

Igor Kuzmanovski · Zlatko Zografski  
Mira Trpkovska · Bojan Šoptrajanov · Viktor Stefov

## Simultaneous determination of composition of human urinary calculi by use of artificial neural networks

Received: 20 December 2000 / Revised: 30 March 2001 / Accepted: 9 April 2001

**Abstract** A new chemometric method, which uses artificial neural networks (ANN), is presented for determination of the composition of urinary calculi. The selected constituents were whewellite, weddellite, and uric acid from which approximately 40% of the urinary calculi obtained from Macedonia patients are composed. The results for the synthetic mixtures were better than those obtained by partial least squares (PLS) regression or by the principal component regression (PCR), because neural networks have better prediction capacity. The generalization abilities of the optimized neural networks were checked using the standard addition method on carefully selected real natural samples.

### Introduction

Knowing the exact chemical composition of urinary calculi is of great importance for clinical chemists, not only because of its relationship with dietary and other health factors, but also for prevention of further urolithiasis. The procedures of classical chemical analysis are more time-consuming than instrumental techniques [1, 2] and can lead to incorrect conclusions because of the low selectivity of the reagents used.

Infrared spectroscopy is one of the most suitable instrumental techniques for the analysis of calculi, and the

most often used. Its advantages compared with classical chemical analysis, and with other instrumental techniques are:

1. it is possible to obtain information about the exact chemical nature of the constituents;
2. the analysis takes less than 10 min; and
3. only approximately 1 mg of sample is required. (This can, if necessary, be reduced to 20–30 µg), which is particularly important for studying nucleation processes.

Several computerized methods based on infrared spectroscopy has been developed with the aim of minimizing the role of the analyst in the decision-making process [2–6]. Among chemometric techniques, partial least squares (PLS) regression has been used for the determination of the composition of the calculi [7]. The techniques of partial least squares (PLS) and of principal component regression (PCR) give the best results when the relationship between absorbance and the mass fraction of the constituents of a given mixture is almost linear [8–10]. If any non-linearity appears in the spectral response this type of regression produces less reliable results. In such cases artificial neural networks (ANN) have proven to be more efficient [11, 12].

The work described herein was performed to develop a method for determination of the composition of urinary calculi consisting of whewellite ( $\text{CaC}_2\text{O}_4 \cdot \text{H}_2\text{O}$ ), weddellite ( $\text{CaC}_2\text{O}_4 \cdot 2\text{H}_2\text{O}$ ) and uric acid, using ANN as a tool. These three constituents were chosen because our recent study [13] has shown that approximately 40% of analyzed urinary calculi from patients from Macedonia consist of whewellite, weddellite and/or uric acid<sup>1</sup>. The results obtained using ANN were compared with those obtained using PCR and PLS regression.

At a later date it is planned to extend the studies to other combinations of urolithic constituents also.

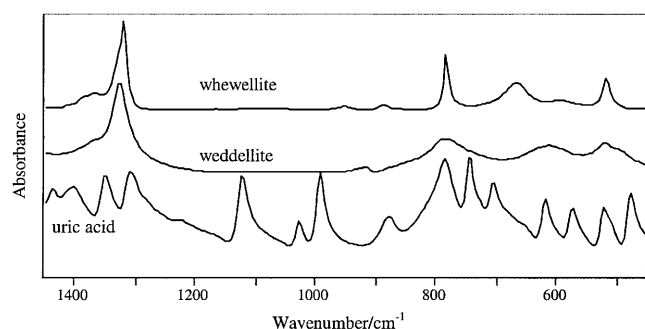
Presented at EUROANALYSIS XI, European Conference on Analytical Chemistry, Lisbon, Portugal, September 3–8, 2001

I. Kuzmanovski (✉) · M. Trpkovska · B. Šoptrajanov · V. Stefov  
Institut za hemija, PMF, Univerzitet “Sv. Kiril i Metodij”,  
PO Box 162, 1001 Skopje, Macedonia  
e-mail: shigor@iunona.pmf.ukim.edu.mk

B. Šoptrajanov  
Makedonska akademija na naukite i umetnostite, 1000 Skopje,  
Macedonia

Z. Zografski  
Elektrotehnički fakultet, Univerzitet “Sv. Kiril i Metodij”,  
PO Box 574, 1001 Skopje, Macedonia

<sup>1</sup> Some of the calculi analyzed contained just one, whereas the rest were binary or ternary mixtures of these compounds.



