CUNOŞTINȚE DIN DATE INDUSTRIALE: APLICAȚIE. PARTEA I - ANALIZA
KNOWLEDGE DISCOVERY IN INDUSTRIAL DATASETS: APPLICATION. PART I - ANALYSIS

ZENO GHIZDĂVEȚ[1*]

[1] *Universitatea POLITEHNICA Bucureşti, Str. G. Polizu nr. 1, sector 1, Bucureşti, România*

O combinație de tehnici de analiză şi de predicție a fost folosită în lucrare pentru a putea extrage cunoştinte dintr-o bază de date industrială, bidimensională. S-a intenționat ca intervenția umană asupra alegerii mărimilor dependente şi independente, precum şi a înregistrărilor care conțin informație relevantă pentru proces să fie minimă. In acest scop, nu a fost încorporată în modele şi experiența prealabilă. Au fost reținute în lucrare doar rezultatele parțiale/finale care pot fi interpretate in mod obiectiv (corelatiile obținute pot fi totuşi verificate pe baze practice). Deoarece articolul are un caracter practic, partea teoretică a fost omisă dar, acolo unde s-a impus, a fost referită. În acelaşi mod s-a procedat şi cu modul de lucru cu mediile software pe care le-am utilizat şi care sunt menționate la sfârşitul articolului. Tehnicile de lucru au constituit componentele unei abordări iterative, bazată pe un mod de lucru euristic. Această parte a lucrării este afectată analizei bazei de date şi selecției pertinenete a parametrilor de proces care vor fi utilizați în cea de-a doua parte a articolului.

Various analysis and prediction techniques were corroborated in the paper to extract knowledge from a bi-dimensional industrial dataset. It was intended to consider the least human intervention possible in all, the selection of the system predictors and dependant variables and of the data records that hold valuable information on the process. In that purpose, no prior experience was used and only results that could not be subjectively interpreted were elected (however, extracted correlations between parameters could be confirmed on an experiential base). As the paper addresses [a large amount of] practical issues, the theoretical part was omitted, though sometimes referred. This is also the case for the way several software packages mentioned at the end of the paper were used on this particular real-life analysis and prediction task. Techniques were manually iterated within a heuristic approach. In this paper we deal with the analysis of the database and the selection of the pertinent process parameters to be used in the second part of the article.

Keywords: *knowledge discovery, multivariate analysis, industrial processes*

Extragerea de cunoştințe din baze de date este o metodologie de analiză şi predicție care este deja folosită în special în domeniul economic şi va continua, probabil, să fie implicată în activitățile de modelare a sistemelor complexe care pot fi caracterizate printr-un comportament neliniar, nestaționar şi afectat de zgomot de fond. Astfel de sisteme pot fi cel mai corect descrise drept conținând atât informații explicite cât şi mai puțin evidente, corespunzătoare interrelațiilor dintre parametrii de proces. Alături de un mare număr de parametri utilizați în monitorizare, descriere şi control, o serie de factori influențează procesul într-un mod care nu este nici stăpânit, nici nu poate fi inclus într-un model matematic determinist (spre exemplu eroarea umană, de aparat sau condițiile meteorologice). Instalația de clincherizare intră în categoria unor astfel de sisteme complexe, fiind un sistem neliniar, caracterizat prin multiple intrari şi ieşiri între care există conexiuni puțin (sau deloc) evidente şi între care există intervale temporale mari – dar şi variabile – la măsurare. Mai multe

Knowledge discovery in databases is not a recent methodology for analysis and prediction, especially in economics, and will subsist as long as the need for modeling nonlinear, noise-affected and non-stationary complex systems will continue. Such systems could be best described by carrying overt and intrinsic information that are best explained by interrelations between process parameters. Along with a large number of parameters assigned to monitor, describe or control the process, several factors are influencing the process behavior in a manner that could not be under our command nor mathematically accounted (such as human and apparatus errors and weather conditions). A clinker plant could enter this category of complex systems, by being a multiple-input/multiple-output (MIMO) system characterized with couplings, nonlinearities and large and variable time delays. More supporting considerations about these systems could be found in [1].

The aim of the paper is to identify causal relationships and patterns in an industrial database

* Autor corespondent/Corresponding author,
Tel.: +40 21 402.39.25 , e-mail: zghizdavet@gmail.com

considerații despre astfel de sisteme pot fi găsite în [1].

Scopul lucrării este de a identifica relații între parametri, şabloane şi alte informații netriviale, necunoscute anterior, într-o bază de date industrială, care pot fi utilizate în scopul predicției unei stări viitoare, fiind disponibile date relevante în acest scop despre funcționarea recentă a instalației de clincherizare. Extragerea de informații din astfel de baze de date industriale se face fără apel la o cunoaştere anterioară despre procesul de clincherizare şi fără a se impune una sau mai multe variabile de proces drept mărimi țintă. Acestea vor fi identificate cu ajutorul unor metode din domeniul *Statisticii Matematice* şi *Rețelelor Neuronale Artificiale* (*ANN*): *Analiza Componentelor Principale (PCA), Analiza Factorilor (FA), Analiza Corelațiilor (CCA), Analiza de Regresie Multiplă (MRA), Analiza Clusterilor (CA), Analiza Pareto (PA), Self Organized Maps (SOM), predicția cu Rețele Neuronale Artificiale (aproximarea de funcții), Arbori de decizie (DT)* şi *extragerea automată de reguli de inferență (RI), predicția în serii de timp cu Rețele Neuronale Artificiale sau modele ARIMA*. *Algoritmii Genetici* vor fi folosiți pentru a creşte performanțele predicției în serii de timp. Deoarece spațiul necesar pentru prezentarea unui pachet minim de rezultate care să susțină scopul şi concluziile lucrării este important, nu au fost furnizate şi informații teoretice referitoare la metodele matematice folosite. Acestea sunt, însă, disponibile într-un volum impresionant în literatura de specialitate; informațiile de literatură considerate a fi utile din punct de vedere practic, accesibile autorului, sunt însă referite în articol.

Tehnicile menționate au fost utilizate în şapte studii de caz (denumite *Caz 0, 1, 2A, 2B, 2C, 2D, 3*) în cadrul unei abordări euristice, cu scopul de a identifica relațiile existente între parametrii de proces şi, astfel, de a obține capacitatea de predicție optimă. Baza de date folosită provine de la o fabrică de ciment funcționând pe procedeul uscat; selecția acestui tip de proces s-a făcut mai ales datorită complexității sale, dată de un număr mare de parametri precum şi de interconexiunile dintre parametri, dificil de estimat – chiar imposibil în acest moment – pe calea tradițională a modelării matematice deterministe [2]. Informații despre unele dintre tehnicile folosite în lucrare precum şi o serie de alte aplicații ale acestora pot fi găsite în [2].

În lucrare au fost preferate valorile orare ale parametrilor de proces în locul celor instantanee. Este cunoscut faptul că valorile instantanee prezintă decalaje temporale corespunzătoare locației din instalație de unde sunt înregistrate (unele decalaje sunt de ordinul secundelor în timp ce pentru altele se ating zeci de minute); aceste decalaje depind – cel puțin – de fluctuațiile uzual întâlnite în activitatea acestor instalații.

Cu toate acestea, în prezent rulează cu succes aplicații de control de proces care permit

that could be used to predict a future state by knowing the recent, relevant operating history of a clinker plant. The purpose is to extract sound information from an industrial database, regardless of any prior knowledge of how clinkering process works and what its priorities are. The target function(s) will not be stated but found by means of a broad range of techniques coming from Statistics and Artificial Neural Networks (ANN) domains. Principal Components Analysis (PCA), Factor Analysis (FA), Canonical Correlations Analysis (CCA), Multiple Regression Analysis (MRA), Cluster Analysis (CA), Pareto Analysis (PA), Self Organized Maps (SOM), ANN prediction (function finding/approximation), Decision Trees (DT) and Rules Induction (RI) techniques, Time Series Prediction by ANN and ARIMA models will provide the tools to discriminate what constitutes relevant information from raw data. Genetic Algorithms will be used to improve the performance of the ANN time series prediction. Due to the amount of space required for displaying a minimum of results that supports the aim of the paper, theoretical information about techniques used here were not given. However, they can be easily found, as literature is already flooded with; as a guide, we provided some useful references found to be mostly related to the subject of that work and to contain a useful, practical meaning.

All these techniques were used within seven case studies (denoted *Case 0, 1, 2A, 2B, 2C, 2D, 3)* within a heuristic approach to obtain the best relationships among process parameters and, thus, the most accurate prediction. The considered database was selected to come from a dry process clinker plant, mainly due to its inherent complexity, showing many process parameters but not less important mutual influences between them, difficult or even impossible at that stage to be modeled in a deterministic way [2]. In addition, information about some of the techniques used in the paper and several other case studies on industrial plants could be found in [2].

Hourly averaged values were used instead of instant values. The reason is, instant values recorded over the entire plant are time delayed (for some of them it is a matter of seconds while for others is of tens of minutes, depending on the location of the sensors within the plant and on the plant's design), and delays are dependant on – at least – the normal fluctuations each plant encounters in its operation. That makes the time delays fluctuant as well, thus increasing the complexity dimension. However, to our best knowledge, process control applications that correlate, for example, flame temperatures with free CaO in clinker or NO_x levels in exhaust fumes are running for a number of years with good results [3, 4]. It has to be stated that these ones are considering subsystems of the clinkering plant (such as cyclones – calciner – rotary kiln or rotary

corelarea – spre exemplu – a temperaturilor înregistrate în zona de ardere cu CaO liber în clincher sau cu concentrația de NO_x în gazele de ardere evacuate din schimbatorul de căldură cu cicloane [3, 4]. Trebuie însă menționat că aplicațiile controlează subsisteme ale instalației de clincherizare (cum ar fi schimbătorul de căldură – calcinator – cuptor rotativ sau cuptor rotativ – răcitor grătar) şi nu întregul sistem. Problema se complică cel puțin datorită oscilațiilor ciclice ale debitelor de material înregistrate în operare [5] dar şi datorită fluctuațiilor normale ale valorilor mărimilor de intrare.

