APPLIED COMPUTER SCIENCE

The Journal is a peer-reviewed, international, multidisciplinary journal covering a broad spectrum of topics of computer application in production engineering, technology, management and economy.

The main purpose of Applied Computer Science is to publish the results of cutting-edge research advancing the concepts, theories and implementation of novel solutions in computer technology. Papers presenting original research results related to applications of computer technology in production engineering, management, economy and technology are welcomed.

We welcome original papers written in English. The Journal also publishes technical briefs, discussions of previously published papers, book reviews, and editorials. Especially we welcome papers which deals with the problem of computer applications in such areas as:

- manufacturing,
- engineering,
- technology,
- designing,
- organization,
- management,
- economics,
- innovations,
- competitiveness,
- quality and costs.

The Journal is published quarterly and is indexed in: BazTech, Cabell's Directory, CNKI Scholar (China National Knowledge Infrastucture), ERIH PLUS, Index Copernicus, J-Gate, Google Scholar, TEMA Technik und Management.

Letters to the Editor-in-Chief or Editorial Secretary are highly encouraged.

CONTENTS

Applied Computer Science, vol. 15, no. 4, pp. 5–15
doi:10.23743/acs-2019-25

Submitted: 2019-09-19
Revised: 2019-09-24
Accepted: 2019-12-01

MRP II, linear optimization, relational databases, ECL^{i}PS^e-CLP.

Jarosław WIKAREK [0000-0003-4758-1490]*,
Paweł SITEK [0000-0001-6108-0241]*, *Mieczysław JAGODZIŃSKI*****

A DECLARATIVE APPROACH TO SHOP ORDERS OPTIMIZATION

Abstract

The paper presents the problem of material requirements planning with optimization of load distribution between work centers and workers' groups. Moreover, it discusses the computational example for shop orders optimization. The data for this example were taken from the relational database. The method of Constraint Logic Programming (CLP) for shop orders optimization has been suggested. Using Constraint Logic Programming, the constraints may be directly introduced to the problem declaration, which is equivalent to the source code of the program. The ECL^{i}PS^e-CLP software system has been presented. It allows for solving optimization problems concerning dimensions greater than in the case of the professional mathematical programming solver "LINGO". The application of ECL^{i}PS^e-CLP in accessing data from relational databases has been presented.

1. INTRODUCTION

Information systems have long been an important part of the manufacturing environment. In the 1960s, manufacturers developed Material Requirements Planning (MRP). According to the American Production and Inventory Control Society, Inc. (APICS), MRP is a set of techniques that uses bill of material data, inventory data, and the master production schedule to calculate requirements for materials. It makes recommendations for reordering materials. Furthermore, because it is time-phased, it makes recommendations for rescheduling open

* Technical University of Kielce, Control and Management Systems Section, 1000-PP 7, Kielce, Poland, j.wikarek@tu.kielce.pl, sitek@tu.kielce.pl
** Silesian University of Technology, Department of Automatics, Akademicka 16, 44-101 Gliwice, Poland, mieczyslaw.jagodzinski@polsl.pl

orders when due dates and need dates are not in phase. Time-phased MRP begins with the items being listed on the Master Production Schedule and the determination of the quantity of all components and materials required to fabricate those items, and the date that the components and materials are required. Time-phased MRP is accomplished by exploding the bill of material, adjusting for inventory quantities on hand or on order and offsetting the net requirements by the appropriate lead times.

In the late 1970s and early 1980s, manufacturers integrated MRP (Material Requirement Planning) and other manufacturing and business functions. This renaissance is commonly known as Manufacturing Resource Planning (MRP II). According to the American Production and Inventory Control Society, Inc. (APICS), MRP II is a method for the effective planning of all the resources of a manufacturing company. Ideally, it addresses operational planning in units, financial planning in currency, and has a simulation capability to answer "what if" questions. It is made up of a variety of functions, each linked together:

- business planning,
- sales and operation planning,
- demand planning,
- master scheduling,
- material planning,
- capacity planning,

and the execution support system for capacity and material. Output from these systems can be integrated with financial reports (Landvater & Gray, 1989). MRP II has also been defined as a set of software modules based on an integrated database. Within this paradigm, the main modules are MPS (Master Production Schedule), MRP (Material Requirements Planning), CRP (Capacity Requirements Planning) and SFC (Shop Flow Control). Decisions for each layer of the manufacturing system are generated and processed in these modules. In the MPS module, master production schedule orders are generated based on sales forecasts and information about real customer orders. In the MRP module, planned orders are generated based on master production schedule orders, inventory stock, bill of material, etc. The SFC module, together with the CRP module, generates shop orders on the ground of planned orders and information about work centers' capacity, routings and real inventory stock. These decisions may be more or less detailed, according to their layers and areas of influence.

2. DECISIONS IN MRP II SYSTEMS

The MRP II standard does not assure optimum values of generated orders. What is more, their feasibility is not assured automatically. Before the realization of shop orders, the charge of work centers and of workers' groups, which depends on them, is checked in the CRP module. If the work centers' capacity is insufficient, then the planned orders are corrected in dialogue with the user of the computer system.

The user of the MRP II system changes the schedule of allowance of work centers regarding planned orders for as long as the CRP module signals insufficient production capacity. If this procedure is not successful, the user should change the planned orders that have been received previously from MRP module. Sometimes, changes of master production schedule orders, made after negotiations with customers, are inevitable.

3. SHOP ORDERS OPTIMIZATION

For automating and optimizing the decision procedure described in chapter 2, the mathematical model of material requirements planning with optimization of load distribution between work centers and workers' groups has been suggested. This model has been formulated as a linear programming problem. A detailed discussion of its objective function (1) and constraints has been presented in (Sitek, Wikarek & Zaborowski, 2002).

The main part of the goal function (1), which is minimized in the presented model of optimization of load distribution between work centers and workers' groups, is the cost of charge for work centers and workers' groups. A work center is a group of similar/identical machines, each of them capable of executing every shop task from a definite set of tasks allotted for a given center. A workers' group consists of workers with the same specializations and load cost per hour. Each of them could execute every shop task from a definite set of tasks allotted for a given group. The constraints of decision variables in this model are a formal representation of the constraints in the MRP II method. Therefore, every feasible solution of this model (constraints) represents the planned orders which have been executed without needs of later modification. Possible corrections in chance of contradicting the constraints are MPS orders. Therefore, one introduced to the goal function the corrective variable of the MPS order quantity $X_j(\tau)$. The penalties for the product j storage and above corrective variables of the MPS were additional elements of the goal function. These additional elements have to assure agreement among the others with the MRP II standard (Landvater & Gray, 1989).

The area of shop orders optimization has been shown in Figure 1.

The main decision variables of this model have been shown in Table 1. Other variables and indices are listed below:

(j) – an inventory item, $j \in J$, ($J \supset J^M$ – the set of products)
(j,k) – an operation identifier, $k \in Kj$, $j \in J$
(j,r) – a routing identifier, $r \in I_j^R$, $j \in J^M$
(j,r,k) – a shop operation identifier
h – a work center number, $h \in I^{WC}$
p – a workers group number
q – a worker specialization number
t – an operational planning period number, $t \in T$
$G_j(\tau)$ – the quantity of the MPS order (j,τ)
γ_h – the work center load cost per hour, $h \in I^{WC}$
$\Omega_{jrk}(\tau)$ – machine hours of the shop operation (j,r,k) in the period τ
δ_p – the workers group load cost per hour
β_j – the penalty for the MPS order correction of the product j
c_j – the penalty for the product j storage

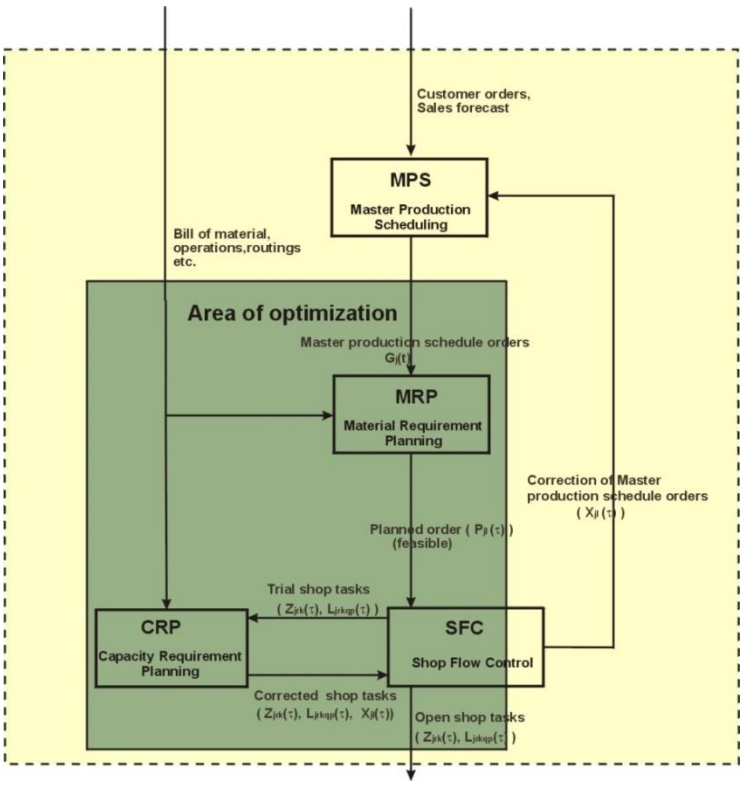

Fig. 1. The area of shop orders optimization in production control in a MRP II system

$$\sum_{\tau=1}^{\tau^h}\sum_{j,r\in J_j^R}\sum_{k\in K_j}\left\{\gamma_h\Omega_{jrk}(\tau)+\sum_{q\in I_{jk}^A}\sum_{p\in I_q^L}\delta_p L_{jrkqp}(\tau)\right\}+\sum_{j\in J^Y}\sum_{t=1}^{\tau^h+L_j^\Sigma}\beta_j X_j(\tau)+\sum_{j\in J}\sum_{t=1}^{\tau^h}c_j V_j(\tau) \qquad (1)$$

Tab. 1. Decision variables of the optimization problem

Symbol	Description
$P_j(\tau)$	Planned order quantity of the order (j,τ)
$Z_{j,r,k}(\tau)$	Shop task quantity of the task (j,r,k,τ)
$X_j(\tau)$	Corrective variable of the MPS order quantity of the order (j,τ)
$L_{j,r,k,q,p}(\tau)$	Labor hours of workers group p with specialization q resulting from the shop task quantity ($Z_{j,r,k}(\tau)$) of the task (j,r,k,τ)
$R_j(t)$	Requirement of the product j in the period t
$V_j(t)$	Requested inventory of the product j in the end of the period t, (inventory $V_j(0)$ is given)

4. COMPUTATIONAL METHODS FOR SHOP ORDERS OPTIMIZATION

Because the considered problem of shop orders optimization is a linear programming problem, the simplex method was first applied solve it (www.lindo.com, 2018). The software package "LINGO" was used to solve several optimization examples (www.lindo.com, 2018).

"LINGO" is a mathematical modeling language. Unlike conventional languages, such as Basic, Pascal or C, is nonprocedural. That is, when you specify a model for "LINGO" to solve, you only tell it what you want, not how it should find the solution. It is "LINGO's" job to worry about the how. In this sense "LINGO" is known as a specification language. "LINGO's" modeling language lets you express your problem in a natural manner which is very similar to standard mathematical notation.

This software is adequate for many examples. Unfortunately, it was not possible to find a solution for greater size problems.

Thus, it was necessary to examine an alternative method of optimization. A subsequent method, named the CLP (Constraint Logic Programming), has been applied. CLP may be defined as a body of techniques used for solving problems with constraints. The main idea of CLP is that (Rossi, Van Beek & Walsh, 2006; Niederliński, 2011):

- The problems to be solved are modeled using elementary logic, in a way that turns a model into a part of the problem-solving program.
- Exploration of the constraints, which should be satisfied by the solutions, generates solutions.

Using Constraint Logic Programming (CLP) for solving the optimization problem (Niederliński, 2011), its constraints, and the objective function (1) may be directly introduced to the problem declaration which is equivalent to the source code of the software program.

5. DATA STRUCTURE

The heart of most software implementations of the MRP II system is an integrated database. There are different database technologies available: hierarchical databases, network databases, relational databases and object-oriented databases. The relational database technology is, by virtue of being both the subject of international standards and a solid theoretical platform, by far the most widely spread database technology today. It is commercially available from a great number of vendors, like Oracle, Informix, Sybase etc. A relational database is organized in tables. A tables represents a real world entity or concept, like a customer, a vendor, orders, invoices, machines etc. Each table is made up by a number of columns, some of which are used for data storage and others for keeping references to other tables. Together, the tables build a more or less complex structure, which will be referred to as the database structure. An exemplified database structure for the MRP II environment has been suggested in (Landvater & Gray 1989).

The data for the shop orders optimization were taken from relational databases. When using the CLP method or the "LINGO" software package for solving optimization problems, it was necessary to organize access to data (Figure 2). In the MRP II system, all data and decisions are stored in relational databases. When using the ECL^{i}PS^e-CLP (www.eclipse.org, 2018) language in the above environment, it was necessary to solve some problems. The most important were:

- Writing procedures for data export from a relation database to text files;
- Writing predicates for data import from text files to lists of coefficients
- in the ECL^{i}PS^e-CLP language;
- Implementing the optimization model to the source code of the software program in ECL^{i}PS^e-CLP;
- Starting optimization;
- Writing predicates for saving output data to text files;
- Writing procedures for data import from text files to a relation database.

