Reduction of the dimension of the subsurface reactive transport problem

Jan Šembera, Lukáš Zedek & Jiří Maryška
Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic

ABSTRACT
The contribution deals with reduction of dimension of transport-reaction problem using Principal Component Analysis – a method that is widely used in signal processing and other technological branches but not standard in the context of transport-reaction problem. The contribution does not bring a complete methodology of reduction of dimension of a general problem. It shows the main ideas applied on one specific problem and shows main questions that should be solved specially for each specific problem.

RÉSUMÉ
La contribution s'occupe de la réduction de la dimension du problème transport réactif utilisant Analyse en composantes principales – une méthode utilisée souvant en traitement du signal et autres disciplines techniques, mais ne pas standarde en contexte de problèmes transports réactives. La contribution n'apporte pas de méthodologie complete de la réduction de la dimension du problème général. Elle démontre les données appliquées à une problème concrete et formules les questions principales qui doivent être résourdrées pour tout problème en particulier.

1 INTRODUCTION

The basic problem of transport problems significantly influenced by chemical reactions is a big dimension (number of solutes) of the problem. The acceptability or unacceptability of the system dimension is given by the specific solved problem. Basic approach to dimension reduction used by chemists is classification of solutes to primary ones and marginal ones and considering only main chemical reactions among primary solutes. The primary solutes significantly influence the problem solution and marginal ones somehow supplement the system or can be completely neglected. Application of such an approach can often reduce the solved problem dimension so that the primary solute concentration and main reactions are simulated and marginal solute concentration can be possibly consequently derived from simulation results.

Such an approach is efficient in case of observation of primary solute concentration development or condition changes influenced by it. If we are interested in phenomena significantly influenced by any marginal specie, we can extend the model by this specie (re-classify it as a primary one) and enlarge the problem dimension. But in case of system with many approximately equally important species application of this approach can efficiently reduce the problem dimension only with parallel reduction of simulation result quality.

Such a system takes place in the problem of long-time prediction of contamination in the site Stráž pod Ralskem after proposed in-situ neutralization. Formerly, at that place there uranium was being leached using injected sulphuric acid. Now, the site is contaminated in several hundreds meter depth by sulphuric acid and secondary leached contaminants, especially heavy metals and other toxic species. The water there is very acidic – in some parts of the site pH is less than 2. S. e. DIAMO now remediates the site. One of preliminarily proposed ways of possible remediation is neutralization in-situ injecting neutralization agent into the contaminated rock to stop further leaching of secondary contaminants. For possibility of meaningful planning of in-situ neutralization, a good model of transport and chemical interactions should be built.

In the underground, there are 22 analysed solution components. Some of them directly control the chemical processes – that is why they should be considered. The marginal solutes are more or less dangerous contaminants and so they should be observed and their spread and balance should be predicted as precisely as possible. The current computational capacity and algorithmic demandingness of the transport problem for such a long-time large-area simulation allows us to simulate transport of at most five solution components.

The number of simulated components should be maintained but the dimension of the problem should be significantly reduced. It can be done using a procedure coming out from linear algebra. Let us consider the set of all executed chemical analyses of solutions as set M of vectors in 22-dimensional linear vector space V whose coordinate axes correspond to concentrations of individual solution components. Let us look for such a n-dimensional linear vector subspace V_n of space V such that it lies "the most close" to the set M, so that it minimize the error of projection E_n^2 defined as the sum of quadrates of distances of all vectors from M from their projections to V_n:

$$E_n^2 = \sum_{x \in M} || \, x - \Pi_{Vn} \, x \, ||^2. \qquad [1]$$

Here Π_{Vn} denotes the operator of orthogonal projection to space V_n. If we find the subspace V_{n0} of enough small dimension n_0 and enough small projection error E_{n0}^2, we can reduce the dimension of the transport

problem from original 22 to n_0 with omitting no solute for chemical simulations.

2 PRINCIPAL COMPONENT ANALYSIS

The problem that we solve is to find such a vector space V_n of given dimension n that is the most close to the data set M in sense of the last paragraph. The best possibility is to have an orthogonal basis of V such that the last basis vector would be of "the least importance" to the set M, i.e. the projections of members of M to the hyperplane orthogonal to it would be minimal from all possibilities. The second last basis vector would be selected from the rest of space V (the mentioned hyperplane) and it would have the same property considering the mentioned hyperplane. And so on. Such a basis would be orthogonal and the optimal space V_n would be generated by the *first* n vectors of this basis.

