Hydrol. Earth Syst. Sci. Discuss., 3, S1425–S1431, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S1425/2006/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map" *by* L. Peeters et al.

L. Peeters

luk.peeters@geo.kuleuven.be

Received and published: 10 November 2006

The authors would like to thank the referee for reviewing the manuscript and providing feedback.

General Remarks:
1.What is the question better answered by GEO3DSOM compared to standard SOM?....

In general the standard SOM clusters similar data patterns, while GEO3DSOM clusters similar data patterns provided they are spatially close. The standard SOM will provide insight into phenomena affecting the entire study area, while the semi-hard constraints on location imposed by GEO3DSOM result in a localization of areas where certain phenomena occur, in addition to a description of the phenomena itself. With regards to the case-study, the results of the standard SOM can indeed be better evaluated and interpreted in terms of the geochemical processes acting in the concerned aquifers. The spatial coherent groups delineated by the GEO3DSOM can however be regarded as distinct hydrochemical facies in the aquifer affected by the same hydrogeological and geochemical conditions, provided the aquifers are relatively homogeneous and no major faults or conduits exist which can lead to abrupt changes in the groundwater chemistry. The GEO3DSOM-delineated groups can contribute in the understanding of the hydrogeological system and subsequently be used in the refinement of hydrogeological models. If the goal of the exploratory data analysis is to evaluate the processes affecting groundwater chemistry, the standard SOM will thus provide the most insightful results, while the GEO3DSOM will provide more coherent results if the goal is to delineate zones with similar groundwater geochemistry. The GEO3DSOM is thus not to be regarded as a substitute for the standard SOM. The combined use of both techniques will provide the most information about the dataset under study.

2.What were the technical details of the SOM and GEO3DSOM?
a)number of grid nodes

The number of grid nodes to be used in a SOM-analysis can be considered as a trade-off between representation accuracy and generalization accuracy. A small number of grid nodes will result in a high quantization error and well-defined clusters, while a large number of nodes result in a low quantization error and, in the most extreme case, a cluster for each data sample. In this study a trial and error-method is used to choose the number of grid nodes. Tables 1a and 1b show the evolution of quantization,

HESSD
3, S1425–S1431, 2006

EGU

topologic and geographic error for maps with different numbers of grid nodes. The grid configuration is rectangular for each of the maps, based on a 3 by 4 configuration. Each of the resulting maps were visually evaluated for their clustering ability. As can seen from tables 1a and 1b, the quantization error rapidly decreases with increasing number of grid nodes, while the topologic and geographic error converge to a stable value. The best compromise between clustering ability and quantization error was found for the 20 by 15 nodes configuration (300 nodes).

Table 1a: Quantization and topologic error for different SOM-grid configurations

units	map configuration	qe	te
12	3 by 4	0.470	0.221
48	6 by 8	0.319	0.137
108	9 by 12	0.230	0.107
192	12 by 16	0.166	0.053
300	15 by 20	0.118	0.069
432	18 by 24	0.069	0.038
588	21 by 28	0.029	0.069
768	24 by 32	0.009	0.046
972	27 by 36	0.005	0.099
1200	30 by 40	0.004	0.069

Table 1b: Quantization, topologic and geographic error for different GEO3DSOM-grid configurations

HESSD

3, S1425–S1431, 2006

units	map configuration	qe	te	ge
12	3 by 4	0.466	0.198	0.113
48	6 by 8	0.324	0.397	0.062
108	9 by 12	0.255	0.450	0.044
192	12 by 16	0.194	0.366	0.034
300	15 by 20	0.135	0.405	0.023
432	18 by 24	0.086	0.336	0.017
588	21 by 28	0.037	0.313	0.007
768	24 by 32	0.015	0.298	0.003
972	27 by 36	0.006	0.359	0.001
1200	30 by 40	0.002	0.374	0.000

b)number of iteration steps

In order to determine the adequate number of iterations for the SOM-analysis, a comparable methodology is applied. The results are shown in tables 2a and 2b. The number of iterations shown are the total number of iterations after rough and fine training. After 1000 iterations (500 rough training and 500 fine training) the topologic and geographic error are stabilized, while the quantization error only decreases slightly.

Table 2a: Quantization and topologic error evolution for number of iterations (SOM)

HESSD

3, S1425–S1431, 2006

iterations	qe	te
200	0.227	0.076
400	0.178	0.053
600	0.152	0.076
800	0.144	0.038
1000	0.115	0.053
1200	0.103	0.046
1400	0.094	0.023
1600	0.083	0.023
1800	0.094	0.038
2000	0.081	0.023

Table 2b: Quantization, topologic and geographic error evolution for number of iterations (GEO3DSOM)

iterations	qe	te	ge
200	0.249	0.412	0.034
400	0.226	0.397	0.035
600	0.172	0.382	0.028
800	0.153	0.420	0.027
1000	0.136	0.389	0.024
1200	0.121	0.382	0.022
1400	0.110	0.366	0.020
1600	0.094	0.435	0.018
1800	0.096	0.435	0.017
2000	0.092	0.351	0.018

c)k-value

A sensitivity analysis with regards to the k-value reveals that with increasing k-value, the geographic error increases rapidly and the quantization error decreases (table 3).

HESSD

3, S1425–S1431, 2006

Once again a compromise has to be found between the weight given to the geographic coordinates and the overall quantization. After visual examination of the ability of the different maps to represent the data, a k-value of 4 produced the best results both with respect to low quantization and geographic error and with respect to visualizing and grouping of the data.

Table 3: Influence of k-value on quantization and geographic error for GEO3DSOM

k	qe	ge
0	0.199	0.015
1	0.149	0.019
2	0.151	0.022
3	0.143	0.023
4	0.137	0.025
5	0.139	0.024
6	0.144	0.025
7	0.145	0.027
8	0.137	0.024
9	0.130	0.024
10	0.133	0.025

The results of the sensitivity analysis of the number of iterations, number of units and the k-values will be incorporated in the revised manuscript together with graphs of tables 1 to 3.

Details:

The modifications suggested by the referee will be carried out in the revised manuscript.

HESSD

3, S1425–S1431, 2006

p.5 table 3: The differences in error measures are indeed rather small. The visualization of the results however shows the improved ability of the GEO3DSOM to distinguish between the pre-defined groups.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 1487, 2006.