Baza de date inițială a cuprins 18 variabile şi aproximativ 2100 înregistrări. Patru variabile care le dublau pe altele au fost eliminate, însă doar după o analiză a influenței lor în comparație cu cea a mărimilor-pereche. Toate aceste variabile au fost înregistrate pe cele două ramuri ale schimbătorului de căldură. După o primă analiză vizuală, alte trei variabile care aveau înregistrări lipsă sau prezentau erori (mult depărtate de restul datelor sau chiar cu semn opus) au fost de asemenea înlăturate. Înregistrările corespunzătoare perioadei finale, caracterizată prin întreruperi frecvente în funcționare, au fost eliminate, rezultând configuratia denumită *Caz 0* (tabelul 1) considerată a fi de referință. Evoluția variabilelor selectate în acest caz este prezentată în figura 1:

CM – cuplul motor, %
COB – concentrația CO în gazele de ardere, ramura B, %
CON – consumul de curent electric al concasorului răcitorului grătar, A
DCX – debitul de combustibil intrat în cuptorul rotativ, t/h
DIRM – depresiunea exhaustorului, ramura B, mbar
NO_xA – concentrația NO_x în gazele de ardere, ramura A, ppm
O_2B – concentrația O_2 în gazele de ardere, ramura B, %
PE – consumul de curent electric al elevatorului cu cupe de alimentare cu făină brută, A
TAE – temperatura aerului în exces, °C
TAS – temperatura aerului secundar, °C
TGVRA – temperatura gazelor evacuate din schimbător, ramura A, °C

O serie de tehnici de analiză şi predicție dintre cele menționate au fost aplicate cazului de referință în scopul identificării variabilelor dependente, fără a se face apel la informații anterioare referitoare la proces, şi al cuantificării – acolo unde a fost posibil – a conexiunilor cu mărimile independente. O procedură tip pas-cu-pas a fost adoptată: cazul curent derivă din precedentul prin extragerea de variabile sau/şi înregistrări, rezultând astfel alte studii de caz. Denumirile acestora, numărul de înregistrări/variabile precum şi mărimile identificate drept dependente se găsesc în tabelul 1.

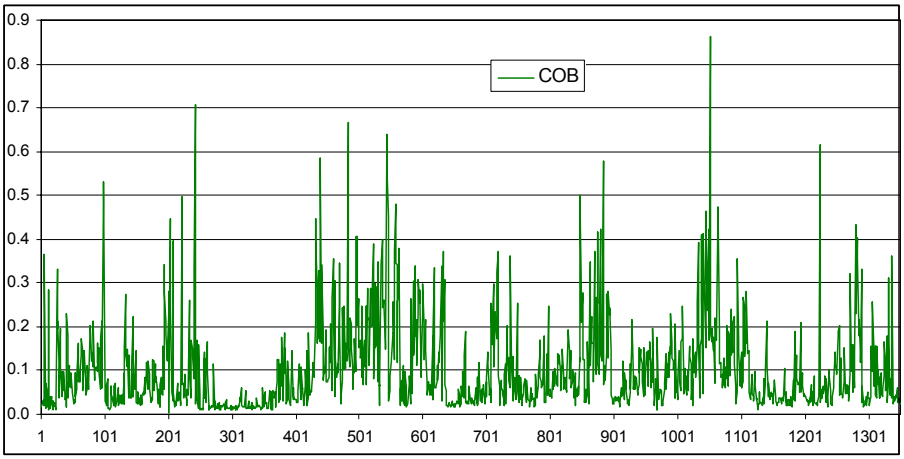
kiln – grate cooler) and not the entire system. Furthermore, the task is more difficult to be accomplished, at least due to the cyclic oscillations recorded in their operation, as considering the mass flow [5] but also due to ordinary non-uniformities in plant's inputs value.

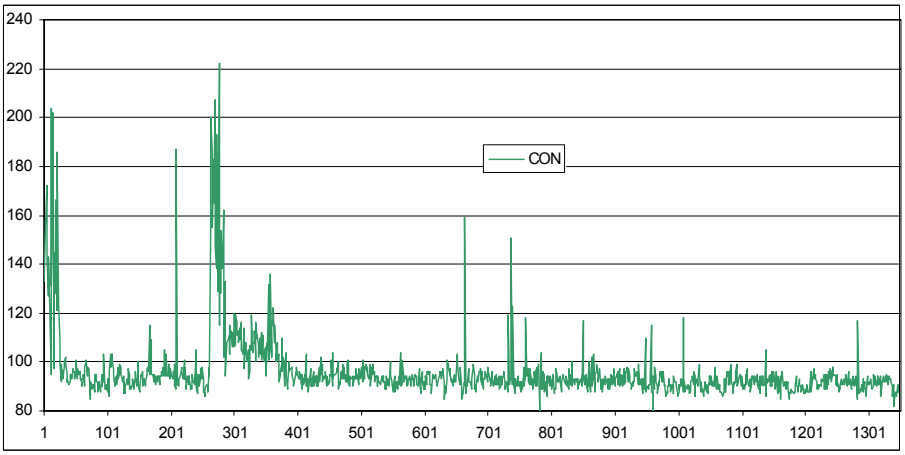
Initially, a database consisting of 18 variables and about 2100 data records was available. Four variables that were doubling other ones were also removed from the dataset but only after completing an analysis of their influence comparing with their twins; the fittest survived. All these four variables were recorded in the two strings of the cyclones tower. After a first visual inspection, other three variables that showed missing values or errors (well outranged or showing an opposite sign on some values than the rest) were removed as well. In addition, records corresponding to a final period characterized by corrupted data due to frequent plant upsets were detached, also. Consequently, it resulted the configuration designated *Case 0* (table 1) and considered here to be the reference. The evolution of these variables over time is given in Figure 1. The selected variables are:

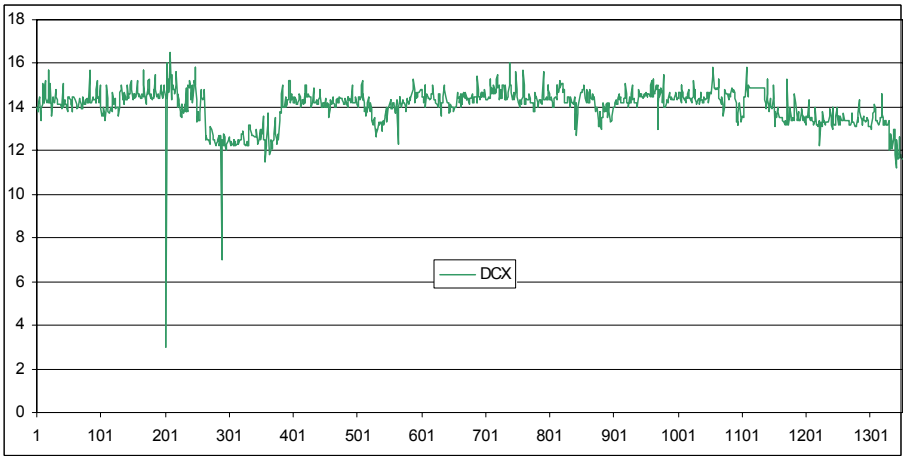
CM – kiln torque, %
COB – CO concentration in fumes, B string, %
CON – grate cooler crusher electric consumption, A
DCX – rotary kiln fuel intake, to/h
DIRM – exhaust fan pressure drop, mbar
NO_xA – NO_x content in exhaust fumes, A string, ppm
O_2B – O_2 content in exhaust fumes, B string, %
PE – raw meal bucket elevator electric consumption, A
TAE – grate cooler excess air temperature, °C
TAS – secondary air temperature, °C
TGVRA – exhaust fumes temperature, A string, °C

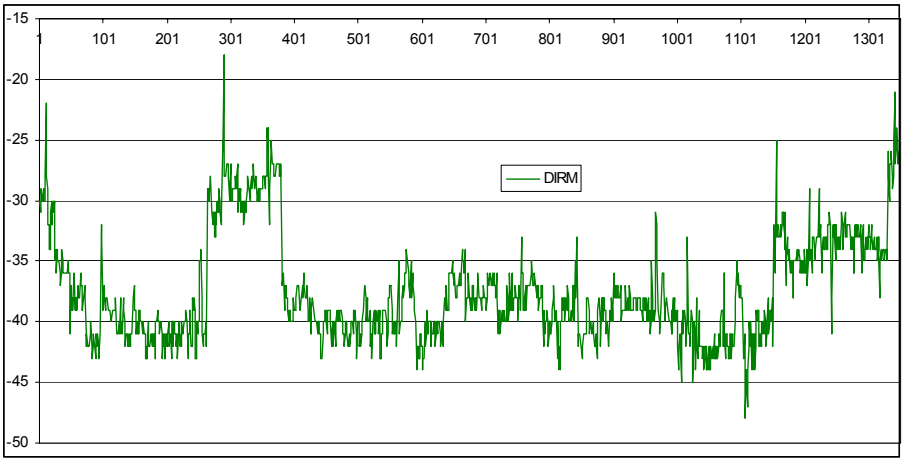
Several analysis and prediction procedures were applied to the reference case with the aim of identifying the dependant variable(s), by excluding a priori knowledge of the process, and to quantify its/their interrelations with the predictors. A stepwise procedure was followed, meaning: by extracting records and/or variables from the reference case, according to the results, we had derived other six cases studies. Their denomination, number of records/variables and identified outputs (dependant variables, target functions) could be seen in table 1.