Looking for such a basis can be realized using Principal Component Analysis (PCA). It is a method of dimension reduction with minimal information loss often used for solution of various technological problems (e.g. data decorrelation in signal processing). It is being applied also in economy or medicine. It is based on coordinate system transformation – finding a special orthonormal basis of the space where the data are placed. The basis vectors are arranged so that the first one determines the direction containing the most possible one-dimensional information in data and the direction of last basis vector includes the minimal possible information contain.

The algorithm of Principal Component Analysis – see e.g. Smith (2002) – looks like this:

1. Arrange the data into matrix X of type $r \times s$. Each row contains one of r chemical analyses, the columns correspond to measured solutes.

2. Compute average data vector (average row of X): $x^{avg}=(x_1^{avg},...,x_s^{avg})^T$, $x_i^{avg}=(1/r)\sum_{j\in\{1,...,r\}}X_{ji}$ (X_{ji} means the element of matrix X in row j and column i); constitute the matrix $X^*=X-\mathbf{1}\cdot x^{avgT}$, where $\mathbf{1}=(1,...,1)^T$.

3. Compute the covariance matrix $C=1/(r-1)X^{*T}\cdot X^*$.

4. Compute the eigenvalues and unite eigenvectors of C, sort the eigenvectors descendent to their corresponding eigenvalues and constitute the transformation matrix T (of type $r \times s$) so that in the i^{th} column of T there is the i^{th} eigenvector of C.

5. Constitute the transformation matrix T_n (of type $r \times n$) containing n principal components (first eigenvectors of C) omitting last $s-n$ columns of matrix T.

6. Reduce the centred data (orthogonally project them to the subspace generated by the first n eigenvectors of C): $Z_n = X^* \cdot T_n$.

7. Reconstruct reduced centred data $Y^*=Z\cdot T_n^T$.

8. Reconstruct reduced original data $Y=Y^*+\mathbf{1}\cdot x^{avgT}$.

The matrix Y then includes the original data orthogonally projected to the affine space of dimension n that is the best one in the sense of minimization of the projection error E_n^2 (Eq. 1). Simplicity of projection step 6 comes from that the covariance matrix C is always symmetric positive semidefinite and so its unite eigenvectors form an orthonormal system. Matrix T is then orthogonal and its inverse is equivalent to its transposition. Similarly pseudoinverse matrix to T_n is its transposition.

Table 1. Table of projection errors E_n^2, E_n, and p_{En} for set of analyses M_{22}.

n	21	20	19	18	17	16	15
E_n^2	0.135	25.8	60.0	155	291	559	1432
E_n	0.367	5.08	7.75	12.5	17.1	23.6	37.8
p_{En} (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00

n	6	5	4	3	2	1	0
E_n^2	$3\cdot10^6$	$6\cdot10^6$	$1\cdot10^7$	$3\cdot10^7$	$4\cdot10^8$	$9\cdot10^8$	$6\cdot10^{11}$
E_n	$2\cdot10^3$	$2\cdot10^3$	$3\cdot10^3$	$6\cdot10^3$	$2\cdot10^4$	$3\cdot10^4$	$8\cdot10^5$
p_{En} (%)	0.21	0.31	0.44	0.70	2.60	3.81	100

The steps 2 and 8 are executed only to obtain better approximation of uncentred data looking for best affine subspace and they are not principal for the method. If we substitute the steps 2 and 8 by steps

2a. $X^* = X$,

8a. $Y = Y^*$,

the method will not find the best affine subspace but the best linear subspace – precisely the one that we called V_n.

3 RESULTS OF APPLICATION OF PRINCIPAL COMPONENT ANALYSIS

We have applied the algorithm to two sets of chemical analyses of solutions taken from various parts of Stráž pod Ralskem site in various times. The first set (denoted M_{22}) includes 90 complete analyses of 22 components. The second set (denoted M_6) included 638 analyses of six primary solutes.

Table 2. Table of projection errors E_n^2, E_n, and p_{En} for set of analyses M_6.

n	5	4	3	2	1	0
E_n^2	$6.3\cdot10^5$	$1.3\cdot10^7$	$6.7\cdot10^7$	$1.6\cdot10^8$	$2\cdot10^9$	$3\cdot10^{12}$
E_n	793	3591	8156	$1.2\cdot10^4$	$5\cdot10^4$	$2\cdot10^6$
p_{En} (%)	0.05	0.22	0.50	0.76	2.82	100

We have applied Principal Component Analysis to each of the two sets and enumerated the projection error E_n^2 (Eq. 1) for each dimension n. Tables 1 and 2 include the results. Here E_n is the square root of E_n^2 and p_{En} is fraction of E_n and E_0 which is a measure of the whole information in the set M_{22}.