**Fig. 1** Infrared spectra of whewellite, weddellite, and uric acid in the 1450–450  $\text{cm}^{-1}$  region

## Experimental

The mixtures used for optimization of the method were prepared from whewellite, weddellite, and uric acid. Whewellite and weddellite were synthesized according to procedures found in the literature [14, 15] and uric acid was a Merck product. Their spectra (Fig. 1) were compared to those in the digital database of infrared spectra of constituents of urinary calculi from Dao and Daudon [16]. Fifty-one different mixtures were prepared containing one to three of these components. The design of the mixtures was carefully chosen to cover the entire range from pure compounds to two- and three-component mixtures containing different amounts of each constituent.

The samples for recording the infrared spectra were prepared as KBr pellets. Each pellet was prepared from 1 mg homogenized mixture and 250 mg spectroscopy-grade KBr. The same procedure was used for recording the spectra of real samples. The digital spectra were recorded at room temperature in the 1450–450  $\text{cm}^{-1}$  region on a Perkin–Elmer System 2000 FTIR instrument (with resolution of 4  $\text{cm}^{-1}$  and sampling interval of 4  $\text{cm}^{-1}$ ), with 32 scans for the samples and 32 scans for the background (atmosphere and 250 mg KBr pellet).

**Data analysis.** Three independent data sets were prepared (training, validation, and test set) according to the mixture design shown in Fig. 2; this was necessary to improve the generalization capabilities of the models used. The first set (training set) was used to compute and update the network weights and biases. The performance of the network, during training, was monitored by use of the validation set, knowing that the validation error normally decreases during the training process. When the network starts to overfit the data, the error in the validation set starts to increase. When the validation error increases in five consecutive iterations, the training of the network stops and the weights and biases with minimum validation error are restored.

To select the optimum number of factors in the PCR and PLS regression models, predictive residual sum of squares (PRESS)

was calculated by use of the training set (cross-validation with the leave-one-out technique) and with the validation set:

$$\text{PRESS} = \sum_{i=1}^n (\bar{w}_i - w_i)^2$$

where  $n$  is the number of samples in the set;  $\bar{w}_i$  and  $w_i$  are, respectively, the calculated and known mass fractions of that constituent in the  $i$ -th sample. The validation set was used only for estimation of the optimum number of factors.

The performances of the optimized networks and the optimum PCR and PLS regression models were tested using the test set comprising artificially prepared mixtures. Another data set prepared from spectra of carefully selected calculi was used to check the performance of the method on real samples. The composition of these samples was determined by the networks with the best performances and then checked using the method of standard additions.

The absorbance values of the recorded spectra were reduced to 50 data points, normalized to unit area, and stored in a single matrix as ASCII data. Each row of the matrix corresponded to different wavenumbers of spectra of the mixtures and each column consisted of different spectra of the mixtures normalized to unit area. The mass fractions of the constituents of the mixtures were also stored in another matrix in which columns corresponded to different mixtures, and rows to the mass fractions of the components.

Three-layered cascade-forward neural networks with sigmoid transfer function in the hidden layer and linear transfer functions were used in the output layer. The training set was employed to adjust the connection between weights and biases using the Levenberg–Marquardt algorithm [17] for back-propagation of error. The validation set was used to monitor the performance of the network during the training process. The training of the network was stopped when the performance goal of  $10^{-3}$  for the sum-squared error (SSE) was achieved, or if in five consecutive iterations the error of the validation set increased. In the latter circumstances the weights and biases of the final network were those with the smallest error in the validation set.