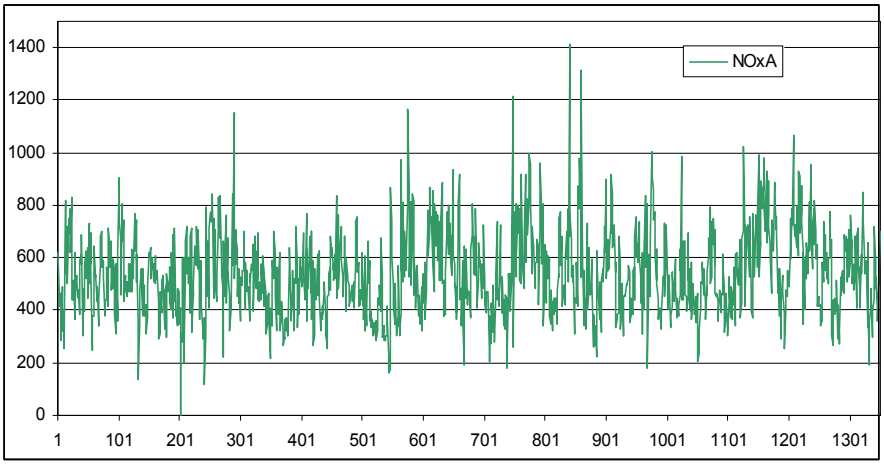
Remark: it should be stated that the initial paper outline of splitting the research efforts within two distinct categories, i.e. analyzing (screening the data through a set of analysis methods and extracting useful information) and than prediction (using the information in order to get the knowledge out of an unknown dataset) proved to be ineffective, as some methods falling into these two categories had to be iterated for almost each of the case studies considered, with no respect for

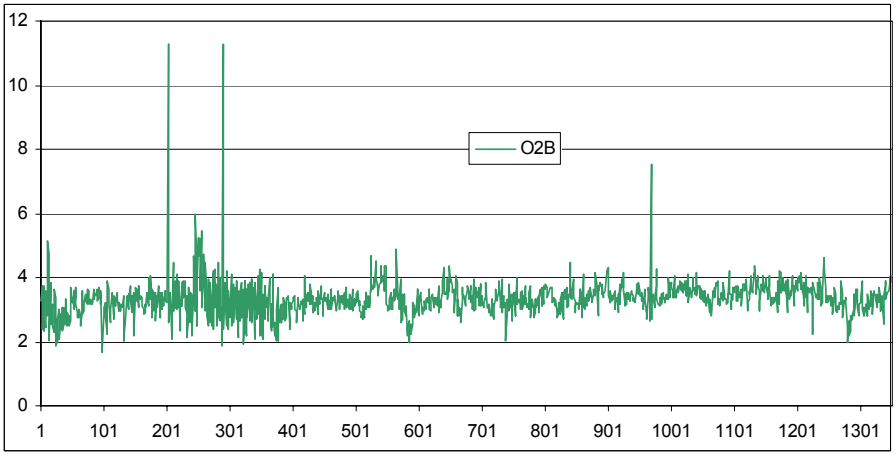

COB
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301

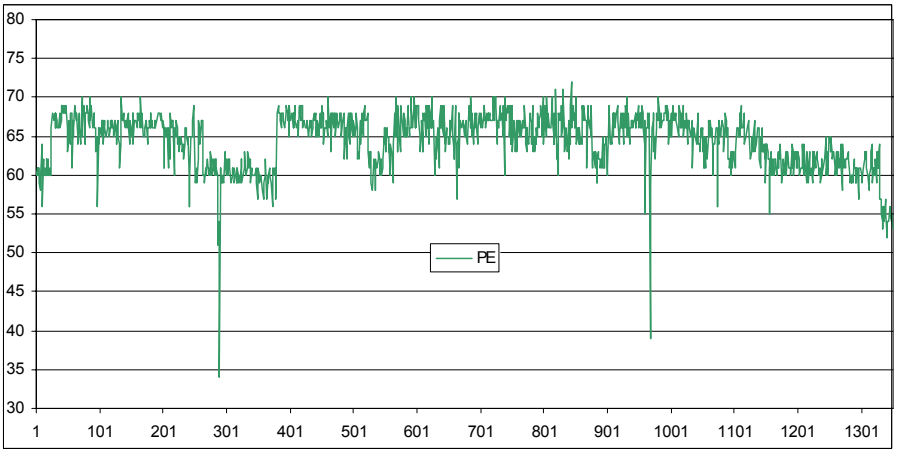

CON
240
220
200
180
160
140
120
100
80
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301

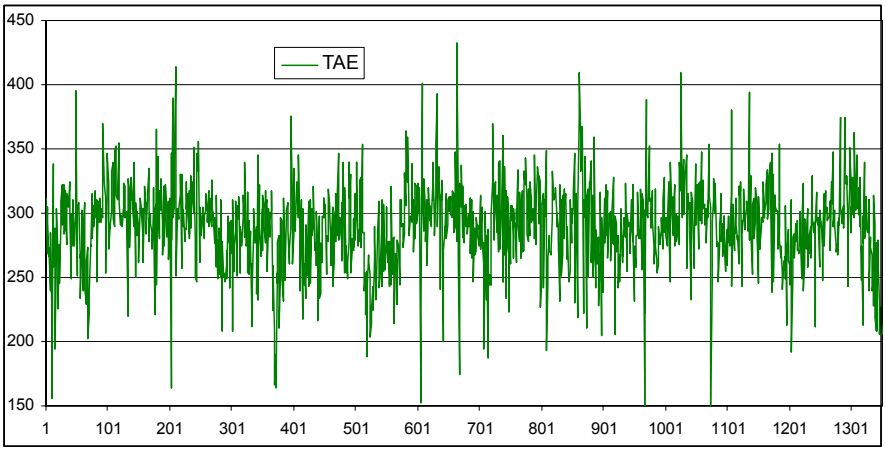

DCX
18
16
14
12
10
8
6
4
2
0
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301


DIRM
-15
-20
-25
-30
-35
-40
-45
-50
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301


NOxA
1400
1200
1000
800
600
400
200
0
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301


O2B
12
10
8
6
4
2
0
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301


PE
80
75
70
65
60
55
50
45
40
35
30
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301


TAE
450
400
350
300
250
200
150
1
101
201
301
401
501
601
701
801
901
1001
1101
1201
1301

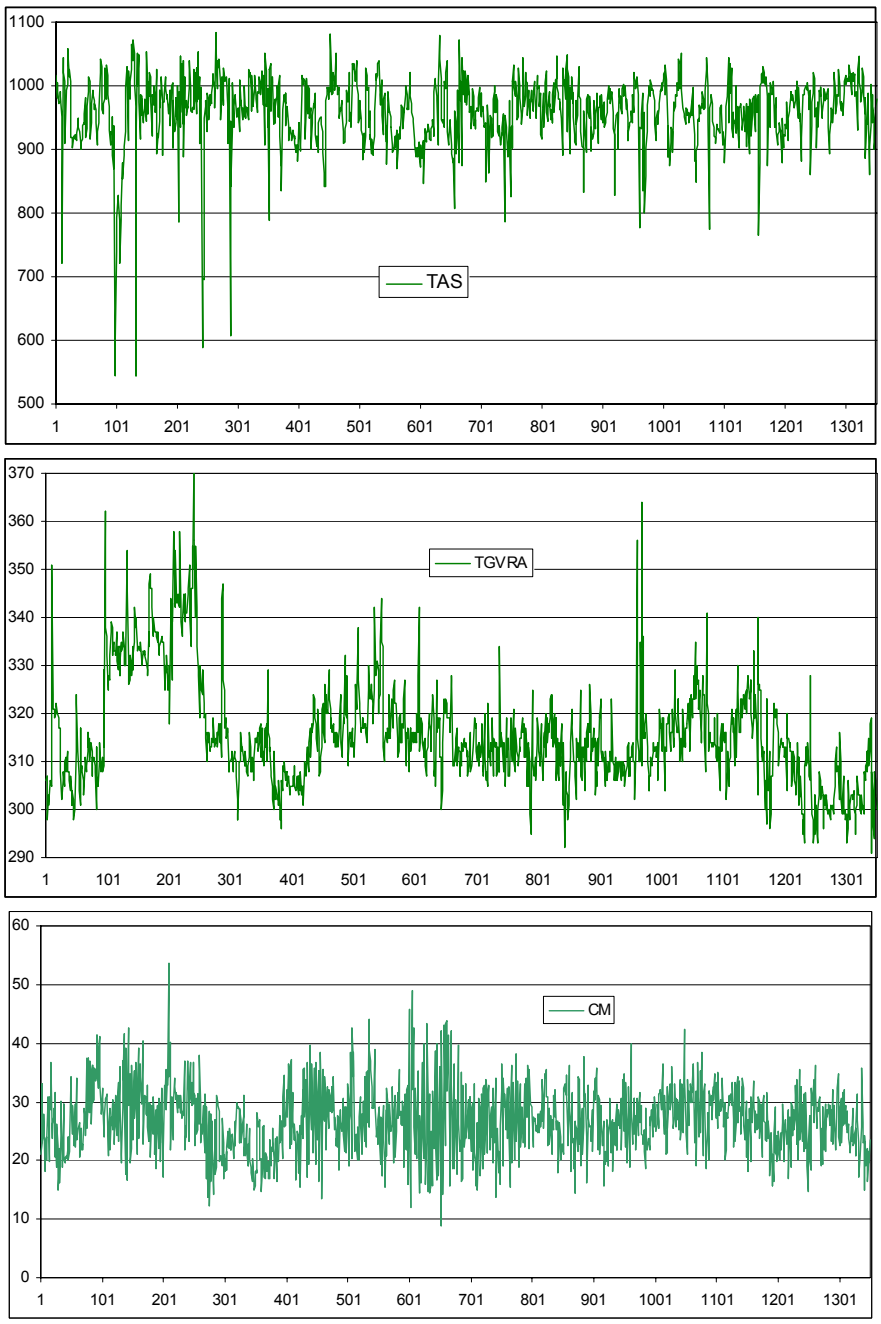


Fig 1- Evoluția a 11 parametri de proces pentru cele 1350 înregistrări (*Caz 0*) / *11 process parameters plots over 1350 data records (Case 0).*

Observație: inițial s-a intenționat împărțirea articolului în două părți: ***Analiza*** (căutarea de informații utile în datele disponibile) şi ***Predicția*** (utilizarea acestor informații pentru predicție pe un set de date necunoscut). Această abordare s-a dovedit rapid ineficientă deoarece metode aparținând celor două categorii au intrat împreună într-o procedură iterativă de lucru – fără a putea face o distincție de acest tip – pentru aproape orice caz considerat. Totuşi, rezultatele predicției vor fi prezentate în a doua parte a lucrării pentru a sublinia nevoia unei căi logice în extragerea de cunoaştere. Un exemplu este cel al identificării şi eliminării excepțiilor - valori care se plasează cu a didactic separation. [However, prediction results will be presented at the second part of the paper to emphasize the need for a logical, evolving pattern in generating knowledge]. This was the case of identifying and extracting outliers (and also outlying variables) – which was the most intricate part – where analysis methods had to be validated by predictions, as some of the data were showing influencing patterns, yet they were considered outliers and inversely. It must be emphasized that, for these particular case studies, automatic outliers/outlying variables removal could not be proved possible with the available software packages. When done, prediction patterns were

mult în afara norului statistic (dar şi al variabilelor care nu influențează mărimile de ieşire), în care rezultatele analizei necesitau validare prin predicție, deoarece în analiză o parte din date arătau o influență puternică deşi erau considerate a fi neimportante în predicție şi invers. Este util de menționat că, pentru aceste studii de caz, eliminarea automată a înregistrărilor/variabilelor care nu influențează în mod semnificativ mărimile de ieşire nu a fost posibilă cu ajutorul aplicațiilor software disponibile. Când s-a încercat, evoluția valorilor de predicție tindea către cea tipică mersului la întâmplare, fără a se putea identifica pentru mărimile-țintă corelații acceptabile cu celelalte variabile.

approaching random walk, showing no significant correlations with the other variables.