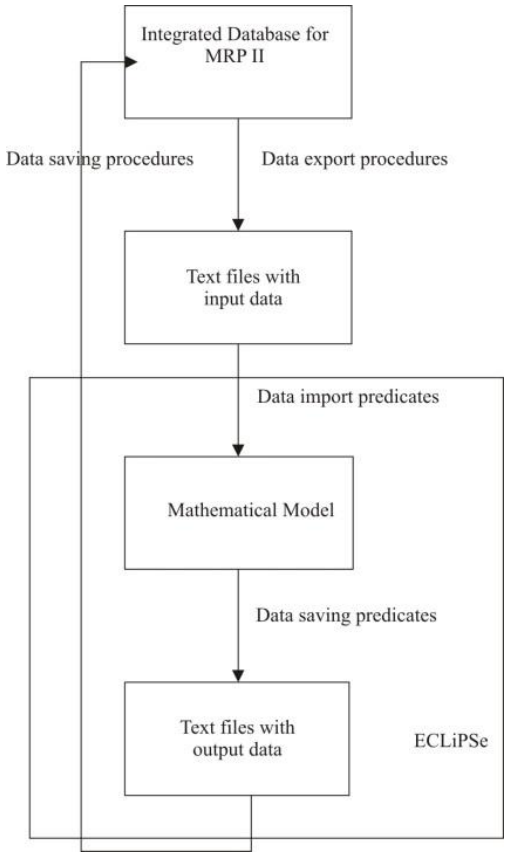

Fig. 2. Access to data for optimization shop orders in MRP II system

6. APPLICATION ECL^{i}PS^e-CLP TO SHOP ORDERS OPTIMIZATION

The implementation of the ECL^{i}PS^e-CLP (Niederliński, 2011; www.eclipse.org 2019) language in decision optimization has been shown in the example of shop orders optimization. One of the most important constraints of the above problem are material balance equations (2), defining net requirements in the given time period, which are equal gross requirements regarding inventory. Planned gross requirements are defined in every time period for every item.

$$V_j(\tau) = V_j(\tau-1) + P_t(\tau) - R_j(\tau) \text{ for } \tau = 1..T + L_j^{\Sigma} \quad j \in J\,, \tag{2}$$

The details of the implementation of the material balance constraint and access to data in the ECLIPSE-CLP language have been presented in the listings (from listing 1 to listing 7).

In addition, when implementing the above problem in ECLIPSE-CLP, it is possible to split the computation into several parts (computational levels). These parts have been shown in Table 2. Using ECLIPSE-CLP and splitting the optimization problem in parts have resulted in a more effective optimization and the possibility of solving problems of size greater than in the case of using the "LINGO" system.

```
% PREDICATE FOR MAKING LIST
% parameters
% 1 - the name of the list
% 2 - the end of previous list
% 3 - the size of the list,
% 4 - domain size,
li_t(L,L,0,0).
 li_t([Zm|R],L,Nr,X):-
  Zm::0:X,
  Nr_1 is Nr-1,
  ( Nr_1#=0 then
   li_t(R,L,Nr_1,0);
   li_t(R,L,Nr_1,X)).
```

Fig. 3. Listing of predicate for making list

```
L_j ::1..300,% number of items,
L_t ::1..20, % number of time periods
L_t1::1..20, % number of time periods
L_jt::1..6000,% product of items and time periods
```

Fig. 4. Listing of declaration of domain sizes

```
stale(L_j,L_t),
 open('!dane\\Stale.txt',S,'r'),
 read(S,Pomo), L_j  #= Pomo,
 read(S,Pomo1),
 L_t  #= Pomo1,
 close(S).
 czytaj(Plik,Lista,D):-
 open(Plik,S,'r'),
 li_c(Lista,[],D,S),
 close(S).
 li_c(L,L,0,0).
 li_c([Zm|R],L,Nr,S):-
 read(S,Pomo), Zm#=Pomo,
 Nr_1 is Nr-1,
 ( Nr_1#=0 then
  li_c(R,L,Nr_1,0);
  li_c(R,L,Nr_1,S)).
```

Fig. 5. Listing of data import predicates

```
og_6(Pjt,Rjt,Vjt,Vjt,L_j,L_t1),
```

Fig. 6. Listing of call predicate for material balance constraint

```
stale(L_j,L_t),
 L_t1 is L_t+1,
 L_jt is L_j*L_t1,
 li_t(Vo,[],L_j,X),
 czytaj('!dane\Vo.txt',Vo,L_j),
 li_t(Rjt,[],L_j*L_t1,X),
 li_t(Vjt,[],L_j*L_t1,X),
 li_t(Pjt,[],L_j*L_t1,X),
 zapas_0(Vjt),
```

Fig. 7. Listing of call predicates

```
% PREDICATE FOR MATERIAL BALANCE CONSTRAINT
%1 - Pjt – net requirements
%2 - Rjt – gross requirements,
%3 - Vjt – inventory of item,
%4 - [A|B] – inventory of item in the next period,
%5 - L_j – number of items,
%6 - L_t1 – numer of time periods,
 og_6(Pjt,Rjt,Vjt,[A|B],L_j,L_t1):-
  og_6p(Pjt,Rjt,Vjt,B,L_j,L_t1,0,0).
 og_6p([C|D],[E|F],[G|H],[I|J],L_j,L_t1,Ob,Cz):-
  Ob1 is Ob+1, Cz1 is Cz+1,
  (Cz1 #< L_t1 then
  C is E + I - G,
  ( C #>0 then
   I is 0;
   (C is 0, E is 0)),
   ( Cz1 #= L_t1 then
   Cz2 is 0;
   Cz2 is Cz1, Pomo is L_j*L_t1-1),
   (Ob1 #< Pomo then
   og_6p(D,F,H,J,L_j,L_t1,Ob1,Cz2);
   og_6p(D))).
   og_6p([A|B]):-
   A is 0.
```

Fig. 8. Listing of predicate for material balance constraint

Tab. 2. Computational levels

Level	Solving sub-problem
1	Set to zero initial values of corrective variables, $X_j(\tau)$, of MPS order quantities.
2	Calculate gross requirements $R_j(\tau)$, planned order quantities $P_j(\tau)$ and planned inventory values $V_j(\tau)$.
3	Distribute planned orders $P_j(\tau)$ between particular work centers and workers groups (calculate variables $Z_{jkw}(\tau)$, $L_{jkwqp}(\tau)$). If work centers load exceeds their capacity then go back to the step 2 with nonzero values of corrective variables $X_j(\tau)$.

```
% PREDICATE FOR REQUESTED
% INVENTORY OF THE ITEM J IN
% THE END OF THE FIRST PERIOD
% 1 - [Vjt] - the list of inventory
%       item,
% 2 - [A|B] - the list of inventory
%       item,
% 3 - L_j - number of items,
% 4 - L_t - number of time periods,
% 5 - Obe - number of current
%      processing item,
zapas_0(Vjt,[A|B],L_j,L_t,Obe):-
 Ob1 is Obe+1,zap_pom(Vjt,A,L_j,L_t,Ob1,0,1,0),
 ( Ob1 #< L_j then
 zapas_0(Vjt,B,L_j,L_t,Ob1)).
```

Fig. 9. Listing of predicate for requested inventory

7. NUMERICAL EXPERIMENTS

Numerous computational experiments were performed to verify the model and the proposed approach. Individual experiments differed in the number of products (N from 5 to 40) performed as part of the order. Two possible execution routes were adopted for each product, which contained from 3 to 5 operations. The availability of 20 employees with 6 specialties was also assumed.

For this data, the proposed problem was modeled and solved in two environments, i.e. the classic mathematical programming environment – LINGO – and in a declarative environment, constraint logic programming – ECL^{i}PS^e-CLP. The results are presented in Table 3. The advantage of using the declarative environment is clearly visible due to the time of calculations.

Tab. 3. Computational levels

N	LINGO		ECL^{i}PS^e-CLP	
	Time [s]	Fc	Time [s]	Fc
5	25	234	10	234
10	56	456	14	456
20	145	934	24	934
30	345	1435	34	1435
40	600	NFSF	45	2034

8. CONCLUSION

The results of calculations for numerical experiments, obtained by using the "LINGO" system and the ECL^{i}PS^e-CLP language, are the same. Other examples proved that using an ECL^{i}PS^e-CLP language software system enables solving optimization problems of dimensions greater than in the case of "LINGO". The application of the ECL^{i}PS^e-CLP language to optimization and also to access data from relational databases proved to be an interesting solution. In the future, to model and solve this problem, it is planned to use a proprietary hybrid approach that integrates both of these environments (Sitek & Wikarek, 2018, 2019).

REFERENCES

Landvater, D. V., & Gray, C. D. (1989). *MRP II Standard System*. Oliver Wight Publications.

Niederliński, A. (2011). *A Quick and Gentle Guide to Constraint Logic Programming via ECLiPSe*. Gliwice.

Rossi, F., Van Beek, P., & Walsh, T. (2006). *Handbook of Constraint Programming (Foundations of Artificial Intelligence)*. New York, NY, USA: Elsevier Science Inc.

Sitek, P., & Wikarek, J. (2019). Capacitated vehicle routing problem with pick-up and alternative delivery (CVRPPAD): model and implementation using hybrid approach. *Annals of Operations Research, 273*, 257. doi:10.1007/s10479-017-2722-x

Sitek, P., & Wikarek, J. (2018). A multi-level approach to ubiquitous modeling and solving constraints in combinatorial optimization problems in production and distribution. *J. Appl. Intell, 48*, 1344. doi:10.1007/s10489-017-1107-9

Sitek, P., Wikarek, J., & Zaborowski, M. (2002). Application of Constraint Logic Programming (CLP) to Shop Orders Optimization in MRP II Systems. In *Conference: 4th Workshop on Constrained Programming for Decision and Control* (pp. 59-66). Gliwice.

www.eclipse.org (n.d.) Retrieved October 19, 2018, from The Eclipse Foundation open source community website www.eclipse.org.

www.lindo.com (n.d.) Retrieved October 19, 2018, from www.lindo.com

Applied Computer Science, vol. 15, no. 4, pp. 16–26
doi:10.23743/acs-2019-26

Submitted: 2019-09-25
Revised: 2019-12-06
Accepted: 2019-12-10

stereolithography, numerical modelling, tensile test, resin

Danuta MIEDZIŃSKA [0000-0003-2503-6600]*, *Ewelina MAŁEK*, *Arkadiusz POPŁAWSKI* [0000-0002-7494-8975]*

NUMERICAL MODELLING OF RESINS USED IN STEREOLITOGRAPHY RAPID PROTOTYPING

Abstract

The presented research deals with the development of the numerical model for resins used for stereolithography (SLA) rapid prototyping. SLA is an additive method of production of models, prototypes, elements or parts of constructions with the use of 3D printing that covers photochemical processes by which light causes chemical monomers to link together to form polymers. Such method is very useful in design visualization, but also can be applied in numerical modelling for the purpose of validation and verification. In this application the resin strength parameters must be described and on the base of them the numerical material model is developed and validated. Such a study for SLA resins was presented in the paper.

1. INTRODUCTION – AIM OF RESEARCH

Stereolithography is one of the methods of rapid prototyping. Models – prototypes – are made of a special photocurable resin. The curing of the resin takes place due to irradiation with a laser light of the appropriate wavelength. The accuracy of model mapping depends on many factors, including the geometric accuracy of the stereolithographic apparatus (Attaran, 2017; Melchels, Feijena & Grijpmaab, 2010; ISO/ASTM International Standard, 2015).

* Military University of Technology, Faculty of Mechanical Engineering, Kaliskiego 2 St., Warsaw, Poland

The initial model for the stereolithographic one is the solid created in the CAD system (Fig. 1a). The accuracy of the body depends on the CAD system used and the modeling method adopted, which is a separate issue. The finished CAD model is exported to the stereolithography format, often referred to as STL (from the file name extension) (Fig. 1b). This format describes each modeled solid by means of flat triangular surfaces and normal vectors for each of them. At this stage, mapping errors may occur. However, these errors can be minimized by adopting small modeling surfaces (Nee, Fuh & Miyazawa, 2001; Kowalska-Bany & Krokosz, 2008).

Then, in the STL model, contour lines, which are used to create the SLA model, are isolated (Fig. 1c, d). The SLA model is made of layers of a predetermined thickness. For example, for 3D Systems SLA-250, the thickness of the layers can be 0.1 or 0.15 mm, respectively.