We can see from the Tables 1 and 2 that using subspace of dimension 3 the projection error is lower than 1%. It means that the analysed data are highly correlated. Opposite observation would discredit intended procedure of dimension reduction.

3.1 Acceptability of identified subspace

After application of PCA we have observed other properties of chemical analysis projections to the selected subspace. No precise requirements for acceptability of results were defined before so we have defined two acceptability parameters: 1) Non-negativity of each component of projected vector. It comes from natural requirement of possibility of interpretation of the projected vectors and their projections as chemical analyses of solutions – concentrations of all solutes should be non-negative. 2) Proximity of the projections to the projected vector not only in l^2 norm but also in some specific scaled maximum norm $|| x-\Pi_{Vn} x ||_\alpha^2 = \max_i \alpha_i |x_i - (\Pi_{Vn} x)_i|$, where components of positive vector α correspond to importance of each solution component. It comes from requirement of restriction of differences in each component separately.

A natural choice of vector α is vector of multiplicative inverse of average concentrations of individual solutes which scales the importance of solutes considering their different typical concentrations.

Table 3. Table of projection errors $\tilde{E}_n$ and $p_{\tilde{E}n}$ for set of analyses M_{22}.

n	21	20	19	18	17	16	15
$\tilde{E}_n$	$9{\cdot}10^3$	$1{\cdot}10^4$	$2{\cdot}10^4$	$2{\cdot}10^4$	$2{\cdot}10^4$	$2{\cdot}10^4$	$3{\cdot}10^4$
$p_{\tilde{E}n}$ (%)	1.22	1.49	2.37	2.39	2.47	2.72	3.31

n	6	5	4	3	2	1	0
$\tilde{E}_n$	$4{\cdot}10^4$	$5{\cdot}10^4$	$5{\cdot}10^4$	$7{\cdot}10^4$	$8{\cdot}10^4$	$9{\cdot}10^4$	$8{\cdot}10^5$
$p_{\tilde{E}n}$ (%)	4.98	5.91	6.36	8.87	9.50	11.70	100

The second part of our definition of acceptability leads us to a slide modification of PCA algorithm: replace of steps 2a and 8a of the algorithm by steps

2b. $X^* = X \cdot \mathrm{diag}(1/x_1^{avg},..., 1/x_s^{avg})$.

8b. $Y = Y^* \cdot \mathrm{diag}(x_1^{avg},..., x_s^{avg})$.

We scale this way the data matrix X to the average concentrations of individual solutes. The algorithm then does not minimize the projection error E_n^2 (Eq. 1) but minimizes another projection error

$$\tilde{\tilde{E}}_n^2 = \sum_{x \in M} || x-\Pi_{Vn} x ||_{2,\alpha}^2 \qquad [2]$$

in scaled norm $|| x ||_{2,\alpha}^2 = \sum_{i \in \{1,...,nj\}} (\alpha_i x_i)^2$. It is not possible to simply modify the PCA algorithm for minimization of any maximum norm, so this is the best possible approximation we have thought.

We have analysed the matrices M_{22} and M_6 using the modified algorithm. Tables 3 and 4 include data comparable with that in Tables 1 and 2. For purpose of comparability the presented errors are not $\tilde{\tilde{E}}_n$ but

$$\tilde{E}_n^2 = || X - Y ||^2 \qquad [3]$$

where X is the data matrix (M_{22} or M_6) and Y is its approximation obtained using PCA with the last modification.

The errors measured this way have evidently risen. This should be so as the first modification of PCA lead always to optimal subspace in non-scaled norm l^2.

Table 4. Table of projection errors $\tilde{E}_n$ and $p_{\tilde{E}n}$ for set of analyses M_6.

n	5	4	3	2	1	0
$\tilde{E}_n$	$4.7{\cdot}10^4$	$5.8{\cdot}10^4$	$8.3{\cdot}10^4$	$8.5{\cdot}10^4$	$9{\cdot}10^4$	$2{\cdot}10^6$
$p_{\tilde{E}n}$ (%)	2.87	3.51	5.07	5.16	5.58	100

We have also mutually compared the results of both modified algorithms considering both parameters of accessibility, i.e. non-negativity of all components of reconstructed data and short distance between original and reconstructed data in scaled maximum norm $|| \cdot ||_\alpha$. The first parameter can be well evaluated from Table 5 where numbers of original data in set M_{22} or M_6 whose projections have at least one negative component.