Starting from *Case 0*, a visual exploration over all data plots was carried in order to identify regions/values that have similar (or opposite) behaviors and to spot isolated values that could be or not correlated with others.

Regions of obvious similar/opposite trends (A to D) could be identified as follows:

Region A – area containing samples 1 to about 30 in DIRM, CON, PE plots;

Region B – area containing samples 250 to about 370 in DIRM, CON, DCX, PE plots;

Region C – area containing samples 505 to about 560 in O_2B, DCX, PE plots;

Tabelul 1

Cazuri considerate în lucrare / *Cases considered in the paper*

Caz *Case*	Înregistrări *No. of Records*	Variabile *No. of Variables*	Mărimi de ieşire *Output(s)*	R^2	Observații *Observations*
0	1350	11	DIRM	0.5139	-
1	1283	11	DIRM	0.7832	-
2A	1263	11	DIRM	0.8121	-
2B	1263	8	DIRM	0.8078	-
2C	1263	11	DIRM şi/sau DCX *DIRM and/or DCX*	0.7374 sau 0.4819 *0.7374 or 0.4819*	1 mărime țintă *1 target function*
				0.7520 şi 0.5352 *0.7520 and 0.5352*	2 mărimi țintă *2 target functions*
2D	1263	11	DCX	0.6477	-
3	1204	6	DIRM	0.7218	-

Începând cu ***Cazul 0***, a fost făcută o analiză vizuală a tuturor evoluțiilor mărimilor cu scopul de a identifica regiuni/valori care prezintă comportamente similare (opuse) şi de a identifica şi extrage valori izolate care pot fi – sau nu – corelate cu altele existente în evoluțiile altor mărimi.

Au fost identificate patru regiuni având evoluții similare/opuse (A - D):

A – înregistrările 1 până la 30 în seriile DIRM, CON, PE;

B – înregistrările 250 până la aproximativ 370 în seriile DIRM, CON, DCX, PE;

C – înregistrările 505 până la aproximativ 560 în seriile O_2B, DCX, PE;

D – înregistrările 1140 până la 1350 în seriile DIRM, DCX, PE.

Picurile observate la aceleaşi înregistrări: 206 în seriile CON, CM, O_2B, DCX; b - 296 în seriile DIRM, CON, O_2B, DCX, PE precum şi altele care au demonstrat o mai mică influență.

Toate aceste regiuni au fost păstrate în modelele obținute deoarece erau în mod evident corelate, pozitiv sau negativ. Unele dintre înregistrări au fost eliminate dacă testele au dovedit influențe slabe sau negative.

O serie de reguli de inferență au fost deduse cu uşurință din investigația vizuală a evoluțiilor surprinse în aceste regiuni, reguli care sunt consistente una în raport cu celelalte:

A – dacă DIRM creşte atunci CON creşte şi PE descreşte;

B – dacă DIRM creşte atunci CON creşte şi DCX descreşte şi PE descreşte;

Region D – area containing samples 1140 to 1350 in DIRM, DCX, PE plots.

Peaks observed in several plots for the same samples: sample 206 in CON, CM, O_2B, DCX; b - sample 296 in DIRM, CON, O_2B, DCX, PE , as well as others of a smaller influence.

All these regions were preserved in the models as they are noticeably correlated. Some of the samples were removed if tests revealed a small or negative influence.

A series of rules were easily inferred from the visual investigation of the trends within these regions, rules that are consistent to each other:

Region A – *if DIRM increases then CON increases and PE decreases;*

Region B – *if DIRM increases then CON increases and DCX decreases and PE decreases;*

Region C – *if DCX decreases and PE decreases then O_2B increases;*

Region D – *if DIRM increases then DCX decreases and PE decreases.*

The opposite also holds true for all rules.

It should be stated that partial knowledge extracted here from the visual analysis could only strengthen or support further results. By their own, these rules would not be considered yet to improve process control (and this is not the aim of the paper), as they only reflect the way the studied process was controlled at that particular time interval, even though some technological aspects of the clinkering process are revealed (the most accurate is the rule extracted from region C). The same statement is underlined by the clear

C – dacă DCX descreşte şi PE descreşte atunci O_2B creşte;

D – dacă DIRM creşte atunci DCX descreşte şi PE descreşte. Opusul este de asemenea valabil pentru toate regulile obținute.

Este bine de menționat că aceste cunoştințe parțiale obținute din analiza vizuală trebuie considerate doar în conjuncție cu altele obținute pe cale analitică sau provenind din experiență. Izolate, aceste reguli nu pot fi considerate în aplicații de control de proces (şi nu este acesta scopul lucrării), deoarece reflectă doar modul în care procesul de obținere a clincherului a fost condus într-o anumită perioadă de timp, deşi sunt surprinse unele dintre aspectele tehnologice ale procesului (un grad mare de acuratețe poate fi atribuit regulii extrase din regiunea C). Prudența în folosirea regulilor deduse este întărită de faptul că mulți alți parametri importanți nu sunt cuprinşi în aceste reguli.

Pe de altă parte, 67 înregistrări care corespund picurilor izolate au fost eliminate din baza de date de referință, rezultând *Cazul 1*. Toate procedurile de analiză şi predicție vor fi aplicate fiecărui caz în parte începând cu acesta; datorită volumului mare de informații extrase, atât utile cât şi doar informative, numai rezultatele relevante vor fi prezentate în lucrare. Prima țintă în această etapă a constituit-o identificarea parametrilor dependenți. O prima indicație este dată de corelațiile între fiecare pereche de variabile (vezi tabelul 2).

observation that many other important, influencing parameters are not involved in these rules.

On the other hand, 67 records corresponding to isolated peaks were detached from the initial dataset, thus resulting *Case 1*. All analysis and prediction procedures will be applied to each case starting from now, yet, due to the huge amount of useful but also unproductive information extracted, only the most relevant ones will be given in the paper. The first aim in that stage was to quantify correlations and to identify the parameters that could be isolated as dependant. A first indication was given by computed correlations for each pair of variables (see table 2). The only parameters that showed *P-values* (the second line of each variable) below 0.05 (which indicates statistically significant non-zero correlations at the 95.0% confidence level) for all variables are DCX and DIRM. It was considered then that these two variables would be adopted as dependant variables while the rest will be predictors. *Multiple Linear Regression Analysis* made on *Case 0* and *Case 1* showed a good improvement of the prediction quality for DIRM (better R^2 value – see table 1 – and also a more compact statistical cloud on the predicted/observed data plot) while for DCX, even if the R^2 value almost doubled, it still remained very low, under 0.4.

The second step was to analyze the datasets in order to give an essential under -

Tabelul 2

Corelații între fiecare pereche de variabile / *Pearson product moment correlations between each pair of variables*

	CM	COB	CON	DCX	DIRM	NO_xA	O_2B	PE	TAE	TAS	TGVRA
CM		0.0630	-0.0764	0.1921	-0.3048	0.1822	0.0712	0.0950	0.2252	0.0546	0.2386
		0.0238	0.0061	0.0000	0.0000	0.0000	0.0106	0.0006	0.0000	0.0505	0.0000
COB	0.0630		-0.1438	0.2407	-0.3356	-0.5407	-0.2470	0.1155	-0.0261	-0.1983	0.0761
	0.0238		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3500	0.0000	0.0063
CON	-0.0764	-0.1438		-0.2196	0.3888	-0.0022	-0.0561	-0.2059	0.0197	0.1566	-0.0370
	0.0061	0.0000		0.0000	0.0000	0.9371	0.0445	0.0000	0.4797	0.0000	0.1849
DCX	0.1921	0.2407	-0.2196		-0.6801	-0.1066	-0.2928	0.6050	0.1701	-0.1211	0.2629
	0.0000	0.0000	0.0000		0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
DIRM	-0.3048	-0.3356	0.3888	-0.6801		0.1222	-0.0675	-0.6691	-0.2052	0.0652	-0.3234
	0.0000	0.0000	0.0000	0.0000		0.0000	0.0154	0.0000	0.0000	0.0193	0.0000
NO_xA	0.1822	-0.5407	-0.0022	-0.1066	0.1222		0.2567	-0.0578	0.1728	0.2871	-0.0788
	0.0000	0.0000	0.9371	0.0001	0.0000		0.0000	0.0382	0.0000	0.0000	0.0047
O_2B	0.0712	-0.2470	-0.0561	-0.2928	-0.0675	0.2567		-0.1868	-0.0738	0.0343	0.0638
	0.0106	0.0000	0.0445	0.0000	0.0154	0.0000		0.0000	0.0081	0.2188	0.0220
PE	0.0950	0.1155	-0.2059	0.6050	-0.6691	-0.0578	-0.1868		0.2264	0.0663	-0.0260
	0.0006	0.0000	0.0000	0.0000	0.0000	0.0382	0.0000		0.0000	0.0174	0.3512
TAE	0.2252	-0.0261	0.0197	0.1701	-0.2052	0.1728	-0.0738	0.2264		0.2666	0.0096
	0.0000	0.3500	0.4797	0.0000	0.0000	0.0000	0.0081	0.0000		0.0000	0.7324
TAS	0.0546	-0.1983	0.1566	-0.1211	0.0652	0.2871	0.0343	0.0663	0.2666		-0.2604
	0.0505	0.0000	0.0000	0.0000	0.0193	0.0000	0.2188	0.0174	0.0000		0.0000
TGVRA	0.2386	0.0761	-0.0370	0.2629	-0.3234	-0.0788	0.0638	-0.0260	0.0096	-0.2604	

Singurii parametri care au avut *toate* valorile coeficientului *P* (a doua linie corespunzătoare fiecărei variabile) sub 0,05 (ceea ce indică obținerea de corelații semnificative pentru un interval de încredere de 95,0%) dintre toate variabilele au fost DCX şi DIRM. S-a considerat că cele două variabile pot fi adoptate drept dependente în timp ce restul sunt variabile

understanding of these results. It showed that even the plots of absolute frequencies for all variables were better in the second case, they were still distorted from the normal distribution (NO_xA in figure 2a has the best distribution over all) up to having a bimodal distribution as it occurs for DIRM, figure 2b. Shape parameters (standardized skewness and standardized kurtosis

independente. *Analiza de Regresie Multiplă* efectuată pe *Cazul 0* şi *Cazul 1* a arătat o creştere substanțială a calității predicției pentru DIRM (o valoare mai bună pentru coeficientul R^2 – vezi tabelul 1 – şi un mai compact nor statistic în graficul în care sunt prezentate valorile măsurate/de predicție) în timp ce pentru DCX, chiar dacă valoarea R^2 aproape s-a dublat a rămas încă foarte mică, sub 0,4.