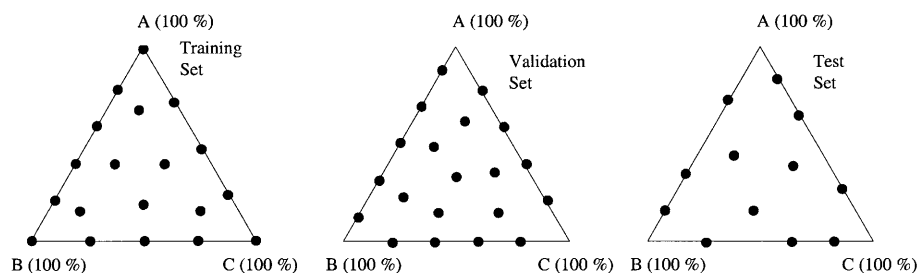
The predictive power of ANN, PCR and PLS regression were compared using the test set which consisted of 12 real samples with exactly known composition. The standard error of prediction (SEP) was used as an indicator of the prediction error. This quantity is defined as

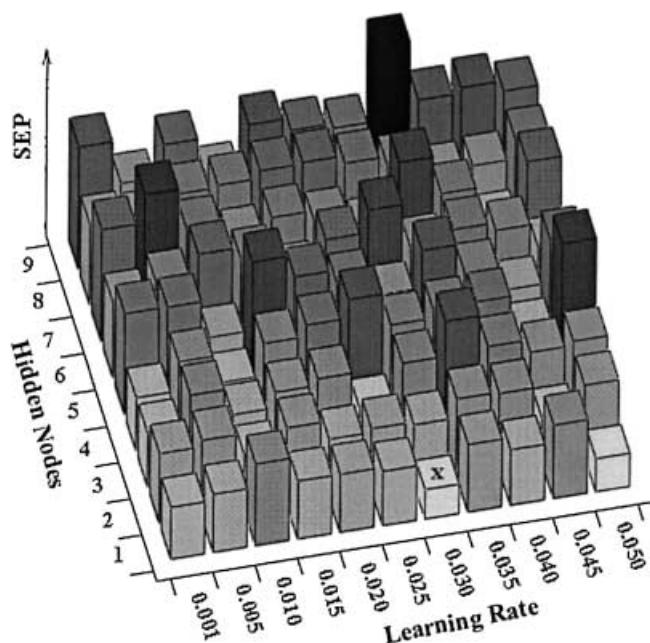
$$\text{SEP} = \left[ \sum_{i=1}^n (\bar{w}_i - w_i)^2 / n \right]^{1/2}$$

where  $\bar{w}_i$  is the predicted mass fraction of the analyzed constituent in the  $i$ -th sample,  $w_i$  is the known mass fraction of that constituent in the  $i$ -th sample, and  $n$  is the number of the samples in the test set.

Data processing was performed with the Matlab programming package [17] using the Neural Network Toolbox for optimization of the neural networks and the Chemometric Toolbox for PCR and PLS regression.

**Fig. 2** Mixture design for the mixtures used for optimization of the neural networks: **A** whewellite, **B** weddellite, **C** uric acid





**Fig. 3** Standard error of prediction (SEP) as a function of the number of nodes in the hidden layer and the learning rate. Lighter color (shorter bar) indicate better performances (smaller SEP value). The combination of the number of hidden nodes and learning rate which gives best performances is marked with “x”

## Results and discussion

### Optimization of the artificial neural networks

Different network architectures were used to find that with the best predictive power. During the optimization the number of nodes in the hidden layer was varied (from 1 to 9) and different learning rates (0.001, 0.005, 0.010, 0.015, ..., 0.050) were employed. The training was repeated ten times for each combination of nodes in the hidden layer and the learning rates. The reason for doing so was the expectation that this would reduce the effect of random initialization of the weights and biases. The values of the standard error of prediction for each constituent in each of the repeated training of the networks with a specific architecture, and average SEP values, were calculated. Figure 3 shows the sum of average SEP values for the mixture constituents as a function of the number of nodes in the hidden layer and the learning rate. It is obvious that networks with smaller number of nodes in the hidden layer have better performances, whereas no change in the network performances as a function of learning rate can be discerned. The network with the best performances (the smallest sum of average SEP for all constituents) for prediction of the content of whewellite, weddellite, and uric acid (Fig. 3) is that with one node in the hidden layer, trained using the learning rate of 0.030 (marked with a sign “x” in the middle).

**Table 1** Comparison of the SEP values obtained by principal component regression (PCR), partial least squares (PLS), and artificial neural networks (ANN)

Constituent	Standard error of prediction (SEP)		
	PCR	PLS	ANN
Whewellite	0.070 (8)	0.069 (3)	0.043
Weddellite	0.084 (5)	0.077 (2)	0.051
Uric acid	0.058 (8)	0.056 (8)	0.019

### PCR and PLS optimization

The same data sets were used for principal component regression, PCR. Each component was analyzed separately from others. The SEP values for each constituent are given in Table 1. The optimum number of factors (determined from the validation set) used for calculation of SEP values for test set is shown in brackets after the corresponding SEP value for each component.

