Variable: U
- b. Predictors: (Constant), MO
- c. Predictors: (Constant), MO, SO4
- d. Predictors: (Constant), MO, SO4, T\_AK
- e. Predictors: (Constant), MO, SO4, T\_AK, B

**Table 7.** R-square when Uranium is a dependent variable

Model	R	R Square	Adjusted R Square
1	.348 <sup>a</sup>	.121	.114
2	.407 <sup>b</sup>	.166	.152
3	.545 <sup>c</sup>	.297	.280
4	.570 <sup>d</sup>	.324	.302

**Table 8.** Step-wise regression models by its producing Horizon codes when Uranium as a dependent variable

Model	F	Sig.
Regression	17.264	.000 <sup>b</sup>
Residual		
Total		
Regression	12.302	.000 <sup>c</sup>
Residual		
Total		
Regression	17.343	.000 <sup>d</sup>
Residual		
Total		
Regression	14.642	.000 <sup>e</sup>
Residual		
Total		

As seen on Table 7 and 8 that the regression model of Uranium for Producing Horizon code of TPO as follows:

$$U = 14.003 + 0.348 MO \tag{6}$$

The model for Producing horizon code of TRD is as follows:

$$U = 9.441 + 0.377 MO + 0.212 SO4 \tag{7}$$

The model for Uranium of Producing Horizon code of POQ is as follows:

The model for Uranium with Producing Horizon code of PGWC is as follows:

$$U = -12.301 + 0.362 MO + 0.530 SO4 + 0.441 T\_Ak - 0.206 B \tag{8}$$

The Uranium regression model of Producing Horizon of PGEB was not created by the SPSS Software. Perhaps, the data related to this PGEB suffered singularity in its variance-covariance matrix, then it could not be found the coefficients of each parameter.

As seen on Table 7 and 8 that all four models were significant for every code of Producing horizon where the adjusted R square is different in accordance with the producing horizon code. However, biggest and involving more variables, such as: Constant, MO, SO4, T\_AK, B.

Where, *U* is Uranium. *MO* is Molybdenum. *SO4* is Sulfate. *T\_AK* is Total Alkalinity. *B* is Boron. The interpretation of equation (8) is that the Uranium was depended upon the existence of Molybdenum, Sulfate, Total Alkalinity, and Boron. However, the Uranium has negative correlation with Boron. Over the last half-century, uranium has become one of the most sought-after minerals on the planet due to from energy to healthcare, this revolutionary substance has applications in a variety of industries. As such, let's take a few moments to consider the uranium mining and extraction process as well as its immense value throughout the world [10].

In Canada [12], Uranium was available in Natural and fortified water as the trace elements in water as also available [12] in Natural Sediments and Soils as trace elements in sediments along with Molybdenum. Most

volcanogenic uranium deposits are deeply buried and very hard to locate. A prospecting method using geoelectrochemical probes has been proven in practice to be an effective way of pinpointing the locations of these hidden deep uranium deposits. The efficacy of the method has been verified through analyses of two carefully selected known uranium deposits: the Xiangshan and Shenyuan uranium deposits in southeastern China [13].

Analyses of geoelectrochemical probe samples presented in transections crossing the ore bodies revealed distinct anomalies of U and Mo distribution [13]. A recent study led by scientists at the Department of Energy's SLAC National Accelerator Laboratory helps describe how the contaminant cycles through the environment at former uranium mining sites and why it can be difficult to remove. Contrary to assumptions that have been used for modeling uranium behavior, researchers found the contaminant binds to organic matter in sediments. The findings provide more accurate information for monitoring and remediation at the sites. Decades after a uranium mine was shuttered, the radioactive element can still persist in groundwater at the site, despite cleanup efforts [5].

## Conclusion

The stepwise regression gave the results that the pH is depended upon Boron for the producing horizon code of TPO. The pH is depended upon Boron and Conductivity (CT) for the producing horizon code TRD, but has negative correlation with Conductivity. The pH in the producing horizon code of POQ is depended upon Boron, Conductivity, and Total Alkalinity. In accordance with the Modus Ponens then if it is available pH in the surrounding of groundwater of allegedly containing uranium, then will be available the Boron, Conductivity, and Total Alkalinity, and Bicarbonate. Also, in this model the pH has negative correlation with Conductivity and T\_Ak., Total alkalinity, and Bicarbonate in the groundwater. The stepwise regression was also giving the results for Uranium as a dependent variable. The Uranium was depended upon Molybdenum, Sulfate, Total alkanity, and Boron. In addition, if uranium is available, then Molybdenum, Sulfate, Total Alkanlinity, and Boron will be available either. However, the association rules were not uncovered the involvement of Uranium in the groundwater samples.

## References

1. ATSDR. *Toxicological profile for uranium*. US Department of Health and Human Services, Public Health Service; Atlanta: Agency for Toxic Substances, Disease Registry (1999)
2. <https://www.mae.gov.nl.ca/waterres/cycle/groundwater/well/uranium.html>. Accessed at 4 November 2018
3. SCGWA. South Carolina Groundwater Association. (2009)

4. <http://scgwa.org/pages/industrynews/economy.html> access on 20 July 2018.
4. S.E.Wagner, J.B. Burch, J. Hussey, T. Temples, S. Bolick-Aldrich, C. Mosely, et al., *Soil zinc content, groundwater usage, and prostate cancer incidence in South Carolina*. *Cancer Causes Control*. (2009); 20(3):345–353. doi:10.1007/s10552-008-9248-0. [PubMed: 18949566]
5. SLAC National Accelerator Laboratory. 2017. <https://www.sciendaily.com/releases/2017/02/170202163234.htm>. Accessed at 4 November, 2018
6. R. Agrawl, T. Imielinski, A. Swami, *Mining association rules between sets of items in large databases*, Proceeding of SIGMOD-1993. Pp.207-216 (1993)
7. I. Lukman, GMDH, *Association Rules and Regression of Environmental Economics Data*, Lambert Academic Publishing (2016)
8. <https://people.duke.edu>. Step-wise regression. Accessed at 25 October 2018
9. I. Lukman, Mengidentifikasi Populasi Airtanah Dengan Menggunakan Pertambangan Data Dari Data Proyek Evaluasi Uranium Departemen Energi Amerika Serikat. Konferensi Nasional Teknik Sipil dan Infrastruktur-I, Jurusan Teknik Sipil Universitas Jember, 30 Oktober (2017)
10. C.E. Nichols, V.E. Kane, M.T. Browning, and G. W. Cagle, *Northwest Texas Pilot Geochemical Survey*, Union Carbide, Nuclear Division Technical Report (1976)
11. D. F. Andrews, and A. M. Herzberg, *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag (1985)
12. <https://portal.a2la.org/scopepdf/2867-01.pdf> accessed at 25 October 2018.
13. P. Liu, X. Luo, M. Wen, J. Zhong, F. Ouyang. 2018. Geoelectrochemical anomaly prospecting for uranium deposits in Southeastern China. *Applied Geochemistry*, Volume 97, October 2018, pp : 226-237. Accessed at 3 November, 2018.
14. <https://www.generalkinematics.com/blog/5-fascinating-facts-world-uranium-mining/> accessed at 25 October 2018