Table 5. Number of elements of M_6 and M_{22} whose projections to the optimal subspace obtained using non-scaled or scaled PCA contain at least one negative component.

n	1	2	3	4	5
M_6 *non-scaled*	0	0	0	3	0
M_6 *scaled*	0	2	7	0	1

n	1	2	3	4	5	6	7	8	9	10	11	12 to 21
M_{22} *ns.*	0	6	2	7	6	7	3	7	10	5	5	0
M_{22} *sc.*	0	6	7	4	5	7	6	1	0	0	0	0

Reduction to dimension 1 cannot show any projection with negative components as positive data are projected to positive first principal component. If we want to project complete M_{22} to dimension higher than 1 we should select dimension higher than 11 or scale the data and use dimension higher than 9. Generally scaling does not make the first parameter of accessibility better as can be seen from the first part of Table 5 where numbers for M_6 are collected.

The second accessibility parameter was investigated by statistical analysis of maximal norm $|| \cdot ||_\alpha$ of difference between original data from M_{22} or M_6 and their projections. We illustrate the results in Table 6 including selected statistical quantities for reduction of M_6 to dimensions 3 and 4.

Table 6. Maximum, average, and median of maximum norm $|| x-\Pi_{Vn} x ||_\alpha$ of all elements of M_6 for n=3 and 4 using both modifications of PCA.

	dimension 3		dimension 4	
	non-scaled	scaled	non-scaled	scaled
maximum	132	149	158	47
average	4.33	1.84	3.91	1.49
median	0.18	0.08	0.17	0.04

Table 6 shows that scaling does not reduce the maximal error (see dimension 3). But typically can be observed decrease of average and median after scaling. Much higher average than median indicates presence of several extremely great errors. Supposed property of scaling is equalizing the importance of individual components of data vectors. This fact can be documented by Tables 7 and 8 including incidence of maximal error in individual vector components of vectors from M_6 for all reduction dimensions. Occurrence of maximal error in i^{th} component of $x \in M_6$ here means that $|| x-\Pi_{Vn} x ||_\alpha = \alpha_i |x_i-(\Pi_{Vn} x)_i|$.

Table 7. Incidence of maximal error in individual components for M_6 non-scaled.

	component					
n	1	2	3	4	5	6
1	0	3	165	39	347	84
2	0	0	168	40	344	86
3	0	0	267	22	141	208
4	0	0	357	2	0	279
5	0	0	0	0	0	638

Table 7 shows that non-scaled algorithm optimizes the subspace so that projection generates higher proportional differences in components with lower average x_i^{avg} (here they are i=6, 3, 5, and 4). Table 8 shows much more balanced error distribution due to scaling.

The realized analyses do not answer generally the question which of the modified PCA algorithms is better for application. The presented analysis should be done for every specific problem again. Suitability of specific selected procedure then depends on specific definition of acceptability and other priorities given by specific problem. For our application, one parameter of acceptability was non-negativity of all components of projections of all data. It needs not to be so strict requirement in case of another application if a part of the data would be omitted or changed. An interesting direction of research can be construction and testing of different scaling vectors α for getting another distribution of importance of individual components considering e.g. various accuracy of determination of individual component concentrations or various importance in the model (in sense of classification to primary and marginal solutes).

Table 8. Incidence of maximal error in individual components for M_6 scaled.

	component					
n	1	2	3	4	5	6
1	0	14	142	59	315	108
2	23	73	336	109	25	72
3	76	171	2	243	17	129
4	163	235	1	239	0	0
5	592	0	1	45	0	0

For presentation of further analysis of our data we choose 3D reduction of the set M_6 obtained by PCA without scaling.

4 LOOKING FOR SUITABLE BASIS VECTORS

Finding of the suitable reduction of original space described in the last section is an important but not last step before simulation of solute transport and chemical reactions. Besides identification of the suitable subspace it is necessary to choose its suitable basis which will be used for decomposition of all vectors of initial conditions and then all vectors used in the model will be reconstructed using this base for interpretation of results.

The selection of the basis is not important from the point of view of mathematics but it is crucial from the point of view of chemical interpretation – some properties of the basis can significantly help a good interpretation and result presentation. Also the basis vectors can be interpreted like all other vectors as chemical analyses of some real solutions. We will call them basis solutions. Considering it we naturally require that the basis vectors would have only non-negative components.