Al doilea pas a fost de a analiza datele în scopul înțelegerii acestor rezultate. S-a dovedit că, chiar dacă graficele frecvențelor absolute au fost mai bune decât în primul caz, ele prezentau totuşi diferențe față de distribuția normală (NO_xA în figura 2a prezintă cea mai apropiată distribuție de acest deziderat dintre toate variabilele) până la a avea o distribuție bimodală aşa cum este cazul variabilei DIRM, figura 2b. Coeficientul de asimetrie bazat pe momente centrate şi coeficientul de boltire – vezi tabelul 3a), chiar în *Cazul 1* au luat valori în afara intervalului (-2, +2) pentru cei mai mulți dintre parametri, ceea ce indică deviații semnificative de la normalitatea statistică.

Tranziția de la *Cazul 1* la *Cazul 2A* deşi nu a fost complicată, a necesitat pentru a fi realizată perioada cea mai îndelungată de timp, chiar daca au fost eliminate doar 20 de înregistrări.

– see table 3a) even in *Case 1* were outside the range of -2 to +2 for most of the parameters, which indicate departures from statistical normality.

Transition from *Case 1* to *Case 2A* was the most time consuming part of the work, even if only 20 data records were removed in that stage. The decision to remove these sets has been made within a heuristic approach, by permanently improving shape parameters (table 3a), correlations (table 3b, figure 3). MLR and ANN were also used to validate these analysis results. For example, automatic outlier identification in figure 4a isolates many points (data records) that should be removed as having a small influence over the variability of the data – see circled points – that was not validated by the prediction quality. In fact, by preserving data record 1152 within the model, it showed that its residual variance decreased from about 2350 in figure 4a to about 500 in figure 4b while its leverage (its degree of influence over the variability of the data) increased about 3 times. It become manifest that mutual relationships at the level of data records, that are intricate to be described mathematically, are influencing - sometimes indirectly - multivariate analysis results and, as a consequence, prediction results. More about parameters in figures 4a, 4b

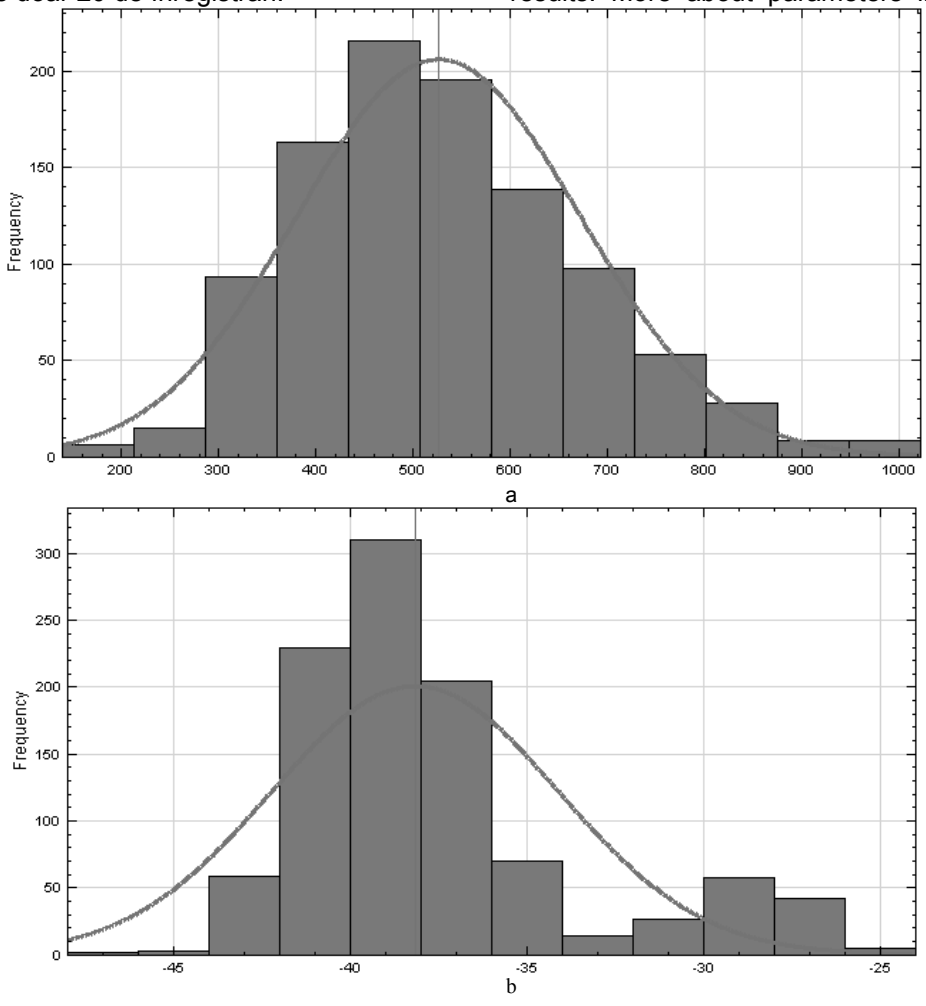


Fig 2 - Frecvențele absolute pentru NO_xA (a) şi DIRM (b) pentru *Cazul 1 I Absolute frequencies for NO_xA (a) and DIRM (b) for Case 1 .*

Eliminarea acestor înregistrări a fost făcută în mod manual, într-un mod euristic, prin îmbunătățirea permanentă a parametrilor de asimetrie şi boltire (tabel 3a), a corelațiilor (tabel 3b, figura 3). *Metoda Regresiei Liniare Multiple* şi *Metoda Rețelelor Neuronale* au fost folosite pentru validarea rezultatelor. Spre exemplu, eliminarea automată a excepțiilor (înregistrărilor neimportante/dăunătoare pentru model) în figura 4a izola un număr mare de puncte (înregistrări) – vezi punctele încercuite – care ar fi trebuit eliminate ca având o mică influență asupra variabilității datelor, fapt care nu a fost validat prin calitatea testelor de predicție. De fapt, prin menținerea înregistrării 1152 în model, se observă în figura 4a o scădere pe ordonată de la o valoare de aproximativ 2350 până la aproximativ 500 în timp ce gradul de influență asupra variabilității datelor a crescut de aproape 3 ori. A devenit evident că relațiile dintre variabile la nivelul fiecarei înregistrări în parte influențează, uneori indirect, rezultatele analizei cu mutiple variabile şi, asadar, şi rezultatele de predicție. Mai multe informații despre parametrii din figurile 4a, 4b, 5a şi 5b pot fi găsite în [6].

O modalitate utilă de apreciere a corelațiilor dintre fiecare pereche de variabile este analiza figurii 3. Corelații liniare, neliniare sau foarte slabe pot fi cu uşurință identificate; de asemenea, puncte (înregistrări) care se găsesc în afara distribuției statistice normale. Prin eliminarea acestora, corelația curentă poate fi îmbunătățită, însă alte corelații pot fi afectate şi nu în mod necesar în mod pozitiv, ceea ce ar putea genera un efect global negativ.

and also 5a and 5b could be found in [6].

A good visual indicator of the way correlations between each pair of variables are made is shown in figure 3. Linear, nonlinear or very little correlations could be easily observed, and, also, points (data records) that are away from the normal distribution. By removing them, the particular correlation we are working on will be improved; however, other correlations will be affected and not always in a positive way, with a possible adverse effect on the overall result.

Based on correlation results (table 3b, figure 3), partial correlations results (not shown here as many other results because the lack of space imposed the presentation of the most important assets) and PCA tests shown in figures 5 a and b, three variables – NO_xA, TAS and TAE - were removed, thus obtaining *Case 2B*, which has the best R^2 value on DIRM and gives the best multivariate prediction results. A simple analysis on table 3b shows that these three variables have small correlations with the two dependant variables or even their values show no growing trend (in fact some of them are decreasing). Moreover, *Principal Components Analysis* results in figures 5a and b places them as being “different” from the others. In fact, PCA results are difficult to be accurately interpreted in other way that in the very absence of any prior knowledge about the modeled process, otherwise the interpretation will be affected by it. However, it is easy to discriminate that O_2B and COB have noticeable opposite behaviors: when one is rising, the other one should decrease, which technologically has a consistent and evident meaning. Several such