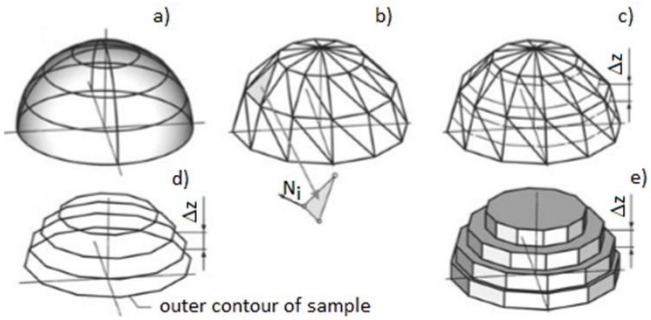

Fig. 1. Stereolithographic model creation process: a) CAD model, b) STL model, c, d) division of STL model into layers, e) ready SLA model; Δz – thickness of single layer of SLA model, N_i – normal vector (Kowalska-Bany & Krokosz, 2008)

The model is shaped on a working platform immersed in a photocurable resin. In the initial stage of model shaping, the first layer of the model is made. The platform is immersed in the resin to the depth of the assumed layer thickness of the model. The laser beam imitates the layer's outline in the resin and then hardens the area inside the overlapping stitches. After building each subsequent layer of the model, the platform is dipped again in the resin to the depth of the assumed layer thickness. The laser beam hardens the next layer of the model. The next model building sequences are similar to each other. The process continues until the entire model is mapped.

SLA technology was established as a cheaper alternative to other manufacturing methods for rapid prototyping. Its advantage is the ability to create models of complex internal and external construction, requiring high accuracy. For example, thanks to models made of a transparent resin, it is possible to create prototypes with a visible internal structure, which allows, for example, to examine traces of cooperation between gears or stresses.

In this work, photocurable resins were tested in a quasi static tensile test, then numerical analyzes were performed validating the material model selected for the resin. These studies will be used for further work, which will be analysis of porous structures with different void distributions.

2. EXPERIMENTAL TESTS

Identification tests of mechanical properties of materials are aimed at determining the basic parameters and characteristics of selected materials to compare them and define a group of materials with the best properties. The obtained data will also allow to create constitutive models of the materials studied.

The tests were carried out for standard samples printed from two different materials: Tough resin and Clear resin. The stretching speed of 5mm/min to determine and compare their mechanical properties was applied.

2.1. Samples

Standard samples for experimental research were made using rapid prototyping technology, which is 3D printing using the SLA method on the printer FormLabs 2. The first stage is to prepare the model in the PreForm software dedicated to the device, which allows you to adjust its quality (the layer height was 0.1 mm – the so-called quick print), the choice of material and the orientation of the element in the workspace. Supporting structures have been added automatically. The next step is the 3D printing process itself (Fig. 2). To ensure maximum repeatability, each process is carried out at a constant, regulated resin temperature. Such a procedure allows to reduce the viscosity of the photopolymer and to remove air bubbles from it.

Fig. 2. Printing in SLA technology before removing from work platform

One of the most demanding activities for which time should be spent in 3D printing in SLA technology is post-processing. The first step is to remove the unbound resin from the model, usually both by bath in isopropanol. The next stage of cleaning the model is required due to the process characteristics. The use of a bathtub with resin forces the use of support structures from the native material, and their removal is done mechanically. To facilitate cleaning the model from the supports, the Form2 device uses a different scan style when building them – such structures have lower mechanical strength. This is usually done by tools that allow cutting, for example, pliers or knives.

After removing the support structures, the traces in which the slides joined the model remain. In order to increase the accuracy of the stereolithography method, the laser power is somewhat limited. Too much power would cause the resin to also harden around a focused laser spot. And this is the reason of applying the last, recommended step. It consists in the final hardening of the photopolymer in the UV chamber. The treatment allows to achieve maximum mechanical and chemical properties for a given material. It should be noted that depending on the size of the model as well as the material used, the time and power of exposure may vary. In addition, the samples after exposure were subjected to temperature in a chamber preheated to 60°C during 1h for the Clear resin and 2h for Tough resin.

2.2. Static tensile test

The presented test consists in axial tensile of the normalized sample until its destruction, with a constant traverse speed at room temperature.

During the test, the force necessary to extend the sample is recorded. On the basis of the obtained data it is possible to determine the basic stress-strain characteristics, which is the basic source of information on the mechanical properties of the material.

The forces and displacements obtained from the tests were converted into engineering stresses and strains according to the following relationships:

– engineering stress:

$$\sigma_{eng} = \frac{P}{A} \tag{1}$$

where: P – applied force,
A – initial area of cross-section.

– engineering strain:

$$\varepsilon_{eng} = \frac{\Delta l}{l} \tag{2}$$

where: l – sample initial length,
Δl – change in length.

Due to the fact that the constitutive models of materials in the LS-Dyna system require the application of real strain values, taking into account the change in the cross-section of the sample due to tensile, they were calculated from the equation:

$$\varepsilon_{real} = ln(1 + \varepsilon_{eng}) \tag{3}$$

The conditions and the method of conducting the tensile testing of plastics are described in the PN-EN ISO 527-1 standard "Plastics. Determination of mechanical properties at static tension". The sample is flat and has the shape of a "paddles" (Fig. 3). The dimensions of the sample should be as follows: thickness 4.0 ± 0.2 mm, width of the gauge length 10 ± 0.2 mm and overall length over 150 mm. Table 1 presents all dimensions of the samples used in the tests.

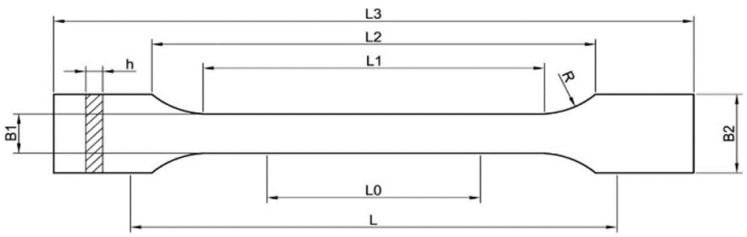

Fig. 3. Scheme of sample in accordance with PN-EN ISO 527 standard

Tab. 1. Characteristic dimensions of samples according to scheme shown in Fig. 3

Symbol	Dimensions [mm]
L1	80 ± 2
L2	20 to 25
L3	150
R	104 to 113
B1	10 ± 0.2
B2	20 ± 0.2
H	4.0 ± 0.2
L0	5.0 ± 0.5
L	115 ± 1

Samples were tensiled using a Zwick Roell Kappa 500 testing machine (Fig. 4) at room temperature of about 20°C. The stand was equipped in videoextenometer, which enabled a non-contact measurement of deformation of the sample in many axes and, thanks to special modules, measurement of the narrowest or widest place of the sample, angle of deflection, distribution of deformations in a given axis, measurement of XY coordinates of points on the plane. Dimensions of standard samples were proportionally reduced by appropriate cross-section scaling.

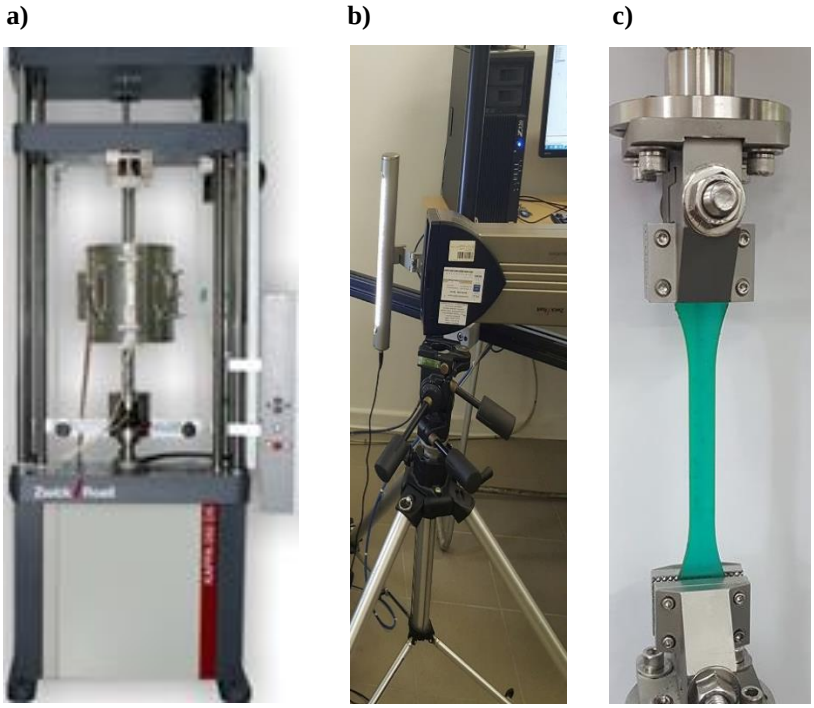

Fig. 4. View of a) Zwick Roell Kappa 500 machine, b) videoextensometer and c) sample placed in jaws of machine

2.3. Experimental tests results

On the basis of conducted experimental tests of uniaxial tensile, the force – displacement characteristics were obtained. Based on the equations (1) and (2), the engineering stress vs. engineering strain curves were determined for the 5mm/min load speed. They are summarized in Fig. 5 and 6.

Based on the curves obtained in the experiment, Young modulus, elongation at break and tensile strength were determined. The data were summarized in Table 2.

Tab. 2. Comparison of material data for Tough and Clear resin

Parameter	Tough resin	Clear resin
Young modulus	2495.72	2907.31
Elongation at break [%]	12.79	6.93
Tensile strength [MPa]	72.04	53.41

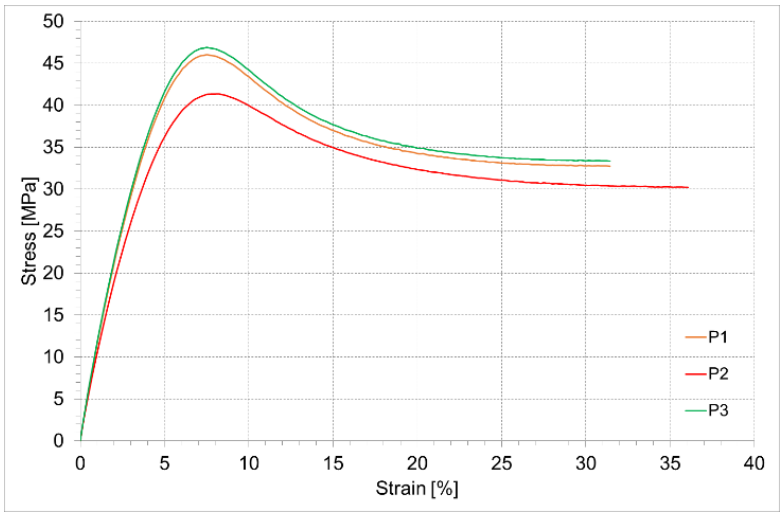

Fig. 5. Engineering stress vs. engineering strain characteristics for tensile test of Tough resin

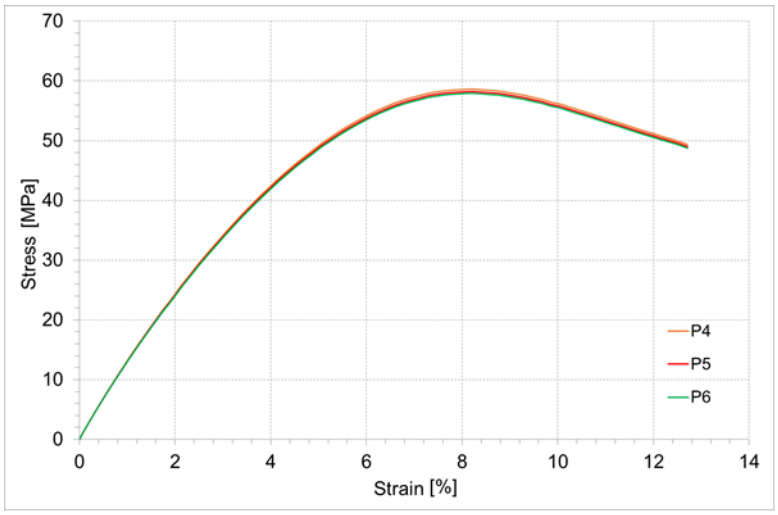

Fig. 6. Engineering stress vs. engineering strain characteristics for tensile test of Clear resin

Experimental research shows that the analyzed materials differ significantly in strength properties. Clear resin has high tensile strength, and low elongation at break. Tough resin, on the other hand, was designed like an ABS material and has better elastic properties and improved impact resistance, but lower Young modulus and tensile strength.

3. NUMERICAL ANALYSES

Simulation of quasi-static tensile was aimed at developing a numerical model of a photocurable resin material. The numerical model of the sample was shown in Fig. 7.

Fig. 7. Numerical model of a paddle sample

The model consisted of 5520 elements of the HEX8 type with the formulation of ELFORM 1. A rigid procedure for counteracting the F-B hourglass effect was used (IHQ = 4). The cross-section of the sample in the measuring area was 5.4 × 1.96 mm.

3.1. Material model selection

Due to the fact that the material will be used to simulate quasi-static compression of microstructural models, the basic elastic-plastic material MAT_024 was selected.

MAT 024 is an elasto-plastic constitutive model based on von Mises yield criteria which is used to model the material behaviour until the point where instability occur. The input parameters implemented in the MAT 024 card are primarily Young modulus, the mass density, Poisson's ratio and the hardening of the material. The hardening curve shall only cover the loading path until instability initiates (Hallquist, 2006).

Furthermore, the model supports more complex material behaviour where the material is strain rate dependent, i.e. a visco-plastic model. Instead of implementing one hardening curve, a table defining different strain rates which are connected to a certain hardening curve has to be implemented to capture the behaviour. However, MAT 024's properties are not able to express the material behaviour beyond the point of uniform expansion.