The coordinates of an individual solution in the basis can be interpreted as mixing ratios of basis solutions. Then we should require that the coordinates of all analysed and all computed solutions in the chosen basis would be non-negative (and additionally their sum would not be higher than one). The second condition is written in brackets because that if we would have the basis with non-negative components so that all measurements had non-negative coordinates in it, fulfilling of that condition in brackets could be arranged by simple multiplication of all basis vectors by suitable positive numbers.

The above mentioned two requirements are not necessary for the model itself but they are important for the interpretation of the modelling results. It is useful to look for such a basis and use it if it exists. Obviously the existence of such a basis is not guaranteed and if it exists, it is not unique. We did not try to make the definition of suitable basis unique as we do not any practical reason for it. We tried to find a method for verification of existence of at least one such a basis. The results of partial automation of process of looking for a suitable basis can be found in Zedek (2008). Here we present the idea of geometrical interpretation of the process of looking for a suitable basis.

An important fact is that the basis obtained by Principal Component Analysis cannot fulfill the first condition (non-negativity of all components of all basis vectors) for dimensions higher than one. The first basis vector has rather always all positive components (it is oriented in some sense to the centre of the set of measured data which is positive) but all other basis vectors are always normal to it, thus they will have at least one negative component.

To be a basis of the optimal subspace V_n and fulfill above mentioned two parameters of suitability, the set of vectors should in geometrical terms fulfill the following conditions:

- They should lie in the subspace V_n and number of them should be n (the dimension of V_n),
- they should be oriented to the positive orthant of space V (their components should be non-negative), and
- the convex cone confined by the vectors must include all projections of measured data to the subspace V_n (the coordinates of all projections of data in that basis should be non-negative).

Here the term "convex cone" means an infinite pyramid of dimension n with its top in the origin of coordinate system and edges parallel with individual basis vectors. For n=2 it is a triangular sector in a plane defined by two basis vectors, for n=3 it is a trihedral pyramid defined by three basis vectors etc.

The first two conditions can be fulfilled simply and a problem defined by them has certainly infinitely many solutions because the subspace V_n always includes at least one positive vector. The first vector of the new basis can be chosen as equal to the first vector obtained by PCA and each other i^{th} vector can be given as a linear combination of the first and i^{th} vectors from PCA containing non-zero multiple of the i^{th} vector and having non-negative components. The last condition technically complicates the problem because the first vector from PCA is oriented into the centre of measured data and so it cannot be a member of any basis fulfilling the third condition.

The meaning of term "suitable basis" should be precisely defined separately for each problem and our definition is not general. That is why the algorithmization of construction of a suitable basis is in general problem with many parameters. Some algorithms for special purposes were proposed in Zedek (2008). Here we only present a possibility of geometrical analysis of the problem for help the expert precisely define his basis requirements or straight propose the basis for his problem. For making an idea about solubility of the problem, make the following steps:

- Apply PCA to your set of measurements M in your space V – choose your optimal subspace V_n and get the set of projections of elements of M to V_n (denote it □).
- Find the set of projections of standard basis vectors of V to V_n (denote it $\acute{E}$).
- Project the sets □ and $\acute{E}$ in V_n to the unit ball with its centre in the coordinate origin (you get sets $\mathbb{M}$ and $\mathbb{E}$ on spherical manifold $\mathbb{V}_n$ of dimension n-1).
- Find the convex hull of $\mathbb{M}$ in $\mathbb{V}_n$ (the set of its vortices denote $\mathbb{M}^*$) and convex hull of $\mathbb{E}$ in $\mathbb{V}_n$ (the set of its vortices denote $\mathbb{E}^*$).

The last operation (looking for the convex hull) can be realised using e.g. Incremental Method described and demonstrated on the web page Lambert (1999).

The choice of a suitable basis then can be realized on n-1-dimensional manifold $\mathbb{V}_n$ so that we choose a set $\mathbb{B}$ including n convexly independent vectors from $\mathbb{V}_n$ so that the whole convex hull of $\mathbb{B}$ in $\mathbb{V}_n$ lies in the convex hull of $\mathbb{E}^*$ and it includes the whole convex hull of $\mathbb{M}^*$. Then we propose the new basis as a set of positive multiples of vectors of $\mathbb{B}$ in V_n so that the sum of all coordinates of each member of □ in the new basis is lower than one.