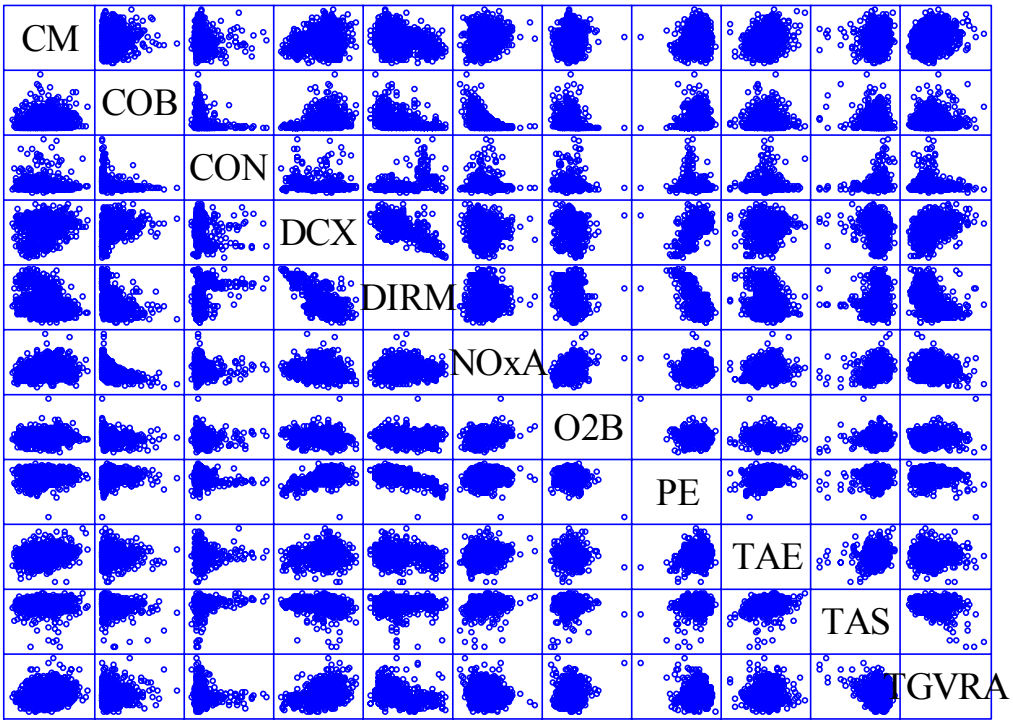


Fig. 3 - *Cazul 2A* corelații ale fiecărei perechi de variabile / *Case 2A bivariate correlations plot.*

Din acest motiv, fiecare eliminare de înregistrări trebuie abordată cu prudență, ținând seama şi de posibila afectare negativă a coeficientului de autocorelație în fiecare serie de timp.

cases could be further observed (and technologically explained well) especially in figure 5b, considering the position of the points over the two axes. However, not all information extracted from that analysis is reliable, for example: DCX

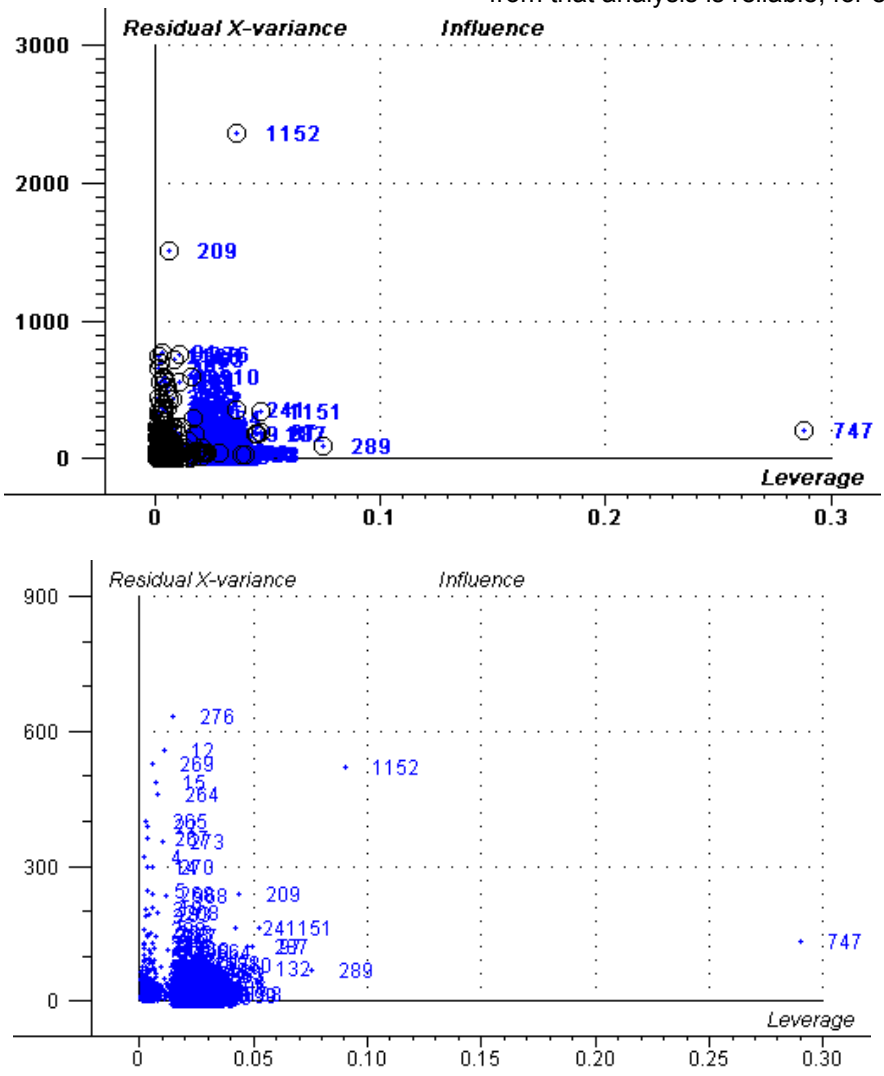


Fig. 4 - Rezultate ale analizei statistice a) *Cazul 1*; b) *Cazul 2A* / *Influence plots a) Case 1; b) Case 2A.*

Tabelul 3

a) Coeficienți de formă / *Shape parameters*

Cazul *Case*		CM	COB	CON	DCX	DIRM	NO_xA	O_2B	PE	TAE	TAS	TGVRA
1	Coeficienți de asimetrie *Stnd. skewness*	1.92	34.35	73.10	-41.63	15.11	9.25	64.45	-21.20	-4.31	-35.07	16.07
	Coeficienți de boltire *Stnd. kurtosis*	-2.35	57.01	226.63	198.40	4.48	9.04	428.06	48.83	13.48	96.89	13.27
2A	Coeficienți de asimetrie *Stnd. skewness*	1.99	34.15	69.69	-12.59	14.79	8.91	6.55	-15.20	-3.54	-33.02	16.0
	Coeficienți de boltire *Stnd. kurtosis*	-2.30	57.18	213.92	4.85	2.96	8.61	44.80	20.10	12.50	95.05	13.39

b) Corelații / *Correlation*

Cazul *Case*		CM	COB	CON	DCX	NO_xA	O_2B	PE	TAE	TAS	TGVRA
1	DCX	0.1921	0.2407	-0.2196		-0.1066	-0.2926	0.6050	0.1701	-0.1211	0.2629
	DIRM	-0.3048	-0.3356	0.3888	-0.6801	0.1222	-0.0676	-0.6691	-0.2052	0.0652	-0.3234
2A	DCX	0.2204	0.2488	-0.2244		-0.1175	-0.0957	0.6204	0.1347	-0.1920	0.3084
	DIRM	-0.3078	-0.3409	0.3963	-0.7523	0.1030	-0.1440	-0.6661	-0.1993	0.0976	-0.3509

Pe baza rezultatelor calculelor de corelație (tabelul 3b, figura 3) şi corelație parțială (care nu sunt prezentate aici ca şi multe alte rezultate datorită necesarului mare de spațiu) şi a testelor PCA arătate în figura 5 a şi b, au fost eliminate trei variabile: NO_xA, TAS şi TAE, obținându-se astfel *Cazul 2B* caracterizat prin cea mai bună valoare a coeficientului R^2 pentru variabila DIRM şi care furnizează şi cele mai bune rezultate de predicție cu metoda regresiei liniare multiple. O analiză simplă a tabelului 3b arată că cele trei variabile prezintă corelații slabe cu cele două mărimi dependente sau chiar că valorile nu prezintă o evoluție ascendentă (de fapt, unele dintre ele sunt în scădere). Mai mult, *Analiza Componentelor Principale* în figura 5a arată că aceste trei variabile sunt "diferite" de celelalte. De fapt, rezultatele PCA sunt dificil de interpretat corect în alt mod decât din perspectiva experienței anterioare asupra procesului modelat, ceea ce poate induce un anumit grad de subiectivism. Totuşi, este usor de observat că O_2B şi COB prezintă comportamente contrare: când unul dintre ele creşte, celalalt scade, ceea ce este principial corect şi din punct de vedere tehnologic.

and TGVRA should have a similar behavior being close to each other, which can not be observed in figure 1. They may have the same influence over the variability of the data; for that reason, PCA was treated with care and information gathered were not overestimated in the paper. On the other hand, the intricacy of the existing relationships can affect the results of the PCA: take for example the subset CON and PE. In *Case 1* (figure 5a) they could be assigned with a similar behavior. It is clearly visible in figure 1 that, even if they both react at mostly the same data records, the manner they react is in fact the opposite, which becomes correct from the analysis only in figure 5b.

Cluster Analysis is another technique for displaying where similarities lie. The essence is the same, similar variables/individuals will tend to react similarly (more in-depth information about techniques used here could be found in [7]). First, variables were automatically grouped within clusters – see figure 6 – by using *Furthest Neighbor Method*. Two major groups resulted, groups that were split further. Some correspondences with results given in figure 5b, judged by position from the center of the figure