Deviatoric stresses are determined that satisfy the yield function (Hallquist, 2006):

$$\phi = \frac{1}{2} S_{ij} S_{ij} - \frac{\sigma_y^2}{3} \leq 0 \tag{4}$$

where:

$$\sigma_y = \beta\left[\sigma_0 + f_h\left(\varepsilon_{eff}^P\right)\right] \tag{5}$$

and $f_h\left(\varepsilon_{eff}^P\right)$ is the hardening function. The parameter β accounts for strain rate effects.

For complete generality a table defining the yield stress versus plastic strain may be defined for various levels of effective strain rate. In the implementation of this material model, the deviatoric stresses are updated elastically, the yield function is checked, and if it is satisfied the deviatoric stresses are accepted. If it is not, an increment in plastic strain is computed (Hallquist, 2006):

$$\Delta\varepsilon_{eff}^{P} = \frac{\left(\frac{3}{2}S_{ij}^{*}S_{ij}^{*}\right)^{1/2} - \sigma_y}{3G + E_p} \tag{6}$$

G is the shear modulus and E_p is the current plastic hardening modulus. The trial deviatoric stress state S_{ij}^{*} is scaled back (Hallquist, 2006):

$$S_{ij}^{n+1} = \frac{\sigma_y}{\left(\frac{3}{2}S_{ij}^{*}S_{ij}^{*}\right)^{1/2}} S_{ij}^{*} \tag{7}$$

Material parameters were adopted from the experimental studies presented above. Parameter responsible for damage – effective plastic strain were calculated from formula (3). For Tough resin, it was 1.153, for Clear resin – 0.67.

3.2. Numerical calculations results

Exemplary model deformations for Tough resin and stress – strain chart were shown in Fig. 8, for Clear resin – in Fig. 9,

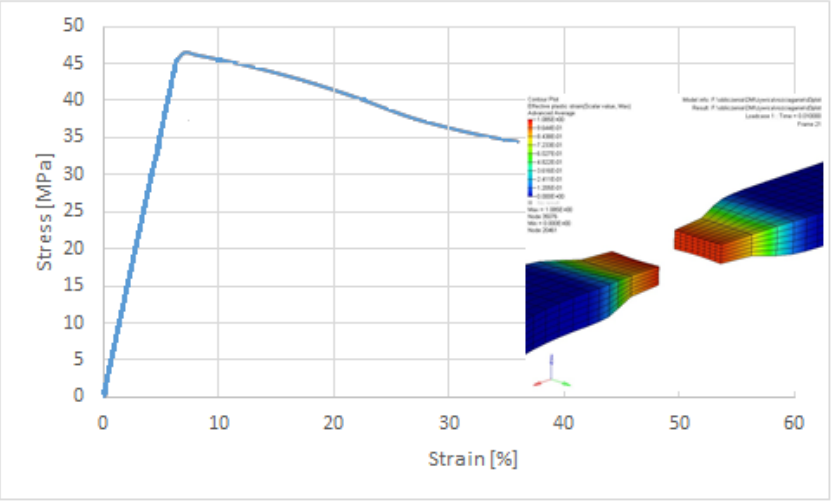

Fig. 8. Deformations and stress-strain chart for Tough resin model

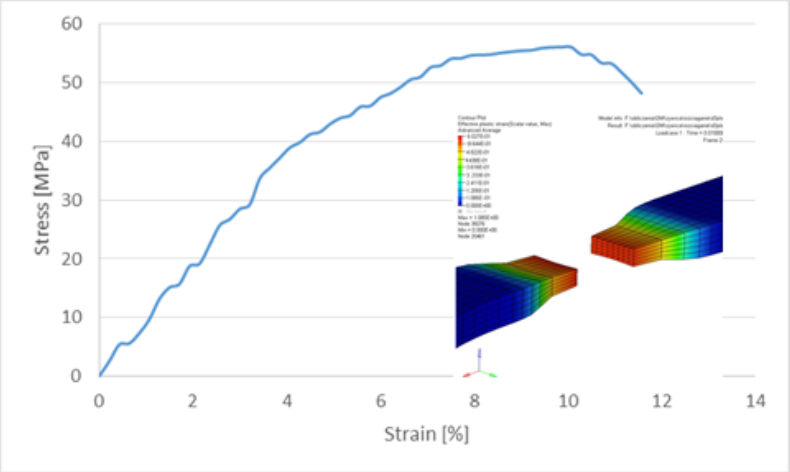

Fig. 9. Deformations and stress-strain chart for Clear resin model

4. DISCUSSION

The results of numerical and experimental tests were compared in Fig. 10.

a) **b)**

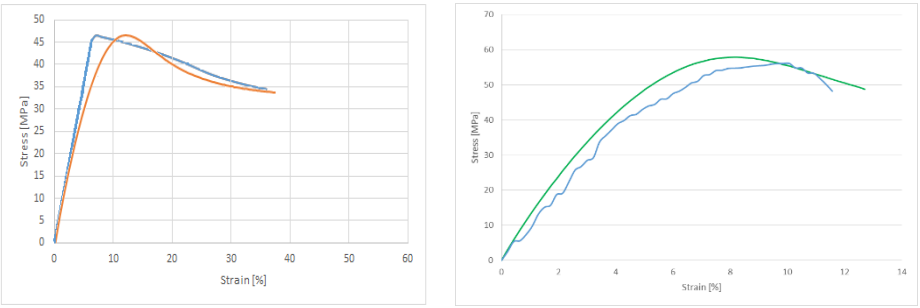

Fig. 10. Comparison of experimental and numerical tensile tests for:
a) Tough resin, b) Clear resin

The relative error for the proposed material model was calculated as the quotient of the absolute error and the exact value x_0:

$$\delta_x = \frac{\sum_{i=1}^{n} \frac{\Delta x}{x_0}}{n} = \frac{\sum_{i=1}^{n} \left(\frac{x - x_0}{x_0}\right)}{n} \tag{8}$$

where: x – measured value,
Δx – absolute error,
n – number of measuring points.

This error for the Tough resin was 12.1% for the Clear resin – 18.7 %.

After the analysis, it can be concluded that the choice of material constants is correct. Errors in numerical analysis result from simplifications of the introduced model of elastic-plastic material.

The numerical tensile test of the sample in a qualitative and quantitative manner corresponds to the real phenomenon.

ACKNOWLEDGEMENTS

The research presented in the paper was supported by grant No BG2/DIOX4SHELL/14 titled "Development of guidelines for design of innovative technology of shale gas recovery with the use of liquid CO_2 on the base of numerical and experimental research – DIOX4SHELL", supported by the National Centre for Research and Development (NCBR) in years 2014–2018.

REFERENCES

Attaran, M. (2017). The rise of 3-D printing: The advantages of additive manufacturing over traditional manufacturing. *Business Horizons*, *60*(5), 677–688.

Hallquist, J. O. (2006). *LS-DYNA Theory Manual*. Livermore Software Technology Corporation.

ISO/ASTM International Standard. (2015) *Additive manufacturing – General Principles – Overview of process categories and feedstock* (17296-2:2015(E)).

Kowalska-Bany, K., & Krokosz, J. (2008). Rapid Prototyping – Historical Overview. *Odlewnictwo współczesne – Polska i Świat*, 112–118.

Melchels, F. P. W., Feijena, J., & Grijpmaab, D. W. (2010). A Review on Stereolithography and Its Applications in Biomedical Engineering. *Biomaterials, 31*, 6121–6130.

Nee, A. Y. C., & Fuh, J. Y. H., & Miyazawa, T. (2001). On the improvement of the stereolithography (SL) process. *Journal of Materials Processing Technology, 113*(1–3), 262–268.

Applied Computer Science, vol. 15, no. 4, pp. 27–36
doi:10.23743/acs-2019-27

Submitted: 2019-09-17
Revised: 2019-11-06
Accepted: 2019-12-06

audio fade, fade-down, fade-up, real-time processing

Lucian LUPŞA-TĂTARU [0000-0002-3320-9850]*

CUSTOMIZING AUDIO FADES WITH A VIEW TO REAL-TIME PROCESSING

Abstract

To a large extent, an audio fade is distinctly acknowledged as a strict increase (fade-up) or decrease (fade-down) of the volume of an audio content. In this broad context, the widely used fade-in and fade-out sound effects, applied to receive smooth transitions from and down to silence, respectively, appear to be restrictive. Taking into account the increasing demand for multimedia techniques adapted for real-time computing, the present investigation advances straightforward procedures intended for customizing the audio fade-up and fade-down profiles, having at hand well-proven techniques of shaping the fade-in and fade-out audio effects, suitable for fast computing.

1. INTRODUCTION

In audio production, both fade-in and fade-out sound effects are extensively employed not only to smooth the beginning and the ending parts of the audio recordings but also to cross-fade various audio sections (Case, 2007; Langford, 2014; Reiss & McPherson, 2015). On the other hand, the applying of a fade-up or a fade-down sound effect simply results in a strict increase or decrease of the audio volume by a specified amount. Furthermore, the successive applications of fade-up and fade-down effects enable one to control the amplitude envelope of an audio content (Jackson, 2015; Langford, 2014; Schroder, 2011). However, similar to the case of customizing the audio fade-in and fade-out shapes, the adjustable audio fades i.e. the fade-up and fade-down sound effects are usually implemented in the off-line mode, by making use of different transcendental functions (exponential, logarithm, sine) to enforce the time-related evolution of the audio volume between imposed amplitude levels.

* "Transilvania" University of Braşov, Faculty of Electrical Engineering and Computer Science, Department of Electrical Engineering and Applied Physics, Bd. Eroilor No. 29, Braşov, Romania, lupsa@programmer.net, lucian.lupsa@unitbv.ro

To boost the computational capabilities with a view to real-time processing, which is actually required by numerous interactive products, the present approach to customizing the audio fades puts forward persuasive methods based on efficient techniques of shaping the audio fade-out and fade-in profiles, which have previously been validated for effectiveness by means of plain JavaScript implementations (Lupsa-Tataru, 2018, 2019).

2. THE AUDIO FADE-DOWN CUSTOMIZING

Assuming that the audio level decreases from the initial value $v_{D,0}$, occurring at the fade initiation, down to the final value $v_{D,f}$, showing at the fade ending, we consider that the evolution of the audio volume during fading-down is given by the output of the following function:

$$v_D(\tau_D) = v_{D,f} + \delta_D(\tau_D), \quad \tau_D \in [0, \tau_{D,f}], \tag{1}$$

wherein the variable

$$\tau_D = t_D - t_{D,ref} \tag{2}$$

is yielded by the difference of the current playback time and the chosen instant of fade-down initiation, whilst $\tau_{D,f}$ stands for the fade-down length. Since quantity $v_{D,f}$ in (1), as the final audio volume, is a constant term, it follows that the rate of change of audio level during fading-down is identical to the rate of change of δ_D in (1).

Thus, customizing the profile of the fade-down effect, provided by the output of (1), is equivalent to shaping the output of function $\delta_D(\tau_D)$ that, from the technical point of view, has to portray a fade-out audio effect. With the purpose of real-time implementation, we plainly consider (Lupsa-Tataru, 2018)

$$\delta_D(\tau_D) = \frac{\tau_D - \alpha_D}{\beta_D \tau_D - \gamma_D}, \quad \tau_D \in [0, \tau_{D,f}]. \tag{3}$$

In order for rational function (3) to describe a fade-out audio effect and, implicitly, in order for (1) to depict a fade-down audio effect, the coefficients of algebraic fraction defining (3) receive the appropriate expressions in terms of imposed maximum (initial) and minimum (final) audio volumes, denoted here by $v_{D,0}$ and $v_{D,f}$, respectively. Hence, one gets (Lupsa-Tataru, 2018)

$$\begin{aligned}
&\alpha_D = \tau_{D,f},\\
&\beta_D = \frac{2\rho_D - 1}{\rho_D \delta_{D,0}} = \left(2 - \frac{1}{\rho_D}\right)/\left(v_{D,0} - v_{D,f}\right),\\
&\gamma_D = \tau_{D,f}/\delta_{D,0} = \tau_{D,f}/\left(v_{D,0} - v_{D,f}\right),
\end{aligned} \tag{4}$$

wherein

$$\rho_D = \frac{\delta_D(\tau_{D,f}/2)}{\delta_D(0)} = \frac{\delta_{D,h}}{\delta_{D,0}}; \quad 0 < \rho_D < 1 \tag{5}$$

or, having in view (1),

$$\rho_D = \frac{v_D(\tau_{D,f}/2) - v_{D,f}}{v_D(0) - v_{D,f}} = \frac{v_{D,h} - v_{D,f}}{v_{D,0} - v_{D,f}}. \tag{6}$$

One can easily observe that, within (4)–(6), we have employed the following auxiliary notations:

$$\begin{aligned}
&\delta_{D,0} = \delta_D(0), \qquad && v_{D,0} = v_D(0),\\
&\delta_{D,h} = \delta_D\left(\tau_{D,f}/2\right), && v_{D,h} = v_D\left(\tau_{D,f}/2\right).
\end{aligned}$$

The technique of customizing the audio fade-out profile by means of rational function (3), which serves as groundwork for the suggested method of shaping the fade-down audio effect, has been validated by a previously advanced implementation in plain ("vanilla") JavaScript (Lupsa-Tataru, 2018). In the present context, it comes to be obvious that implementing the fade-down audio effect by valuating function (1) to generate the fade profile in real-time is structurally similar to implementing a fade-out audio effect that requires the computation of the output of rational function (3). Generically, a JavaScript implementation of the proposed method of fade-down shaping should include the construction given next.

