Nonlinear Processes in Geosciences Division EGU

Vienna 2020

Application of non-paramertic trend analysis to concentration time series data

Artúr Kőhler, kohler@ftr2000.hu

UNIVERSITAS BUDAPESTINENSIS DE ROLANDO EÖTVÖS NOMINATA · 1635

1 – Data

Groundwater contamination often proves to be a persistent feature of the affected groundwater regime. Persistent plumes regularly monitored. Concentration data gained by repeated sampling of monitoring wells and laboratory analyses of the samples. Concentration data are ordered into time series. Plume behaviour is assessed by concentration trend evaluation. Data of a 14-year-long monitoring of a halogenated aliphatic hydrocarbon plume is used

-Nondetect values replaced by 65% value of detection limit[1]

2- Trend evaluation

Trends are evaluated using either parametric (like linear regression using least squares) or non-parametric methods.

Mann-Kendall statistic is one of the most widely used non-parametric method.

Mann-Kendall statistics is generated comparing the data in the time series to each other, as follows:

3- Mann-Kendall test

$n = number\ of\ data\ in\ time\ serie$

$j, k\ denotes\ the\ jth\ and\ the\ k^{th}\ member\ of\ the\ serie$

$j = 2,3 \dots n, \qquad k = 1,2 \dots n-1$

$for\ all\ j > k$:

$sign = 1\ if\ xj - xk > 0$

$sign = -1\ if\ xj - xk < 0$

$sign = 0\ if\ xj - xk = 0$

$$S = \sum sign(xj - xk)$$

Significance of S is determined using ‚Z' statistic. Var (S) is calculated, and corrected for tied groups

$$VAR(S) = \frac{1}{18}\left[n(n-1)(2n-5) - \sum t_p(tp-1)(2tp+5)\right]$$

Where t_p is the number of data in a tied group (i.e. a group of equal values)

$$Z = \frac{S-1}{\sqrt{VAR(S)}}\ if\ S > 0;\quad Z = \frac{S+1}{\sqrt{VAR(S)}}\ if\ S < 0;\quad Z = 0\ if\ S=0$$

4- Method

-Concentrations reported by a laboratory are not deterministic numbers. They represent a range, specified by laboratory-given measurement precision. This is usually a ± percentage of the reported concentration

-a 10% precision means that the reported concentration is within a ± 10% percentage range of the reported concentration

Laboratory reported concentrations are considered equal, if their ± 10% ranges overlap.

If laboratory reported concentrations

$$C_{LR1} < C_{LR2}$$

Then if

$$C_{LR1} \times 1{,}1 > C_{LR2} \times 0{,}9$$

Then

$$C_{LR1} \sim C_{LR2}$$

-Comparing reported concentrations with or without the precision range may result in different S statistic, and Z statistic, for two reasons

- S will be directly different because the comparison differences

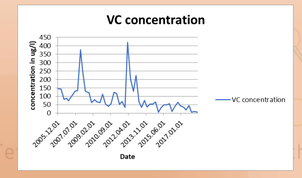
- Var (S) will be different because of the difference in tied groups

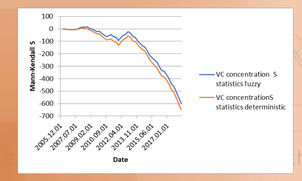
Therefore significance may be different wich results in differences in trend perceptions

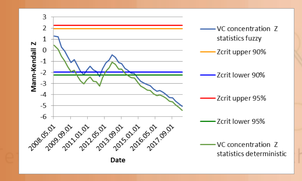
The two different approach are named here in this presentation „deterministic" and „fuzzy"

5- Results (examples)