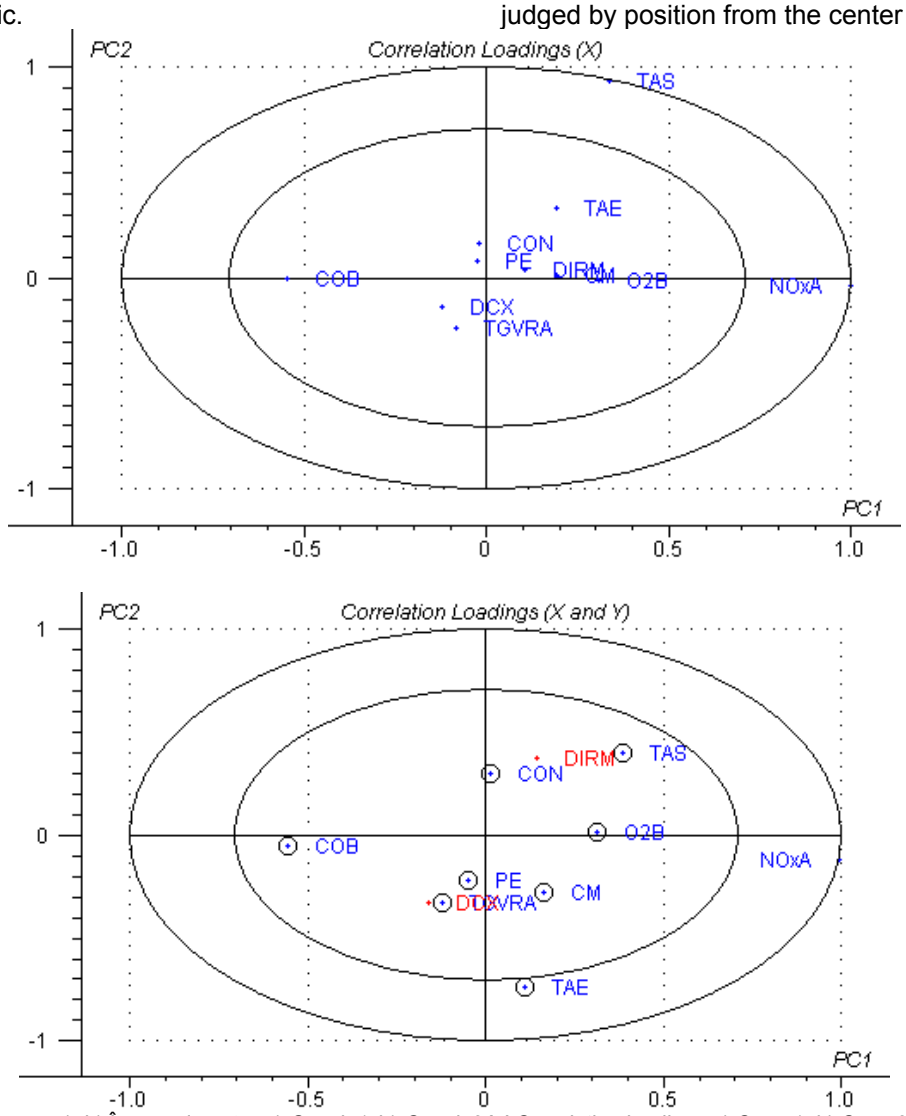


Fig. 5 - a), b) Încărcări pentru a) *Cazul 1;* b) *Cazul 2A I Correlation loadings a) Case 1; b) Case 2A.*

Alte câteva astfel de observații se pot identifica în figura 5b – şi explica sub aspect tehnologic – prin simpla apreciere a poziției punctelor în raport cu cele două axe. Este de subliniat totuşi că nu toate informațiile extrase din aceasta analiză sunt de încredere, de exemplu: DCX şi TGVRA ar trebui să aibă comportamente asemănătoare, fiind apropiate, ceea ce nu poate fi observat în figura 1 (cel putin la o analiză vizuală). Aceste variabile ar putea avea aceeaşi influență asupra variabillității datelor; din acest motiv, PCA a fost tratată cu prudență iar informațiile obținute nu au fost supraestimate în lucrare. Pe de altă parte, complexitatea relațiilor existente între parametrii considerați poate afecta rezultatele PCA; spre exemplu, se consideră două mărimi: CON and PE. În *Cazul 1* (figura 5a) aceştia pot fi considerați a avea un comportament similar. Este vizibil, însă, în figura 1 că, chiar dacă modificările pe care aceştia le suportă aparțin, în mare parte, aceloraşi înregistrări, sensul de variație al lor este de fapt opus, ceea ce devine cunoscut doar prin analiza figurii 5b.

Analiza clusterilor este o altă tehnică pentru a identica similaritățile. Principiul este acelaşi, conform căruia variabilele/valorile individuale apropiate vor avea un comportament similar (informații detaliate se pot găsi în [7]). Variabilele au fost grupate în clusteri în mod automat cu ajutorul metodei *Furthest Neighbor* (vezi figura 6). Au rezultat două grupuri majore care se ramifică în continuare. Se pot extrage o serie de corespondețe cu rezultatele date în figura 5b, judecând după poziția față de centrul elipsei şi după gradul de proximitate al variabilelor. Spre exemplu, DIRM şi CON sunt apropiate în ambele figuri; aceeaşi afirmație poate fi făcută pentru cuplurile DCX şi PE, TGVRA şi CM.

Se confirmă din analiza vizuală (vezi figura 1) că primele două perechi prezintă comportamente asemănătoare, în timp ce pentru a treia o astfel de apreciere este dificil de susținut. Distanța dintre variabile în interiorul structurii de clusteri este cea mai mare pentru acest ultim cuplu dintre toate patru perechile formate, incluzând printre acestea şi NO_xA şi TAS.

În ceea ce priveşte înregistrările, acestea au fost grupate în clusteri cu ajutorul metodei *Self Organized Maps* (SOM), pe baza unui scop similar. Cazul curent, denumit *Cazul 3,* a rezultat ca fiind cel mai coerent şi consistent cu rezultatele analizei vizuale care a putut fi obținut (corespunzător unui număr minim de parametri) şi a fost urmărit doar pentru se obține informații (de orice natură) sau confirmări posibile prin această metodă.

O procedură iterativă, euristică, a fost utilizată pentru eliminarea variabilelor şi înregistrărilor, conducând la o configurație de şase variabile şi 1204 înregistrări care a dat cel mai bun răspuns. Informații despre metoda SOM pot fi găsite în [8].

and/or proximity criteria, could be drawn. For example, DIRM and CON are close together in both figures and the same for DCX and PE, TGVRA and CM. It is confirmed from the visual analysis (see figure 1) that the first two pairs have very close behaviors, while for the third one is difficult to substantiate. The distance between variables within clusters is the biggest among all four pairs initially formed, considering NO_xA and TAS as well.

Data records were grouped in clusters with the aid of *Self Organized Maps* (SOM) with the same aim in mind. The present case, designated *Case 3*, was the most coherent and consistent with visual analysis results we obtained, and it was made mostly to observe if any information or confirmation could be acquired by this method. An iterative heuristic procedure was used here, which led to a configuration of 6 variables and 1204 records that gave the best response. Information about SOM could be found in [8].

The trained SOM is given in figure 7. In this example, SOM nodes contain 0 (left edge on the

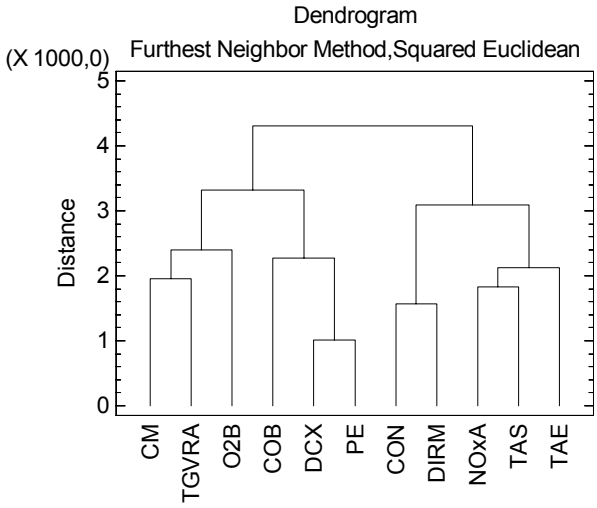


Fig. 6 - Dendrograma corespunzătoare *Cazului 2A* / *Cluster plot in Case 2A.*

middle, approximately) to up to 23 samples (left edge on the lower corner). Four clusters were automatically built.

The information we can gather from figure 7 could be resumes as: PE and DCX have very similar behaviors (that observation is only another confirmation of the different prior analysis results); DIRM and CON could be assigned with a certain resemblance in their behavior. The overall behavior of these four variables acquired from the SOM analysis is confirmed by the rules extracted from the visual analysis. It now became appealing to pinpoint the influence of the other two parameters, COB and CM. It was observed that, when removing one or both variables from the analysis, although an overall resemblance with figure 7 still subsist, regions (islands) of opposite colors appear within existing well defined domains, though lessening or even obliterating the observed

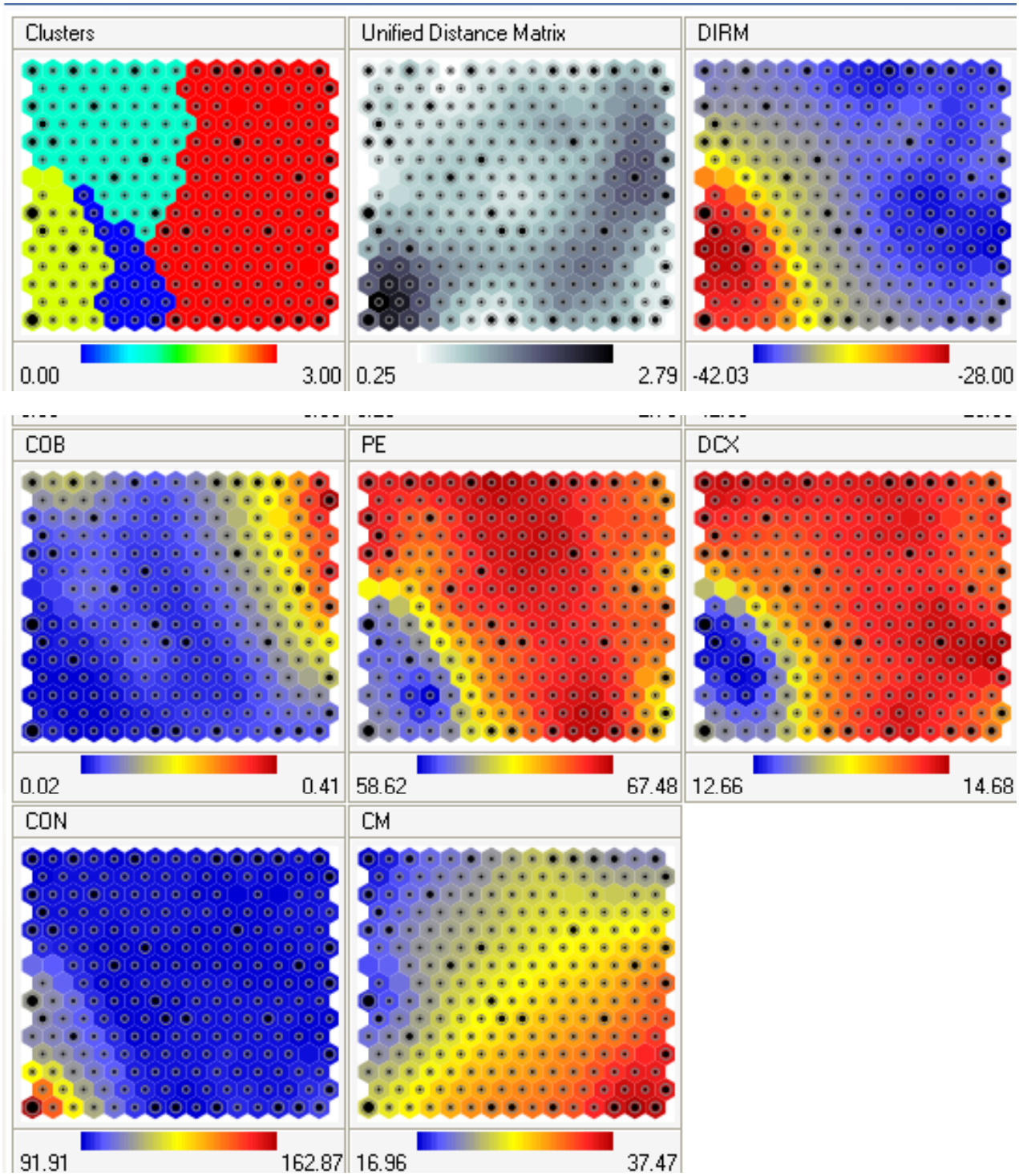


Fig. 7 - SOM pentru *Cazul 3* / SOM for *Case 3*.