Listing 1. The function designed for audio fading-down.

```
/* global scope: var ae, alphaD, betaD, gammaD;
   var fadeDown = false; */

function setVolD( tDref, tauDf, vDf, rhoD ) {
   var tauD = ae.currentTime - tDref;
   var vD0 = ae.volume;

   if ( fadeDown ) {
      var deltaD = ( tauD - alphaD ) / ( betaD * tauD - gammaD );
      var vD = vDf + deltaD;
      if ( vD > vDf ) { ae.volume = vD; }
      else { ae.volume = vDf; fadeDown = false; }
   }
   else if ( tauD >= 0.0 && vD0 > vDf ) {
      var deltaD0 = vD0 - vDf;
      alphaD = tauDf;
      betaD = ( 2.0 - 1.0 / rhoD ) / deltaD0;
      gammaD = tauDf / deltaD0;
      fadeDown = true;
   }
}
```

Since global variable “ae” of the provided code is created to refer the audio element (object), the invocation of function “setVolD()” will result in an audio volume updating whenever the playback position within the audio content is greater than the (expected) instant $t_{D,ref}$ of fade-down initiation and the output of function (1), denoted within the code by variable “vD”, is greater than the imposed final volume $v_{D,f}$, designated here by means of parameter “vDf” of function “setVolD()”. One perceives that when the value of (1), i.e. the value of local variable “vD”, is found less than or equal to the imposed final level that is the value of parameter “vDf”, the audio volume is set up just to the imposed final level and the fading-down process is stopped.

To avoid unnecessary valuations of (1), the structure encompasses the global variable “fadeDown”, which receives the value of “true” only when the playback position comes to be greater than or equal to the requested instant of fade-down initiation, denoted here by parameter “tDref”, and the detected audio volume, returned by the “volume” property of the audio object “ae”, remains greater than the value of parameter “vDf” that stores the imposed final volume.

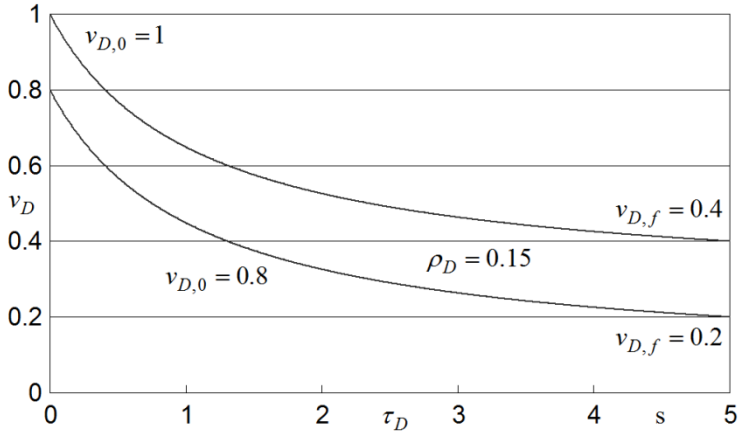

Fig. 1. Fade-down curves for fade length of 5 s, and ratio (5) of 0.15.

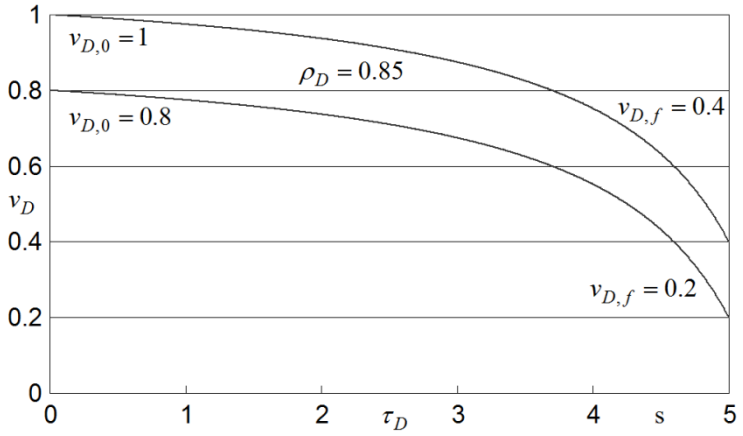

Fig. 2. Fade-down curves for fade length of 5 s, and ratio (5) of 0.85.

Taking into account (1), where function $\delta_D(\tau_D)$ is now provided by relation (3), in Fig. 1 and Fig. 2 we have plotted the fade-down profiles obtained for the fade length of 5 s, the initial levels $v_{D,0} = 0.8$ and $v_{D,0} = 1$, the final levels $v_{D,f} = 0.2$ and $v_{D,f} = 0.4$, and ratio (5) of value $\rho_D = 0.15$ and $\rho_D = 0.85$, respectively. One observes that, regardless of the imposed final level i.e. the value of $v_{D,f}$ in (1), the generated shapes of the fade-down sound effect are crucially decided by the value of quantity (5) that is the ratio between the value of (3) at the fade-down midpoint and the initial value of function (3), occurring at the fade-down initiation.

3. THE AUDIO FADE-UP CUSTOMIZING

Similar to the case of customizing the fade-down shape, we take into account that the evolution of the audio level during fading-up can be represented as the summation of a constant quantity and a function of the playback time. More precisely, we consider here that the function shaping the fade-up profile is brought forth by the summation of the initial audio volume $v_{U,0}$ and a function that technically describes a fade-in audio effect, i.e.

$$v_U(\tau_U) = v_{U,0} + \delta_U(\tau_U), \quad \tau_U \in [0, \tau_{U,f}], \tag{7}$$

where $\tau_{U,f}$ is the length of fade-up audio effect, while the relationship defining the independent variable

$$\tau_U = t_U - t_{U,ref} \tag{8}$$

plainly indicates that, with a view to software implementation, the instant of fade-up initiation has to be subtracted from the current playback time within the audio content (Lupsa-Tataru, 2019).

Since function $\delta_U(\tau_U)$ in (7) has to designate a fade-in audio effect i.e. a strict increasing of the audio level, starting from silence, we straightforwardly employ a rational function that proved to be suitable for real-time implementing. Thus, we deal with the following relation (Lupsa-Tataru, 2019)

$$\delta_U(\tau_U) = \frac{\alpha_U \tau_U^k}{\tau_U + \beta_U}, \quad \tau_U \in [0, \tau_{U,f}], \tag{9}$$

with $k \in \{1,2,3\}$.

To customize the shape of the fade-up effect, we account here for the ratio between the value of (9) at the fade midpoint and the value of (9) at the end of fading-up i.e.

$$\rho_U = \frac{\delta_U(\tau_{U,f}/2)}{\delta_U(\tau_{U,f})} = \frac{\delta_{U,h}}{\delta_{U,f}}; \quad 0 < \rho_U < 1 \tag{10}$$

or, considering relation (7),

$$\rho_U = \frac{v_U(\tau_{U,f}/2) - v_{U,0}}{v_U(\tau_{U,f}) - v_{U,0}} = \frac{v_{U,h} - v_{U,0}}{v_{U,f} - v_{U,0}}. \tag{11}$$

In order for rational function (9) and, implicitly, function (7) to be strictly increasing, the encompassed parameters get the specific expressions in terms of ratio (10) and the initial and final audio levels $v_{U,0}$ and $v_{U,f}$, where $v_{U,0} < v_{U,f}$ (Lupsa-Tataru, 2019):

$$k = \begin{cases} 3, & 1/8 < \rho_U < 1/4\,; \\ 2, & 1/4 < \rho_U < 1/2\,; \\ 1, & 1/2 < \rho_U < 1; \end{cases} \tag{12}$$

$$\alpha_U(\rho_U) = \begin{cases} \frac{4\rho_U}{8\rho_U - 1} \frac{v_{U,f} - v_{U,0}}{\tau_{U,f}^2}, & 1/8 < \rho_U < 1/4\,; \\ \frac{2\rho_U}{4\rho_U - 1} \frac{v_{U,f} - v_{U,0}}{\tau_{U,f}}, & 1/4 < \rho_U < 1/2\,; \\ \frac{\rho_U}{2\rho_U - 1}\left(v_{U,f} - v_{U,0}\right), & 1/2 < \rho_U < 1; \end{cases} \tag{13}$$

$$\beta_U(\rho_U) = \begin{cases} \frac{1 - 4\rho_U}{8\rho_U - 1}\tau_{U,f}, & 1/8 < \rho_U < 1/4\,; \\ \frac{1 - 2\rho_U}{4\rho_U - 1}\tau_{U,f}, & 1/4 < \rho_U < 1/2\,; \\ \frac{1 - \rho_U}{2\rho_U - 1}\tau_{U,f}, & 1/2 < \rho_U < 1. \end{cases} \tag{14}$$

It can be observed that, similar to the case of customizing the fade-down audio effect, we have performed several auxiliary notations i.e.

$$\delta_{U,h} = \delta_U\left(\tau_{U,f}/2\right), \quad v_{U,h} = v_U\left(\tau_{U,f}/2\right),$$
$$\delta_{U,f} = \delta_U\left(\tau_{U,f}\right), \quad v_{U,f} = v_U\left(\tau_{U,f}\right).$$

To facilitate the understanding, in Fig. 3 we have illustrated the fade-up curves received for the fade length of 5 s, the initial audio levels $v_{U,0} = 0.2$ and $v_{U,0} = 0.4$, respectively, the final audio levels $v_{U,f} = 0.8$ and $v_{U,f} = 1$, respectively, and ratio (11) of value $\rho_U = 0.85$.

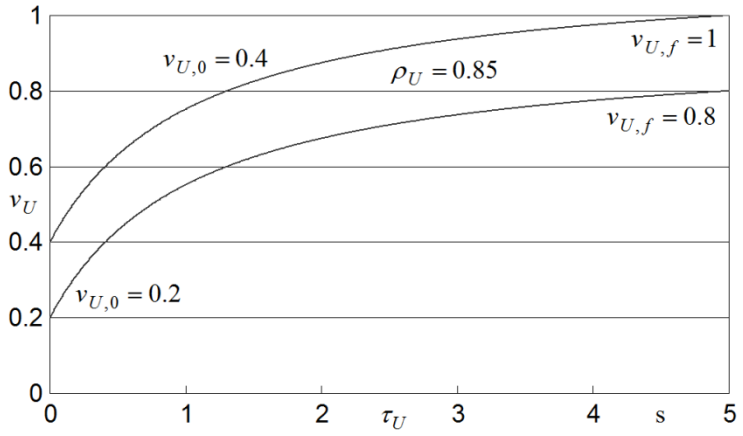

Fig. 3. Fade-up curves for fade length of 5 s, and ratio (11) of 0.85.

Structurally, the implementation of the fade-up audio effect by valuating the output of (7) is analogous to implementing the fade-in audio effect by employing the rational function (9), which proved to be suitable for fast processing in real-time (Lupsa-Tataru, 2019). Nevertheless, in the present case, the audio volume has to be updated each time the playback position of the audio content is greater than the instant $t_{U,ref}$ of fade-up initiation and the value of function (7) is less than the imposed final audio level $v_{U,f}$. As soon as the output of (7) comes to be greater than or equal to the final audio level $v_{U,f}$, the audio volume has to be set precisely to $v_{U,f}$, and the fading-up process has to be stopped in order to avoid subsequent evaluations of function (7).

Obviously, the computation of parameters (12)–(14), expressed now in terms of ratio (11), and the initial and final audio levels, has to be carried out only once that is the first time the playback position is found greater than or equal to the assumed instant of fade-up initiation. For instance, the JavaScript construction given next has been designed for the case of $1/2 < \rho_U < 1$ in representations (12)–(14). Anyhow, one easily perceives that the calling of function "setVolU()" leads to the computation of coefficients (13), (14), denoted by global variables "alphaU" and "betaU", respectively, only if the fading process is not yet started and the playback position, returned by the audio object "currentTime" property, comes to be greater than or equal to the requested instant of fade-up initiation.

Listing 2. The function designed for audio fading-up.

```
/* global scope: var ae, alphaU, betaU;
   var fadeUp = false; */

function setVolU( tUref, tauUf, vUf, rhoU ) {
   var tauU = ae.currentTime - tUref;
   var vU0 = ae.volume;

   if ( fadeUp ) {
      var deltaU = alphaU * tauU / ( tauU + betaU );
      var vU = vU0 + deltaU;
      if ( vU < vUf ) { ae.volume = vU; }
      else { ae.volume = vUf; fadeUp = false; }
   }
   else if ( tauU >= 0.0 && vU0 < vUf ) {
      var deltaUf = vUf - vU0;
      var auxVar = rhoU + rhoU - 1.0;
      alphaU = rhoU / auxVar * deltaUf;
      betaU = ( 1.0 - rhoU ) / auxVar * tauUf;
      fadeUp = true;
   }
}
```

4. CONCLUSIONS

The present investigation emphasizes the feasibility of customizing and implementing in real-time the fade-down and fade-up audio effects, having at hand techniques of shaping the fade-out and fade-in audio effects, which have been verified for the suitability with real-time computing (Lupsa-Tataru, 2018, 2019).

The audio fades customization is carried out here by taking into account that the evolution of the audio volume during a fading process can straightforwardly be described by the output of a function of the type:

$$v(\tau) = V + \delta(\tau) \tag{15}$$

where: τ – the difference of the playback time and the instant of fade initiation,
V – constant term,
$\delta(\tau)$ – rational function that depicts a fade-out or a fade-in audio effect.