### Comparison of the results

An overall comparison of the results is presented in Table 1. The results imply that, of the two factor-based methods, PLS regression has slightly better predictive power than PCR. The superiority of the ANN is, on the other hand, obvious and improvement of the predictive power for weddellite is at least 30%. Also apparent from Table 1 is that the differences between SEP values for uric acid and the oxalates are not as substantial as for of PCR and PLS regression. One possible explanation of this could be that for highly overlapped spectra, such as those of calcium oxalate monohydrate and calcium oxalate dihydrate, PCR and PLS regression methods are not as good as ANN at “recognizing” the mass fractions of the constituents.

### Analysis of real samples

The results obtained were checked using the standard addition method. Exactly known masses of the pure substances (whewellite, weddellite, and/or uric acid) were added to each of ten selected calculi. For illustration, the calculated and predicted (using ANN) mass fractions of the analyzed constituents after standard addition are presented in Table 2.

The absolute differences between the calculated and predicted mass fractions are presented in Table 3. It is apparent that the value of the absolute difference between the calculated and predicted mass fraction is usually smaller than 0.05. For samples 1, 3, 4, 8, and 9 and some of the components the prediction discrepancies of the methods were larger than 0.05. This finding for samples 1, 4, and 9 might be because all three methods were optimized using pure synthesized substances and their artificial mixtures, which are only models for real urinary calculi synthesized

**Table 2** Comparison of the calculated and predicted mass fractions (with network giving best performances) for the ten analyzed samples after standard addition

Sample number	1	2	3	4	5	6	7	8	9	10
Calculated mass fractions after standard addition										
Whewellite	0.089	0.031	0.033	0.291	0.285	0.668	0.094	0.216	0.297	0.775
Weddellite	0.488	0.171	0.554	0.574	0.283	0.245	0.057	0.298	0.675	0.211
Uric acid	0.423	0.798	0.478	0.135	0.432	0.087	0.849	0.487	0.029	0.014
Predicted mass fractions by ANN after standard addition										
Whewellite	0.128	0.035	0.038	0.339	0.322	0.627	0.106	0.143	0.351	0.742
Weddellite	0.498	0.195	0.523	0.532	0.320	0.278	0.065	0.336	0.643	0.245
Uric acid	0.387	0.816	0.561	0.157	0.385	0.099	0.854	0.549	0.034	0.016

**Table 3** Differences between calculated and predicted mass fractions for the real samples after standard addition

Sample number	1	2	3	4	5	6	7	8	9	10
ANN										
Whewellite	-0.039	-0.004	-0.006	-0.048	-0.038	0.041	-0.012	0.072	-0.054	0.032
Weddellite	-0.010	-0.024	0.032	0.041	-0.037	-0.033	-0.007	-0.038	0.031	-0.034
Uric acid	0.036	-0.018	-0.082	-0.022	0.047	-0.012	-0.006	-0.063	-0.005	-0.002
PCR										
Whewellite	0.039	-0.005	-0.037	-0.068	-0.029	0.001	0.012	0.076	-0.031	-0.049
Weddellite	-0.062	-0.006	0.045	0.033	-0.034	-0.041	0.014	-0.020	-0.032	0.009
Uric acid	0.029	0.029	-0.057	-0.053	-0.022	-0.050	0.006	-0.058	-0.023	0.033
PLS										
Whewellite	-0.072	-0.012	-0.017	-0.033	-0.032	0.045	-0.015	0.061	-0.038	0.050
Weddellite	0.022	-0.009	0.072	0.053	-0.020	-0.032	-0.003	-0.014	0.027	-0.037
Uric acid	0.027	0.006	-0.075	-0.030	0.035	-0.027	0.009	-0.067	-0.016	-0.049

**Table 4** Relative standard deviation (%) of the predicted mass fractions of real samples, obtained using all three techniques, after standard addition