Rețeaua SOM antrenată este prezentată în figura 7. În acest exemplu, nodurile rețelei SOM conțin 0 (pe latura stângă, aproximativ în mijloc) până la 23 înregistrări (latura stângă, colțul inferior al figurii). Patru clusteri au fost construiți automat. Informațiile pe care le putem obține din examinarea figurii 7 sunt: PE şi DCX prezintă comportamente asemănătoare (această observație reprezintă o altă confirmare a diferitelor rezultate date în articol); DIRM şi CON pot fi considerate a avea un relativ grad de asemănare în comportamentul lor.

correlation. Note that, except for a small region in the lower left corner, most of the unified matrix has light gray regions which indicate that records are both physically close to each other and similar. The opposite – the black region – indicates that even close, records are not similar. Bearing in mind these observations drawn from the SOM, it was considered that this analysis tool could support the conclusion that this configuration is the smallest, still balanced one (in terms of variables and data records numbers) to provide the least

Comportamentul general al acestor patru variabile, extras din analiza SOM este confirmat de către regulile rezultate din analiza vizuală. Este interesant de urmărit şi influența celorlați doi parametri, COB şi CM. [S-a observat că, atunci când sunt eliminate una sau ambele variabile din analiza SOM, deşi subzistă un o asemănare generală cu figura 7, apar regiuni (insule) de culori distincte în interiorul unor domenii bine definite, slăbind sau chiar blocând corelațiile existente]. Este de apreciat în mod pozitiv că, în afara unei regiuni de arie mică, existentă în colțul inferior-stânga, cea mai mare parte a matricii unificate prezintă regiuni cenuşiu-deschis, ceea ce indică faptul că înregistrările sunt nu numai apropiate din punct de vedere fizic ci au şi comportament similar. Considerând aceste observații deduse din analiza SOM, se poate concluziona că cea mai mică dar echilibrată configurație posibilă, care poate genera o corelație acceptabilă pentru variabila-obiectiv DIRM (vezi de asemenea tabelul 1) este dată chiar de către *Cazul 3*. S-a constatat în mod evident că dacă se elimină variabile sau/şi înregistrări din configurația corespunzătoare *Cazului 2A*, modelele obținute vor avea performanțe inferioare acestuia, datorită diminuării coeficienților de corelație şi autocorelație (apar discontinuități în secvențele temporale ale tuturor variabilelor).

Variabila DIRM poate constitui un bun exemplu al principiului Pareto, care aplicat în aceste cazuri, afirmă că un volum mic de date poate avea un efect decisiv asupra eficienței modelului. Se poate observa în Figura 8 faptul ca ceva mai putin de 20% dintre înregistrări (categorie numită “Other”) şi care iau valori în intervalul 25 până la 33-34 mbar corespund picului secundar din Figura 2b şi regiunilor A, B şi parțial D in Figura 1. S-ar putea extrage de aici în mod subiectiv concluzia că acest mic pic secundar poate influența în mod nesemnificativ variabilitatea datelor. Totuşi, prin eliminarea nejustificată printr-o analiză pertinentă a acestor înregistrări, se va pierde şi informație importantă – aşa cum rezultă din analiza vizuală; aceste înregistrări oferă o corelație importantă cu celelelalte variabile.

DCX are o valoare R^2 nesemnificativă în *Cazul 3*; corelațiile pe perechi între DCX, respectiv DIRM, şi cei doi parametri (COB şi CM), care începând cu *Cazul 1* au favorizat deja DIRM, au încă valori mari pentru DIRM şi foarte scăzute pentru DCX în comparație cu *Cazul 1*.

Concluzii

În lucrare se relevă faptul că, prin utilizarea a mai mult de o singură tehnică de analiză se obțin rezultate cu un grad mai mare de siguranță în ceea ce priveşte extragerea de cunoştințe dintr-o bază de date bidimensională. O îmbunătățire suplimentară, dar absolut necesară, în alegerea corectă a mărimilor dependente şi independente decent correlation for DIRM (see also table 1). It was seen that, when extracting variables and/or data records from *Case 2A* configuration, models will drop in performance along with a decrease in correlation and autocorrelation coefficients (discontinuities in all variables’ time sequences start to appear).

DIRM could be a good example of the Pareto principle, when a small amount of data could have a decisive effect on the model’s efficiency. It could be seen in figure 8 that less than 20% of the records (called “Other”) ranging from 25 to 33-34 mbar correspond to the secondary peak in figure 2b and to regions A, B and partially D in figure 1. It could be theoretically inferred that this small secondary peak has a small significance over the variability of the data. However, by indiscriminately removing these records, information will be lost - as it becomes obvious from the visual analysis; they provide an important correlation with other variables.

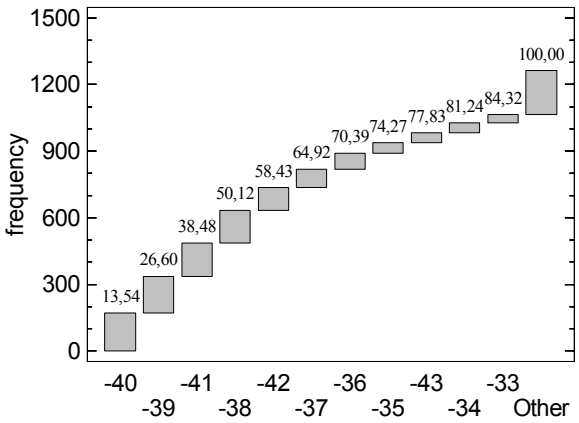


Fig. 8 - Graficul Pareto pentru variabila DIRM, *Cazul 1* / DIRM Pareto chart for *Case 1*.

DCX had an insignificant R^2 value in *Case 3*; computed bivariate correlations between DCX and DIRM and the two parameters (COB and CM), which starting from *Case 1* already favored DIRM, were still high for DIRM and very low for DCX as compared to *Case 1*.

Conclusions

The employment of more than one analysis techniques could be a more secure approach to distill knowledge from a given 2-D array of numbers, as it was shown in this paper. However, the proper selection of the predictors and regressors, and also of the samples to be used when building sound models should be censored by constantly reporting to the efficiency of the prediction itself. Analysis should be considered as being a step – a necessary one as it is able to give a better perception on the connections that exist between process parameters, yet not sufficient –

precum şi a înregistrărilor poate fi făcută doar prin corelarea permanentă a rezultatelor de analiză cu cele de predicție. Analiza trebuie considerată doar o etapă necesară, deoarece asigură o percepție îmbunătățită a conexiunilor existente între parametrii de proces, dar nu şi suficientă şi trebuie completată prin etapa de predicție. Această etapă constituie scopul următoarei parți a articolului.

În lucrare au fost folosite următoarele pachete software: Centurion Statgraphics pentru MVA şi dendograme, CAMO The Unscrambler pentru MVA, Peltarion Synapse pentru analiza SOM.

and must be completed by prediction. This step will constitute the aim of the second part of the article.

Software packages used in the paper: Centurion Statgraphics for MVA and dendrograms, CAMO The Unscrambler for MVA, Peltarion Synapse for SOM analysis.

REFERENCES

1. Xi-song Chen, Shi-hua Li, Jun-yong Zhai and Qi Li, Expert system based adaptive dynamic matrix control for ball mill grinding circuit, Expert Systems with Applications, 2009 **36**, 716.
2. Z, Ghizdăveț, Mathematical modeling in cement industry, PhD Thesis, University Politehnica Bucharest, 2006.
3. T. Aizawa, and T. Yokoshita, Applications of neural network technology to the operation and control of cement plants, ZKG 1995, (10), 532.
4. D. Schmidt, Online prediction of the free lime content in the sintering zone and the use of neural networks for process optimization, ZKG International, 2001, (9).
5. H.F. Meier, K. Ropelato, M. Mori, K.J.J. Iess, and H. Forster, Computational Fluid Dynamics for cyclon evaluation and design, ZKG International, 2002, **55** (4), 37.
6. xxx, CAMO Software, The Unscrambler Appendices: Method References, 2008.
7. D.W. Stockburger, Multivariate statistics: concepts, models, and applications, Web version, 1998.
8. T. Kohonen, Self-organizing maps, Springer, 2001.

****

MANIFESTĂRI ŞTIINȚIFICE / SCIENTIFIC EVENTS

The 11th International Conference on Ceramic Processing Science - ICCPS - 11
Zürich, 29. Aug. - 1. Sep. 2010

The activities at the conference will be enhanced with plenary and invited lectures given by internationally distinguished scientists, oral and poster presentations, covering a broad range of fundamental and applied topics, such as:

- **ceramic films**
- **ceramic shaping**
- **metal/ceramic interfaces**
- **powder synthesis**
- **forming and sintering of ceramics**
- **novel characterization methods**
- **modeling and simulation**

For further information mail to: **iccpsinfo@ethz.ch**

http://www.iccps11.ethz.ch/home

****