The set $\mathbb{B}$ needs not exist for any set M and any subspace V_n. If it exists, it needs not to be unique.

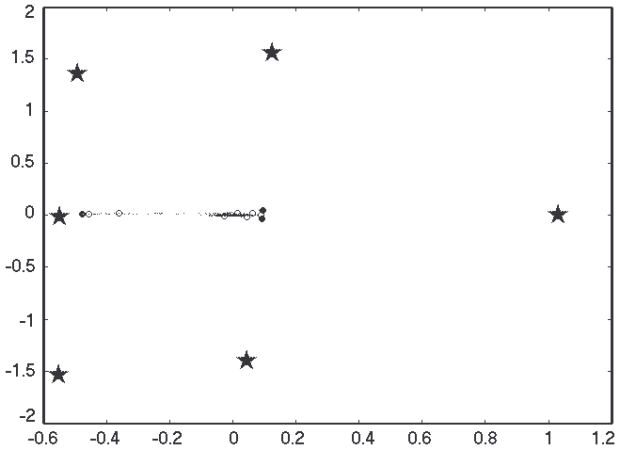

Figure 1. Projection to the unit ball around coordinate origin in the space V_3 obtained by PCA without scaling. Small points: the set $\mathbb{M}_6$, big stars: the set $\mathbb{E}=\mathbb{E}^*$, empty circles: vertices of the convex hull of the set $\mathbb{M}_6$, filled circles: chosen basis of optimal 3D subspace $\mathbb{B}$.

Figure 1 shows the result of proposed procedure applied to the set M_6 reduced to the optimal 3-dimensional subspace obtained by PCA without scaling. The picture shows that in this case we are very free choosing the basis. The realised choice is made so that one of the basis vectors corresponds to one projection of measured data (the member of □$_6$ corresponding to solution with the lowest content of solutes) and the other two basis vectors are chosen as close to other measured data as possible.

So we have done an efficient decomposition of simulated solutions to a three-member basis. Each solution in analysed set can be interpreted as a mixture of the three basis solutions (mixing ratio of each of them is given by the corresponding coordinate in the new basis) and distilled water (its chemical analysis corresponds to zero vector and its mixing ratio is given as one minus sum of all coordinates in the new basis).

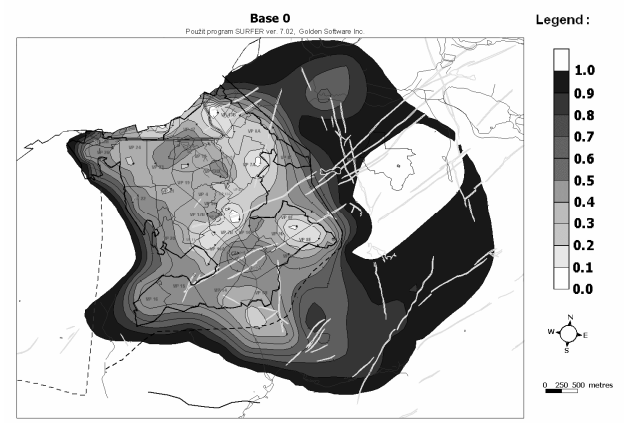

Figure 2. First coordinate of chemically analysed solutions in 4-componental basis pictured in a map. The first basis vector corresponds to composition of original Cenoman water.

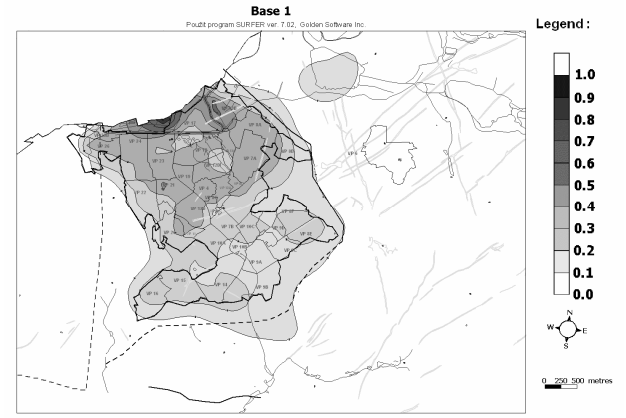

Figure 3. Second coordinate of chemically analysed solutions in 4-componental basis pictured in a map. The second basis vector corresponds to composition of technology solution formerly injected to north-western part of the site.