Number of sample	1	2	3	4	5	6	7	8	9	10
ANN										
Whewellite	3.13	0.42	6.87	0.05	1.40	0.75	0.66	0.67	1.22	0.37
Weddellite	0.31	1.40	0.89	1.22	2.04	2.10	14.51	0.92	1.73	6.21
Uric acid	0.70	0.08	1.47	7.41	0.11	17.62	0.08	0.88	27.03	24.50
PCR										
Whewellite	13.60	2.05	10.35	2.00	2.97	2.62	2.95	4.24	2.65	2.62
Weddellite	2.67	7.65	3.26	5.67	4.14	5.54	4.22	11.67	4.28	8.77
Uric acid	3.30	1.64	4.24	11.03	1.24	23.19	1.15	5.57	31.30	47.44
PLS										
Whewellite	13.60	2.02	12.30	3.19	8.03	3.43	4.06	3.34	3.35	9.71
Weddellite	2.67	6.94	3.81	5.05	5.29	5.18	6.16	14.80	2.64	6.65
Uric acid	1.92	4.84	6.14	5.13	3.86	18.98	18.98	6.41	25.24	47.04

in the human urinary tract under physiological conditions. More significant discrepancies are found for all three methods for samples 3 and 8; this might be because of weighing inaccuracy in the process of preparing the samples with standard addition.

The precision of the three methods, estimated as the relative standard deviation, was also assayed (Table 4). The results show the clear superiority of the ANN over factorial-based methods; values of the relative standard deviation were smaller than 5% for most of the components.

These values for the relative standard deviation obtained by the ANN method satisfy the criteria for analytical methods used for routine analyses according to Püschel [18].

## Conclusion

Artificial neural networks were used as a tool for prediction of the mass fraction of the constituents of three-component urinary calculi composed of whewellite, weddellite,

and uric acid. Networks with smaller numbers of gave better prediction results (smaller average SEP). No clear trend in the performances of the networks as a function of the learning rate was noticed. When compared with factor-based methods use of ANN led to better results, especially for mixtures of components with highly overlapping spectra (whewellite, weddellite). In our study ANN were proven to be superior and suitable for application in determining the composition of urinary calculi composed of whewellite, weddellite, and uric acid. The extension of the work to other combinations of constituents expected or known to be present in uroliths is planned.

---

## References

1. Daudon M, Reveillaud RJ (1987) *Presse Med* 16:627–631
2. Kandil SH, Abou El Azm TA, Gad AM, Abdou MM (1986) *Comput Enhanced Spectrosc* 3:171–177
3. Hesse A, Gergeleit M, Schüller P, Möller K (1989) *J Clin Chem Clin Biochem* 27:639–642
4. Hobert H, Meyer K (1992) *Fresenius J Anal Chem* 334:178–185
5. Peuchant E, Heches X, Sess D, Clerc M (1992) *Clin Chim Acta* 205:19–30
6. Kuzmanovski I, Trpkovska M, Šoptrajanov B, Stefov V (1999) *Vib Spectrosc* 19:249–253
7. Volmer M, Block A, Wolthers BG, de Ruiter AJ, Doornbos DA, van der Slik W (1993) *Clin Chem* 39:948–954
8. Borggaard C, Thodberg HH (1992) *Anal Chem* 64:545–551
9. Thomas EV, Haaland DM (1990) *Anal Chem* 62:1091–1099
10. Beebe KR, Kowalski BR (1987) *Anal Chem* 59:1007A–1017A
11. Gemperline PJ, Long JR, Gregoriou VG (1991) *Anal Chem* 63:2313–2323
12. Zupan J, Gasteiger J (1999) *Neural Networks in Chemistry and Drug Design*. Wiley, Weinheim New York Chichester Brisbane Singapore Toronto
13. Kuzmanovski I, Trpkovska M, Šoptrajanov B (1999) *Maked Med Pregled*, 53:251–255
14. Brown P, Ackermann D, Finlayson B (1989) *J Cryst Growth* 98:285–292
15. Ackermann D, Brown P, Finlayson B (1988) *Urol Res* 16:219–220
16. Dao NQ, Daudon M, (1997) *Infrared and Raman Spectra of Calculi*. Elsevier, Paris Amsterdam Oxford New York Tokyo
17. Matlab 5.2, 1984–1998 Mathworks
18. Püschel R (1968) *Mikrochim Acta* 4:783