It is pointed out that, by employing a relation of type (15) in order to shape the fade profile with the purpose of real-time processing, the implementation of the fade-down audio effect is similar to implementing the fade-out audio effect whilst the implementation of the fade-up audio effect becomes analogous to implementing the fade-in audio effect. In this context, the essential tasks required by the appropriate real-time implementations are highlighted in the course of presentation.

REFERENCES

Case, A. U. (2007). *Sound FX: Unlocking the Creative Potential of Recording Studio Effects.* Burlington, MA, USA: Focal Press.

Jackson, W. (2015). *Digital Audio Editing Fundamentals: Get started with digital audio development and distribution*. Berkeley, CA, USA: Apress Media. doi:10.1007/978-1-4842-1648-4

Langford, S. (2014). *Digital Audio Editing: Correcting and Enhancing Audio in Pro Tools, Logic Pro, Cubase, and Studio One.* Burlington, MA, USA: Focal Press.

Lupsa-Tataru, L. (2018). Novel technique of customizing the audio fade-out shape. *Applied Computer Science*, *14*(3), 5–14. doi:10.23743/acs-2018-17

Lupsa-Tataru, L. (2019). Implementing the fade-in audio effect for real-time computing. *Applied Computer Science*, *15*(2), 5–18. doi:10.23743/acs-2019-09

Reiss, J. D., & McPherson, A. (2015). *Audio Effects: Theory, Implementation and Application.* Boca Raton, FL, USA: CRC Press.

Schroder, C. (2011). *The Book of Audacity: Record, Edit, Mix, and Master with the Free Audio Editor*. San Francisco, CA, USA: No Starch Press.

Applied Computer Science, vol. 15, no. 4, pp. 37–48
doi:10.23743/acs-2019-28

Submitted: 2019-10-24
Revised: 2019-11-25
Accepted: 2019-12-05

Information Systems IS, Information Systems Technology IST,
Integrated Management Information Systems IMIS, Enterprise Resource Planning ERP

Bartosz CIEŚLA[*]*, Grzegorz GUNIA*[**]

DEVELOPMENT OF INTEGRATED MANAGEMENT INFORMATION SYSTEMS IN THE CONTEXT OF INDUSTRY 4.0

Abstract

In this paper, development trends of information systems, information systems technology and enterprise information management were analyzed in the context of Industry 4.0 tools. In the first part (par. 1–2), fundamental definitions referred to the subject were presented as well as historic background of Integrated Management Information Systems. In the second part (par. 3), evolution and trends in ERP class systems, electronic economy tools and Product Lifecycle Management software were described. In the third part (par. 4–5), observed trends in information systems technology, in relation to Industry 4.0 tools, were discussed including manufacturing resources, production objects and novel management strategies approach. Many conclusions were related with actual manufacturing practices observed by the authors.

1. ENTERPISES INFORMATION SYSTEM TECHNOLOGY

1.1. Enterprises Information System (EIS)

For enabling an effective information flow between source and user as well as its efficient utilization, system for collecting, storing and transferring information internally and externally should be designed. It is called Information System (IS).

[*] Redor Ltd., Grażyńskiego Street 2, 43-300 Bielsko-Biała, Poland, b.ciesla@hotmail.com
[**] University of Bielsko-Biała, Faculty of Mechanical Engineering and Computer Sciences, Production Engineering Department, Willowa Street 2, 43-309 Bielsko-Biała, Poland, ggunia@ath.bielsko.pl

IS can be defined as multilevel structure, which is used for transforming certain input information/data into desired output information, by means of dedicated procedures and models. On the basis of this output information, particular decisions are made (Kisielnicki & Sroka, 2005).

1.2. Enterprises Information System Technology (EIST)

Information System Technology (IST) is a separate part of IS, which is computerized for achieving relevant, particular objectives, so it can be defined as information system based on information technology (Kisielnicki & Sroka, 2005). In the other words, it is a set of interrelated elements, designed for transforming data and implementing communication flow with application of computer technology.

Enterprises Information System Technology is a part of EIS which is responsible for generating and collecting source data, as well as transforming, analyzing and visualizing this data. It is done with application of methods, techniques and tools of Information (Computer) Technology (Januszewski, 2001).

Essential components of EIST are: equipment (workstations, servers, ICT infrastructure, …), software (operating systems OS, application software App, database systems, …), human resources (system administrators and users), organization (procedures, rules, EIST instructions and laws).

Integrated Management Information Systems are modularly organized EISTs, operating all enterprises activities: marketing, resources planning and supplies (logistic), technical preparation of production, manufacturing processes managing, distribution, sales, maintenance, finance and human resources (Adamczewski, 2014).

From the point of view of technical solutions, Integrated Management Information System, is a system in which (Lech, 2003), (Januszewski, 2008):

- user with his own workstation, is able to use any system function,
- all users of the same system, are working on the same system interface,
- input of data into the system is done only once, and this automatically upgrades state of the system as well is visible for all system users.

2. EVOLUTION OF INTEGRATED MANAGEMENT INFORMATION SYSTEMS

Development of ERP (Enterprise Resource Planning) class systems has begun in the 60's. Its first version was MRP I (Material Requirements Planning). This system was designed for calculating precise amount of materials and components needs in time, so effective supplies schedule for fluctuating demands of various products was enabled to create. Working principle of this system is to transform production plan to components and materials requirements with precision amount of it and access to information about requirement date, for every production batch.

Upgrade of MRP I model was involving closing information loop (Closed Loop MRP). With this technique reaction for fluctuating demands and changing production parameters (especially authentic supply and products amounts) in real time was enabled.

In case of this model, despite of planning aspect (material and production resources, production processes, internal orders, production scheduling), for the first time ever, quality control aspect was considered (for production processes, internal orders execution, turnover and deployment of: materials, components and products in warehouses). Information given by controlling elements of model was next returned to central element of material requirements planning (and intermediately to production capacity planning element). Authentic materials, components and products balance are then available for system users, so better plan restructuring decision process could be done. This return information flow was closing information loop of MRP model. Closed Loop MRP was also improved on the previous version MRP I with functionality of planning necessary resources other tan components and materials (workforce, machines, equipment, tools, transportation etc.).

Next model development was MRP II called Manufacturing Resource Planning. In this model, Closed Loop concept was fully adopted, as well as new elements connected with sales processes and more strategic than operating (as it was before) decision-making processes like marketing planning, strategic planning or finance planning.

MRP II standard was defined and published in year 1989 by The Association for Operations Managements APICS (historically American Production and Inventory Control Society).

Besides controlling materials and components requirements as well as its stock and orders level, MRP II was also designed for workforce, machines and equipment planning. It is done by allocating subsequent jobs to manufacturing resources according with designed production process of single parts and final products (production scheduling). Moreover, throughput level for each manufacturing resource was shown, creating new controlling tool for managers.

Finally, as an upgrade of MRP II, Enterprise Resource Planning ERP model was developed. Sometimes it is also called MRP III – Money Resource Planning. Term ERP was proposed for the first time in year 1990 by analytical consulting corporation Gartner Group. These systems are adopting MRP II model and procedures, and further developing them with solutions supporting decision making processes, using enterprises know-how, along with artificial intelligence concept. Major task for ERP systems is to fully integrate all enterprises business activities: production, marketing, finance, logistics, strategic management, etc. Moreover, with mechanisms available in ERP systems, simulation of different business scenarios is given. These scenarios can be further analyzed from various points of view, also financial one (Gunia, 2010), (Gartner, 2004).

3. TRENDS IN THE DEVELOPMENT OF MANAGEMENT INFORMATION SYSTEMS

3.1. ERP II class systems

ERP II class systems are the next development stage of integrated MRP/ERP systems after level of ERP systems. Modern trends in enterprises structure evolution, connected with vertical integration and optimization-focused internal functions and processes, are development basis for systems class ERP II. Moreover, organizations endeavor to preserve flexibility of their core activities, and optimal positioning themselves, along the supply chain and value network. Fundamental aspect of this positioning, is not only participating in trading using computer networks (e-commerce, B2B, B2C), but rather commitment in c-commerce processes. C-commerce (collaborative commerce) means supporting electronic business interactions between enterprise personnel, clients and business partners in the context of one trading society. This society could be any industry, or one of their segments, as well as supply chain or even a part of a supply chain. In world of global collaboration, enterprises have to compete with each other's not only in the field of availability, quality or price of their products and services, but also in field of information. Especially speed of information and quality of information supported to co-operating partners is crucial.

Gartner Group has defined ERP II as a business strategy and collection of specific program applications for particular sectors, which generates value for clients and shareholders. It is done by mean of availability and optimization of both: internal processes and processes between co-operating companies (Genovese, Bond, Zrimsek & Frey, 2001), (Gartner, 2004) .

ERP II systems are orientated on external integration and developing solutions along with business partners, in contrast to classic ERP systems – orientated on internal business processes. Optimization of resources and process data processing, are in ERP II supported with extra information about resources involvement in companies' efforts, aiming extending cooperation with others enterprises. In this area, traditional ERP systems allows only for managing purchases and sales by means of e-commerce. Last but not least, internet-based, integrated architectures of ERP II products, are such different than monolithic ERP architectures, that they require overall transformation. Data treatment involves not as before collecting data inside enterprise, but managing data distributed along with trading society.

During its evolution, ERP II model absorbed functionality of SCM (Supply Chain Management), which was used for managing business partners supply chain as well as exchanging information within. In ERP II concept, functions of ERP /MRP II systems like production planning, logistic and inventories management, finance management, etc. were complemented with electronic offers, orders and invoices exchange, likewise electronic payments. In the close future,

e-procurement, automation in enterprise office materials and consumables supply, will be also absorbed and integrated in ERP II (Genovese at al., 2001), (Rzewuski, 2002).

In conclusion, it can be said that ERP II class systems are much stronger focused on reflecting market behavior and mutual relationships between cooperating companies, than on supporting and handling enterprises internal business processes. Evolution of Integrated Management Information Systems in the context of industrial revolutions timeline is presented on figure 1.

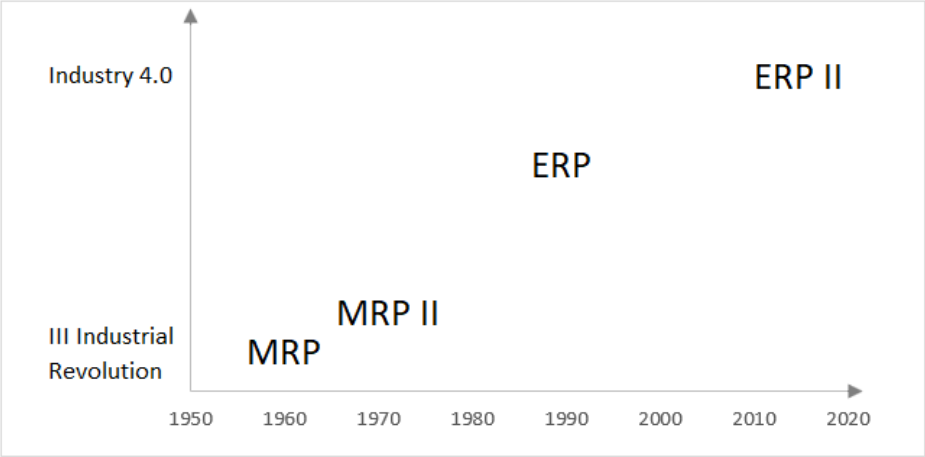

Fig. 1. Evolution of Integrated Management Information Systems

3.2. Electronic economy

With the large market competition and the fierce battle for the customer, enterprises have to seek new possibilities for promoting and advertising products and services, as well as fresh distribution channels and effective communication channels with business partners.

Expansion of worldwide Internet network, allowed for companies arisen, irrespective of their size, in the new digital reality, termed as "New Economy", where e-commerce, e-business, e-economy, etc. were developed (Gregor & Stawiszyński, 2002), (Kolbusz, Olejniczak, & Szyjewski, 2005). There is opinion in science literature, that development of informatic solutions, enabled new direction for economy development, but in the other hand, economy development forced new informatic tools emerging.

Term e-business is defined on many ways. On strategic level, e-business is understood as idea of reconstructing whole enterprise, for maximizing benefits from modern technologies utilization.

On lower management levels, e-business is understood as network involving issues, considering purchase and sales of products and services. Significant feature of e-business is that transactions between two sides are executed on-line, and main object of this trading is information.

E-business could be generally divided into three categories, differing with objectives and target group (Kolbusz at al, 2005): B2C (Business-to-Customer), B2B (Business-to-Business), B2P (Business-to-Public).

Business-to-Customer

B2C is designed for executing transactions between enterprises and customers (consumers). It is more likely organized as electronic trading (e-commerce, e-tailing) performed through online shops.

Fundamental tasks of B2C systems:

- enabling purchasing on-line,
- supporting supply processes,
- providing after-sales services,
- improving distribution,
- lowering transaction costs.

Electronic trading creates competitive environment, in which smaller enterprises are able to compete with giants. Geographical barriers are broken through this global environment, and nearly unlimited choice possibilities for customers are provided.