5 MOTIVATION EXAMPLE FOR APPLICATION OF THE METHODOLOGY

In this paragraph we interpret possible value of proposed methodology on the result of Ing. Vladimír Wasserbauer, CSc. from DIAMO, s. e. He decomposed data from chemical analyses of solutions taken from various parts of Stráž pod Ralskem site during one year to 4 basis solutions. They were not selected using the proposed methodology and it is not reproducible. The results obtained by our methodology are not presented here because we have not enough information about location of individual measurements and so cannot draw maps similar to the presented ones. Here presented maps were drawn in s. e. DIAMO by Ing. Jiří Šrámek using software SURFER and they were provided to authors for purpose of such illustration of possible results of presented method.

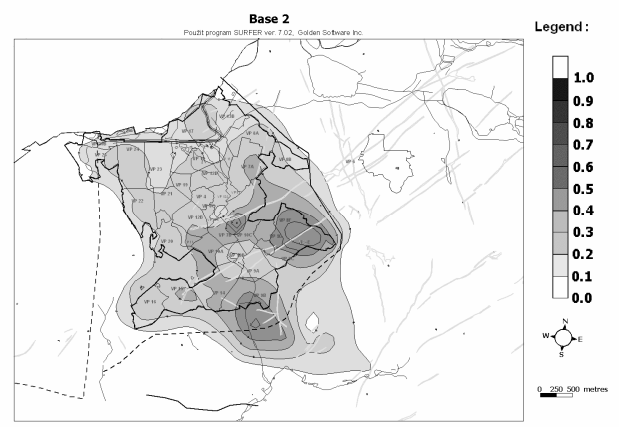

Figure 4. Third coordinate of chemically analysed solutions in 4-componental basis pictured in a map. The third basis vector corresponds to composition of technology solution formerly injected to south-eastern part of the site.

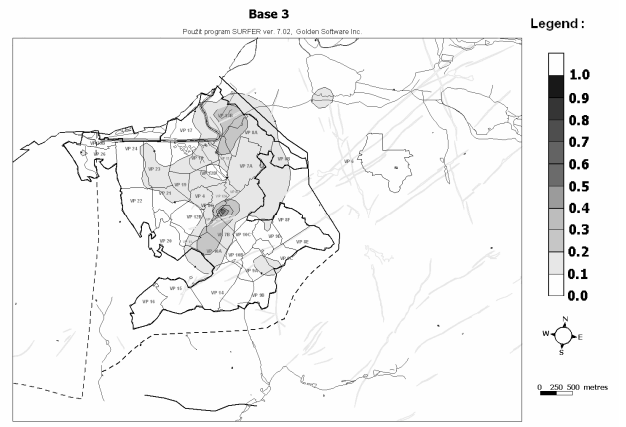

Figure 5. Fourth coordinate of chemically analysed solutions in 4-componental basis pictured in a map. The fourth basis vector probably corresponds to composition of products of reaction between the formerly injected technology solutions.

The chemical analyses were projected to a subspace defined by four basic vectors corresponding to composition of three real solutions present at the site and one more solution obtained by iterative optimization. The first basis vector corresponds to composition of original Cenoman water which was present in the underground before leaching started. Other two basic vectors correspond to composition of two different technology waters formerly injected into the underground for purpose of leaching uranium. Looking for the fourth basis vector was done so that differences of individual measurements from their orthogonal projection to the subspace defined by the four basic vectors were minimized. It does not correspond to any real solution.