Sample ID	Date	VC concentration	S	S_{det}	Z_{fuzzy}	Z_{det}
AE-4-5	2005.12.08	**145**				
AE-4-6	2006.03.29	**143**	0	-1		
AE-4-7	2006.06.13	*82,3*	-2	-3		
AE-4-8	2006.09.19	*89,6*	-4	-4		
AE-4-9	2006.11.30	*74,5*	-6	-8		
AE-4-11	2007.06.12	**129**	-3	-7		
AE-4-12	2007.09.10	**135**	0	-5		
AE-4-13	2007.12.07	**376**	7	2		
AE-4-14	2008.02.29	**264**	13	8		
AE-4-15	2008.05.28	**129**	14	6	1,27069	0,44901
AE-4-16	2008.09.03	**122**	15	2	1,21395	0,07809
AE-4-17	2008.12.03	*62*	5	-9	0,29623	-0,54987
AE-4-18	2009.03.05	*80,2*	-2	-17	-0,0657	-0,97796
AE-4-19	2009.06.11	*67,4*	-11	-28	-0,58857	-1,48033
AE-4-20	2009.09.17	*63,5*	-22	-40	-1,0952	-1,93237
AE-4-21	2009.12.10	**113**	-19	-41	-0,83533	-1,80273
AE-4-22	2010.03.22	*52,8*	-33	-57	-1,35184	-2,30875
AE-4-23	2010.06.17	*41,2*	-50	-74	-1,89587	-2,76706
AE-4-24	2010.09.16	*55,8*	-62	-88	-2,19118	-3,04561
AE-4-25	2010.12.07	**123**	-54	-83	-1,80799	-2,66183
AE-4-26	2011.03.23	**117**	-47	-81	-1,4503	-2,41686



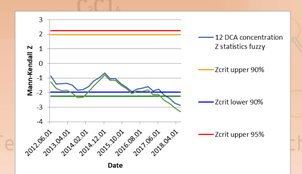




The reason for differences in „S" is obvious, a deterministicly different concetrations will be equal concentrations as fuzzy

Identical colors indicate tied groups. Depending on the comparison base (i.e. the last data in the time serie) the number and the content of earlier established tied groups may change

Sample ID	Date	12 DCA concentration	S	S_{det}	Z_{fuzzy}	Z_{det}
AE-20-22	2010.03.22	*1,3*				
AE-20-23	2010.06.17	*2,7*	1	1		
AE-20-24	2010.09.16	*2,8*	2	3		
AE-20-25	2010.12.07	*2,6*	3	2		
AE-20-26	2011.03.23	*2,7*	4	3		
AE-20-27	2011.06.09	*2,3*	5	0		
AE-20-28	2011.09.22	*1,7*	1	-4		
AE-20-29	2011.12.01	0,195	-6	-11		
AE-20-30	2012.03.22	*1,9*	-8	-13		
AE-20-31	2012.06.22	*1,9*	-10	-15	-0,86736	-1,26234
AE-20-32	2012.09.18	*0,7*	-18	-23	-1,39897	-1,72317
AE-20-33	2012.12.13	*1,6*	-20	-28	-1,38818	-1,86023
AE-20-34	2013.03.27	*1,7*	-22	-31	-1,36892	-1,84057
AE-20-35	2013.06.01	*1,4*	-27	-38	-1,48957	-2,03473
AE-20-36	2013.09.25	*1*	-37	-48	-1,84838	-2,33449
AE-20-37	2013.12.19	*1,6*	-38	-52	-1,77948	-2,30551
AE-20-38	2014.03.18	*1,9*	-37	-48	-1,58738	-1,9471



While tied groups, once established in the deterministic approach will persist and their influence is stable or increasing with increasing ‚n', the fuzzy aproach can produce a smaller correction for larger ‚n'

6- Conclusions

6.1 Fuzzy aproach will results in larger correction terms in VAR (S) calculations, however this will not necessarily reflected in ‚Z' statistic

6.2 There is no obvious reason, why Mann-Kendall trend analysis produces different trends for certain time series using determinsitic and fuzzy approach

6.3 Deviation of fuzzy ‚S' can be both direction from determinsictic ‚S' (regarding zero ‚S' as an absolut no trend), therefore using the fuzzy aproach has not necessarily dampening effect.

6.4 Time series data structure needs to be looked into to explain the fuzzy ‚S' and ‚Z' behaviour

References:

[1] J. A. Martin-Fernandez, C. Barcelo-Vidal,V. Pawlowsky-Glahn Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation Math. Geol 2003 Vol 35 No. 3