Business-to-Business

Model B2B is designed for transactions executing between companies. Analysts are expecting this category to have highest turnovers and profits, eventually even 90% of market income. Nowadays, US market is developing constantly in this direction, and the same could be observed through European Union countries. Integration supply processes named as Supply Chain Management (SCM), is considered to be essential element of B2B solutions. At this field, with properly utilized modern technologies, the best conceivable measurement results for enterprises, like cost reduction or logistics improvement can be obtained. Proper explored opportunities could be required to ensure competitive position preservation.

B2B development trends are driving enterprises effectiveness and efficiency improvement through increasing business processes integration. This results in automated data processing and again costs reduction.

Business-to-Public

At this field of e-business, relations between enterprises and their macro environment (especially social environment) are main objectives.

Fundamental tasks of B2P systems are:

- creating the company's image, profile and brand (not only on the Internet and in social media),
- brand and product promotion,
- creating social ties and improving links between company and its environment,
- last but not least attracting new customers.

Alongside with abovementioned, in the literature one can find spectrum of more specific models e.g. (Simon & Shaffer, 2002), (Koźmiński & Piotrowski, 2002): C2C (Customer-to-Customer), C2B (Customer-to-Business), B2E (Business-to-Employee), G2C (Government-to-Citizen), B2G (Business-to-Government). Nowadays electronic economy tools are linking nearly every possible entity as shown on Figure 2.

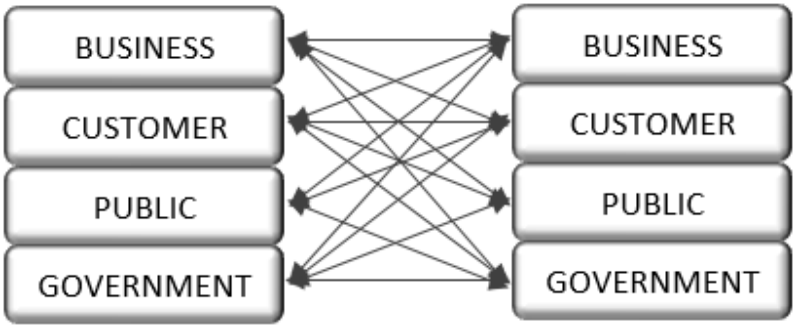

Fig. 2. Modern information flow with electronic economy tools

3.3. Product Lifecycle Management

Set of processes focused on issues related with product, from birth concept, through designing, manufacturing and selling, until aftersales service and disposal, are called Product Lifecycle Management (PLM). It is responsible for integrating key factors in the creation of product: data, resources and designing process. PLM is designed for enabling access to all product information, making them easier to manage. It is very important tool for Product Managers and is applied in many businesses.

PLM could be considered in two aspects:

1. As a concept, for supporting product lifecycle tasks, in a holistic manner, at any stage of its maturity, through:
   - procedures standardization,
   - collection, integration, and sharing data and knowledge about product,
   - automation and support of whole sets of functions and tasks,

- electronic data and documents exchange,
- automation of technical documentation workflow and process management,
- integration of the entities involved in various product lifecycle stages (traders, clients, designers, production processes organizers etc.).

2. As a tool – an informatic system (but more precise a collection of integrated applications). Its main objective is to integrate applications from the various action fields (supporting particular stages of product lifecycle) into well organized and coordinated, efficiently functioning entirety. This tool involves many systems as:
   - Computer-aiding – CAx,
   - Product Data Management – PDM,
   - Document Data Management – DDM,
   - modelling and simulation of production systems,
   - systems for supporting cost calculation,
   - project management systems,
   - etc.

4. INFORMATION SYSTEM TECHNOLOGY IN RELATION TO INDUSTRY 4.0

Definition of Industry 4.0, understood as fourth industrial revolution has very broad sense. It could be described as human, machines and digitally controlled devices network creation and systems integration. It is done by means of widely used Internet and Information System Technology. The rise of Industry 4.0 was made possible through widespread access to computer equipment in various forms. Equipment with very high computing power, where single units are connected with each other and are able to communicate in real time through global network. This idea is focused on information, its availability as well as sharing at any time and at any place.

Approximately at the same time as ERP model was defined, the most characteristic invention of Industry 4.0 – Internet was invented. It was one of main activators for vast electronic devices development and global data digitalization. Integrated Management Systems have benefit from this advancement, especially in the area of information flow and data collection. Progressive availability of informatic infrastructure was used by companies for building interior physical networks for the purpose of ERP software data feeds. From this moment, every single event in the enterprise operation could be recorded using reasonable workload. Later on, collected data is transformed into information links between company's personnel. Basically, Customer Service is informed about orders status and payments, Logistics is informed about needs and deliveries, Production

(orders executors) are informed about demands and plans, and last but not least Management is informed about expenses, costs, sales and profits, all of the above nearly in the real time and simultaneously. In this context, Industry 4.0 trends may be understood as sophistication in data collection and its transformation into information as well as sophistication in information display.

4.1. IST products development

From many areas of Industry 4.0, producers of Integrated Management Information Systems are mainly interested in their product development connected with: IT and automatics integration, Industrial Internet of Things (IIoT), Augmented Reality and cyber-security in industry. But manufacturing companies in developed countries have their own considerable demands. They endeavor to transform into integrated networks in which, with usage of Agile Manufacturing and Mass Customization strategies, their core competences will be united (Brettel, Friederichsen, Keller &Rosenberg, 2014). Purpose for this ongoing transformation lies in emerging competitive advantage resulting from customized products with fast time to market. This trend must be considered by IST producers as these corporations from developed countries constitute a large share of IST market. Moreover, and even more importantly for IST producers, no other enterprises can provide bigger scale funding of new technology applications. Considering all of the above, it is suspected that IST producers will adjust their product mix development strategies to the highlighted market needs. On the other side, there is virtually countless number of Small and Medium Sized Enterprises (SMEs) and less developed bigger companies waiting for budget, simple to implement, novel tools for their stage of advancement, which is mostly between 3rd Industrial Revolution and Industry 4.0. Observations of manufacturing practice have shown that SMEs believes in existence of universal, verified devices and software, with a touch of Industry 4.0 artificial intelligence, which can cure most of their operational difficulties. Obviously, nobody can clearly define what could it be and how would it work. But it should be considered, that maybe everyone in this market – producers and users are missing something crucial in IST development. Arguably it won't be a cure for everything but it could be a breakthrough for SMEs agile at operational level. It is well-known that SMEs are more effective at operational level than Corporations since their compact structure, but problem resides in chaotic execution based on aptitude of individuals. Novel advisory tools in ERP systems could stabilize decision making process at SMEs by means of Big Data and neural networks elements, all based on automatically collected data by IST networks.

When it comes to production supervision, development of IST is focused on integration of all processes in enterprise, executed with application of Internet and mobile technologies. Again, scope of processes is very wide, embracing customer service (quotation, order approval and order confirmation) through

manufacturing planning, throughput level and machines availability, Supply Chain and Logistics, ending with sales and distribution. Corporations are achieving significant progress in this area by thrusting particular software upon their smaller suppliers. For example, it is done with SAP in automotive business involving German enterprises. In many cases these suppliers are treated more favorably during SAP implementation process, receiving greater conceptual support, so it became a win-win situation. Corporation is executing orders, invoices and other necessary information flow with suppliers by means of mobile technology developed by SAP, and suppliers are acquiring best known operational techniques for their internal development (SAP, 2019).

4.2. Challenges for IST producers

First main area of challenges for IST producers are issues involving linking physical world of manufacturing resources with informatic systems or virtual either augmented reality. In context of Integrated Management Information Systems, studies have shown that there is no unified standard and protocol for machine-ERP communication but simultaneously ERP systems are considered technologically and operationally ready for Industry 4.0 (Gunia, 2010). Moreover, producers are obliged to keep up with development and autonomation of manufacturing resources, exacting interfaces improvement for integrated information flow between machines software and production planning systems. In this regard new open standards must be worked out, adjusting communication channels between manufacturing resources, software for production planning and data collection, as well as information exchange between cooperative enterprises. Nowadays IST designed for autonomous data collection from machines and workstations are formations separated from ERP and analysis of given information are hampered in context of production results. Again, scope of problem is different depending on enterprise characteristics. Corporations have their dedicated applications and controlling departments at this area, so analytics issue is not as problematic for them as for SMEs. With SMEs limitation in specialized human resources, their require information flow simple to processed and interpreted, with excluded unnecessary informatic infrastructure to purchase and maintain. In this context, important challenge for IST producers is in implementing visual process analysis and information display in their core products. It can be defined as fusion of Business Intelligence Systems and Management Information systems. Through this development direction, ERP systems could upgrade into flexible structures, adjusted to various unique management models as Theory of Constraints or Activity Based Costing (Cieśla & Kolny. 2019).

In theme of Industrial Internet of Things, two main areas can be highlighted to challenge with: manufacturing resources (machines, equipment, work stations) and production objects (materials, components, parts and products).

First area is connected with direct, both ways communication between manufacturing resources and operating personnel as well as data processing and collecting for major machines issues: maintenance, throughput efficiency and processes condition. The results of this data analysis should be used in the real time, for adjusting production plan, revealing causes of fluctuations of productivity indicators, revealing causes of equipment breakdowns. Very clear, and visual feedback should be given to all users. In this area support for Predictive Maintenance should be given as well. Predictive Maintenance is a strategy for Maintenance Departments, aiming optimal use of machinery and equipment, by eliminating possible breakdowns or performing maintenance tasks on basis of received technical condition data.

Field of production objects is developing in direction of current and automatic quality control of products, with real time feedback. Through this advancement, quality control process should become less expensive and less time consuming than nowadays. It can also enable economically reasonable defects detection after each production step. The effect would be in higher production automation as well as in higher efficiency of material requirement planning and warehouse documents flow connected with defects. Second research direction in field of production objects, connected with Factory of the Future concept, is economically rational technology of production autonomous actuator-sensor networks in the factories. The purpose for these networks is in knowing products history and routes so logistic chain is simplified and autonomously managed and controlled (Zuelhke, 2010). All of the above is mainly done with usage of: Radio-frequency Identification Technology (RFID), image recognition, image analysis and augmented reality.

5. CONCLUSIONS

Integrated Management Information Systems are nowadays on high development level. Through their five decades of evolution very compact and multifunctional structures were established. All of the above was done along with emerging markets needs and economic progress. In Industry 4.0 era, new circumstances and requirements appeared, related with instruments like (Industrial) Internet of Things, Big Date, Business Intelligence Systems etc. Customers, goods and manufacturing becomes global and widely available for instance with application of Electronic Economy tools. Companies producing EIST have opportunity to achieve new competitive advantage by developing product mix framework appropriate for present market 4.0 conditions. Authors believes that it could be done only through close collaboration with particular markets innovative representatives. Like in Lean Gemba Walk the best improvement ideas could come up from real work observation.

REFERENCES

Adamczewski, P. (2014). *Zintegrowane systemy informatyczne w praktyce*. ZNI MIKOM.

Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M. (2014). *How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective.* International Journal of Mechanical, Aerospace, Industrial and Mechatronics Engineering 8(1): 37–44.

Ciesla, B., Kolny, D., (2019). *Visual process analysis in SMEs as a support for management models on example of TOC.* Journal of Systems Integration 2: 19–27.

(2004) *The Gartner Glossary of Information Technology Acronyms and Terms*. Gartner.

Genovese, Y., Bond, B., Zrimsek, B., & Frey, N. (2001). *The transition to ERP II: Meeting the challenges*. Gartner Research. Strategic Analysis Report No. R–14–0612.

Gregor, B., & Stawiszyński, M. (2002). *e-Commerce*. Branta.

Gunia, G. (2010). *Zintegrowane systemy informatyczne zarządzania w praktyce produkcyjnej*. Wydawnictwo Fundacji Centrum Nowych Technologii, Bielsko-Biała.

Haddara, M., Elragal, A. (2015). *The Readiness of ERP Systems for the Factory of the Future.* Procedia Computer Science 64: 724–728.

Januszewski, A. (2001). *Informatyka w przedsiębiorstwie: systemy i proces informatyzacji*. WSZiF.

Januszewski, A. (2008). *Funkcjonalność informatycznych systemów zarządzania: Zintegrowane systemy transakcyjne*. Wydawnictwo Naukowe PWN.

Kisielnicki, J., & Sroka, H. (2005). *Systemy informacyjne biznesu, metody projektowania i wdrażania systemów*. Agencja Wydawnicza Placet.

Kolbusz, E., Olejniczak, W., & Szyjewski, Z. (Eds.). (2005). *Inżynieria systemów informatycznych w e-gospodarce*: praca zbiorowa. Polskie Wydaw. Ekonomiczne.

Koźmiński, A. K., & Piotrowski, W. Z. (2002). *Zarządzanie. Teoria i praktyka*. Wydawnictwo Naukowe PWN. Warszawa.

Lech, P. (2003). Zintegrowane *systemy zarządzania ERP/ERP II. Wykorzystanie w biznesie, wdrażanie*. Difin, Warszawa.

Rzewuski M. (2002): *ERP II – nowy stary gatunek.* Pckurier 20/2002.

Simon, A. R., & Shaffer, S. L. (2002). *Hurtownie danych i systemy informacji gospodarczej: zastosowanie w handlu elektronicznym*. Oficyna Ekonomiczna.