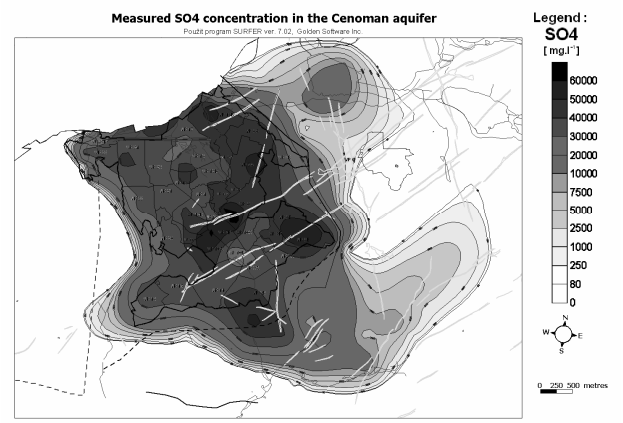

Figure 6. Measured concentrations of SO_4 pictured in a map

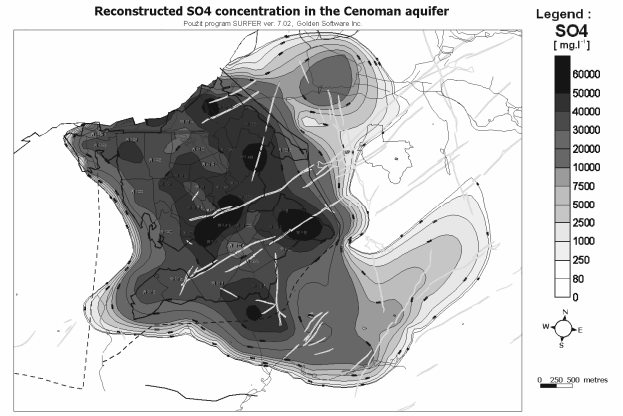

Figure 7. Concentrations of SO_4 reconstructed from data reduced to 4D subspace pictured in a map.

Figures 2 to 5 show decomposition of chemical analyses into above described basis. Each analysed solution was taken from a particular place at the site and corresponds to its respective point in the map. The values were expanded to the plane using software SURFER. The values pictured in Figures 2 to 5 correspond to coordinates of projection of analysed solutions to the four-dimensional subspace in above mentioned basis. Figures 2 to 4 show dominant representation of original Cenoman water around the leaching fields and domination of the two technology waters in the parts of the site where they were injected. Figure 5 shows significant representation of the fourth basis solution at the interface between the regions dominated by technology waters and it can be interpreted so that the fourth basis solution corresponds to products of reactions accompanying mixing of the two technology waters.

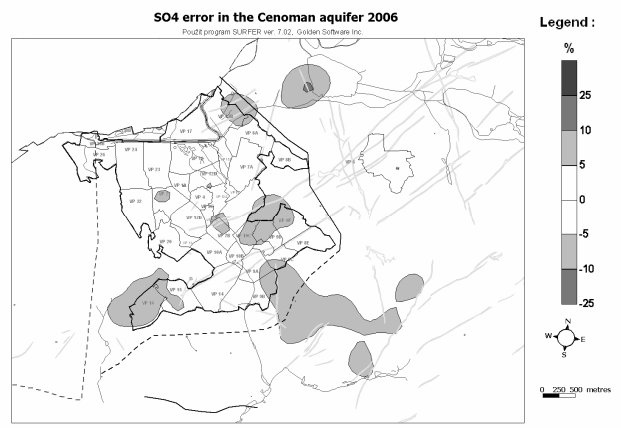

Figure 8. Proportional differences between measured concentration of SO_4 and concentration of SO_4 reconstructed from data reduced to 4D subspace pictured in a map.

Figure 6 pictures the map of measured SO_4 concentrations. Similar map at Figure 7 shows distribution of SO_4 concentration obtained from data projected to four-dimension-subspace. Figure 8 shows the map of differences in concentration of SO_4 between original and projected data. Although the introduced result does not correspond to decomposition to the optimal four-dimension-subspace, the observed differences in SO4 concentration do not exceed 20% and in most samples they are less than 5%.

6 CONCLUSION

We have proposed a new approach to reduction of dimension of a transport-reaction problem for reducing computing time needed for its simulation. The core of the proposed procedure is constituted by application of Principal Component Analysis.

We have not formulated a complete methodology of reduction of dimension of a general problem but we have shown the main ideas applied on one specific problem and formulated the main questions that should be solved specially for each specific problem.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of ing. Vladimír Wasserbauer, CSc. (DIAMO, s. e.) who defined the specific problem, brought lot of ideas and allowed testing on problem specific data, and Ing. Jiří Šrámek (DIAMO, s. e.) who drew the maps presented at Figures 2 to 8.

This result was realized with subvention of the budget of the Czech Republic through the project Advanced Remedial Technologies and Processes No. 1M0554 of the programme Research Centres (PP2-DP01) of the Ministry of Education, Youth, and Sports and with subvention of Grant Agency of the Czech Republic under project No. 102/06/P450.

REFERENCES

Lambert T. 1999. Convex Hull Algorithms, Research project, University of New South Wales, Sydney, Australia. *http://www.cse.unsw.edu.au/~lambert/java/3d/*

Smith, L.I. 2002. A tutorial on Principal Components Analysis. Student Thesis, University of Otago, New Zealand, *http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf*

Zedek, L. 2008. Určení kritérií k optimalizaci báze pro snížení dimenze problému modelování transportu mnohasložkových vodných roztoků [in Czech], Master Thesis, Technical University of Liberec, Czech Republic