Zuelhke, D., (2010). *SmartFactory – Towards a factory-of-things.* Annual Reviews in Control 34: 129–138.

www.sap.com/industries/automotive.html (29.08.2019)

Applied Computer Science, vol. 15, no. 4, pp. 49–62
doi:10.23743/acs-2019-29

Submitted: 2019-09-21
Revised: 2019-11-25
Accepted: 2019-12-06

herb extract, phenolic compounds, antioxidant capacity, VisionPro™ software

Karolina FERYSIUK[*], *Karolina M. WÓJCIAK*[0000-0002-4202-433X]*

THE SPECTROPHOTOMETRIC ANALYSIS OF ANTIOXIDANT PROPERTIES OF SELECTED HERBS IN VISION-PRO™ UV-VIS

Abstract

The aim of the study was to evaluate the influence of type of the solvent (water, aqueous ethanol and ethanol) on the antioxidant properties of four various herbs: couch grass (A. repens), milk thistle (S. marianum), dandelion (T. officinale) and fireweed (E. angustifolium) measurement by three common UV-VIS methods (TPC, ABTS^{+}, DPPH). The results were collected through the Vision-Pro™ UV-VIS spectrophotometer software. Aqueous ethanol was the most effective solvent for extraction for all type of herbs. Fireweed contains the highest amount of polyphenol compounds (0.625 μg GA/ml). The lowest antioxidant capacity was presented by extracts from couch grass (0.019 μg GA/ml).

1. INTRODUCTION

Meat products are exposed to the oxidation process, which is responsible, among others, for deterioration of nutritional value, shortening of shelf-life and creation of off-flavors. To avoid those processes, a synthetic antioxidant (e.g. sodium nitrite) is added to meat products. Unfortunately, generally speaking, consumers associate food chemical additives as a negative factor for their health – while the naturalness of the products is linked with a positive effect on health (Rodríguez-Rojo, Visentin, Maestri & Cocero, 2012; Hung, de Kok & Verbeke, 2016). For that reason, alternatives for typical food additives are needed. One of the substances that could partially or even fully replace synthetic additives is phenolic compound.

[*] University of Life Sciences in Lublin, Faculty of Food Science and Biotechnology, Department of Animals Raw Materials Technology, Skromna 8 Street, 20-704 Lublin, Poland, karolina.ferysiuk@student.up.edu.pl, karolina.wojciak@up.lublin.pl

Phenolic compounds (polyphenols) are secondary plant metabolites; depending on their chemical structure, several classes are highlighted: phenolic acids, tannins, flavonoids, lignans and stilbenes. Polyphenols show good antioxidant activity and can be used as a food preservative. The relations between carboxyl group and numbers and positions of the –OH are factors which determine their antioxidant activity. Moreover, secondary plant metabolites also presented a positive influence on human health by reducing the risk of some pathological disturbances (e.g. reduce the incidence of coronary diseases, present anti-atherosclerotic and anti-carcinogenic effects). They also present antimicrobial properties and can inhibit the growth of pathogenic microorganisms (Falowo, Fayemi & Muchenje, 2014; Ignat, Volf & Popa, 2011; Oroian & Escriche, 2015; Martillanes, Rocha-Pimienta, Cabrera-Bañegil, Martín-Vertedor & Delgado-Adámez, 2017; Wendakoon, Calderon & Gagnon, 2012). As Al-Snafi (2015) pointed out, secondary metabolites found in the plant have a therapeutic and pharmacological effect. Herbs and spices are a rich source of phenolic compounds and they are often applied to the food products – except for the prolonging of the durability of the food they also carry flavor. Antioxidant substances applicated in food industry must be effective at low concentrations, inexpensive, highly stable, non-toxic, colorless, tasteless and odourless. In order to avoid the too intense taste of plant additives in the product, the extraction process is used (Hinneburg, Dorman & Hiltunen, 2006; Shahidi & Ambigaipalan, 2015). But this is not the only reason – in general, this process is considered as one of the best sustainable methods of the biological components extraction (Gupta, Naraniwal & Kothari, 2012). In general, extraction is a process in which the main purpose is the isolation and separation of the specific components through the application of an appropriate, adequate solvent. It is important to remember there is no one, standard method of extraction (Ignat et al., 2011). Various extraction solvents can be used: water, ethanol, methanol, acetone and their mixtures with water (Wendakoon, Calderon & Gagnon, 2012). The choice of reagent depends on, among others, chemical nature of polyphenols and their solubility in the solvent (e.g. methanol or even acidified methanol is usually applied for anthocyanin extractions) (Naczka & Shahidi, 2004). Therefore, compared with other solvents, water and ethanol are recommended as an extraction solvents for food industry due to their safety for human consumption (Wendakoon et al., 2012; Ignat et al., 2011).

As it was mentioned earlier, the lipid oxidation is a very negative process for food, especially for meat products. The oxidation process consists of the following stages: initiation (the lipid-free radical occurs), propagation and termination (occurs of non-radical products). Antioxidant substances can stop this process through the scavenging of initial radicals, braking chain reactions, intercepting singlet oxygen or decreasing concentration of oxygen. Phenols present strong antioxidant activity and should be added to food products in low concentrations; in higher concentrations they can lose their activity and become prooxidants (Shahidi & Ambigaipalan, 2015).

The aim of this article is to compare the influence of the three various solvents on the antioxidant properties of herb extracts from dandelion, couch grass, milk thistle and fireweed, extracted at 40°C by using a different UV-VIS antioxidant measurement methods. The data were collected through the Vision-Pro™ UV-Vis spectrophotometer software version 2.03.

2. MATERIAL AND METHODS

2.1. Material

Plant material was dry, shredded herbs: milk thistle (*S. marianum*), dandelion (*T. officinale*), couch grass (*A. repens*) and fireweed (*E. angustifolium*). Three variants of the extracts were created: aqueous, ethanolic and their mixture (50:50). 30 ml of fresh solvent was added to herbs (5 g) and shaken at 150 rpm for 3 hours at 40°C. Solvent was then changed every hour. The extract was filtered through filter paper Whatman No 1 and prepared for analysis.

2.2. Methods

2.2.1. Antioxidant activity

The analyzes of antioxidant properties included total phenolic content (TPC) which determined by a modified Folin–Ciocalteau method, described by Singleton and Rossi (1965) using a Folin–Ciocalteu reagent. Samples were measured after 30 min of storage at room temperature, in the dark. The TPC values were calculated from a standard curve of gallic acid equivalent and expressed as mg GA/ml. The absorbance was measured by using a UV-VIS spectrophotometer (Nicolet Evolution 300).

2.2.2. Radical scavenging activity

The radical scavenging activity (DPPH and $ABTS^+$ methods) was measured according to Jung et al. (2010) with some modifications. For both methods absorbance was measured after 3 min. Volume of $ABTS^+$ reagent was reduced to 1.8 ml, volume of extracted sample was reduced to 12 µl. The absorbance was measured by using a UV-VIS spectrophotometer (Nicolet Evolution 300) at 734 nm for the ABTS assay and at 517 nm for DPPH assay.

2.3. Statistical analysis

The results were statistically analyzed using KyPlot statistical program and presented as mean±standard deviations using a T-Tukey's range test. The three research series were created and samples were measured in duplicate.

3. THE ANTIOXIDANT MEASUREMENT METHODS

How was mentioned earlier, the polyphenols compounds occurring in plant materials, are considered to be a substances of antioxidant properties. For those reason they can be applied as additives to food products and therefore extend their durability. It should be keep in mind that the antioxidant capacity of plants can be various and this determine their application as food additives. For the measurement of antioxidant potential various methods can be applied, but the most common are: TPC, DPPH and TEAC/$ABTS^+$. All those assays are based on the reaction between substances of antioxidant properties and a special reagent. The common features connecting these methods are simplicity, inexpensive, reproducibility and no need for specialized equipment except ultra violet visible (UV–VIS) spectrophotometer (Karadag, Ozcelik & Saner, 2009; Moniruzzaman, Khalil, Sulaiman & Gan, 2012; Shahidi & Zhong, 2015). UV–VIS spectrometry is considered as a sensitive, fast, environment friendly and simple method for antioxidant potential measurement (Biswas, Sahoo & Chatli, 2011; Yu, Wang, Zhan & Huang, 2018). How explain Yu et al. (2018) in spectroscopic techniques the concentration of the chemical component is predicted through the calibrating a predicting models which, through the proper chemometrics, correlates collected spectral data and reference values of chemical concentrations. The wavelength range for the UV–Vis method is between 200 and 780 nm which corresponded to the X–rays and NIR (near-infrared) range. The spectrophotometer apparatus consist of a proper, dedicated optical spectrometer (light source, detector, sample compartment, monochromator) and a control unit (PC).

In general speaking spectrophotometer methods for antioxidant measurements are based on the measurement of changing the absorbance spectrum of the tested sample against the blank sample (Moniruzzaman, Khalil, Sulaiman & Gan, 2012; Wojdyło, Oszmiański & Czemerys, 2007). Depending on the method, various wavelength are chosen: DPPH reagent have a characteristic, strong absorbance at 515–517 nm (Wojdyło, Oszmiański & Czemerys, 2007; Moon & Shibamoto, 2009) however other authors also pointed out that the reducing ability can be also measured at 518, 520 to 528 nm (Kedare & Singh, 2011; Karadag, Ozcelik & Saner, 2009). Moreover, for the $ABTS^+$ radical scavenging activity measurement, Karadag et. al. (2009) divided wavelength in two groups, depending on the solvent type: for aqueous: 414, 752, 842 nm, for ethanolic: 414, 730 and 873 nm.

The data from measurement are collected through the proper software, which can be built into the device (a computer unit is not needed) or installed separately on the computer. The recommended progremm for the Nicolet Evolution 300 spectrophotometer is the Vision-Pro™ program (Vision-Pro Thermo Electron UV-Visible Spectrometry, version 2.03; Math Version 24.00). The main view of Vision-Pro™ program is presented on Fig. 1. On the toolbar are presented a standards options: file, application, command data store etc. Before the measurement, the user sets the wavelength, performs zeroing of the device and prepares the sample.

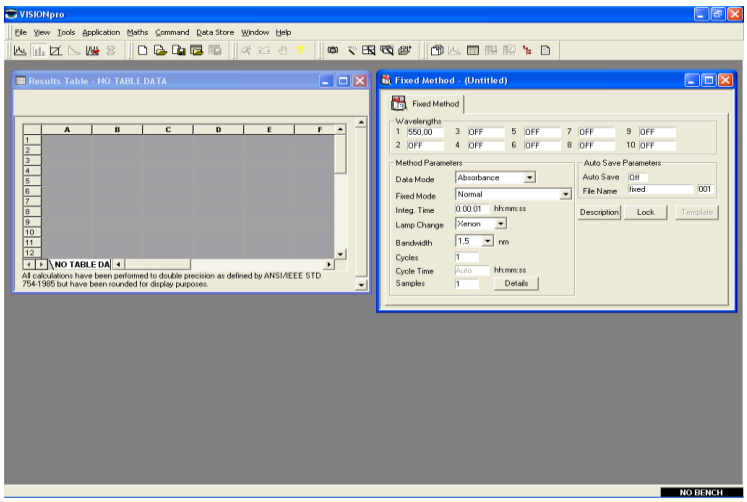

Fig. 1. Main menu of Vision-Pro™

This program allows to perform many different types of measurements e. g. sample measurement in selected cycles and time intervals (method applied e. g. for radical scavenging assays for the Inhibition Concentration IC_{50} determination) or for the calibration graphs drawing (Fig. 2).

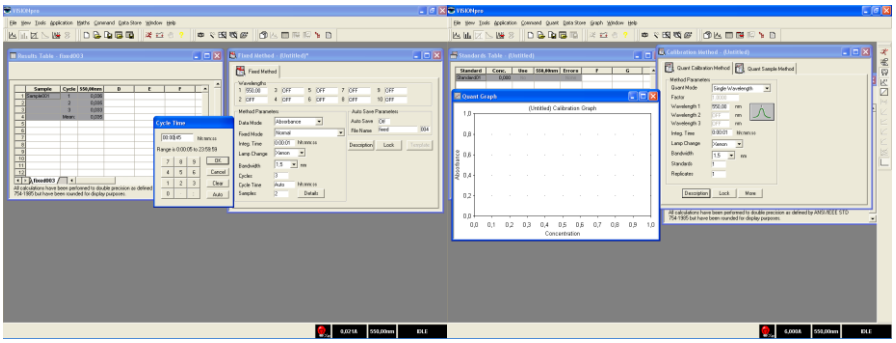

Fig. 2. Various analyses from program

The effect of the carried out measurements are a certain absorbance values. To the standardize the results the absorbance values are recalculated through the proper mathematic formula or equations from the standard calibration curve. For example, the results of DPPH and $ABTS^+$ measurements can be recalculated and presented as a percentage radical scavenging activity according to the formula e.g. for DPPH:

$$\% = \frac{A_{control} - A_{sample}}{A_{control}} \cdot 100 \quad (1)$$

where: *A* – absorbance.

Or can be show as an Inhibition Concentration (IC_{50} or EC_{50}) – concentration of the antioxidant substances which can inhibit of free radical by 50% (Moniruzzaman, Khalil & Sulaiman, 2012; Kedare & Singh, 2011). Although most scientists prefer the presents results as an equivalent of amount of selected antioxidant – standard – per volume (μL or mL). A calibration curve is constructed by reference values of measurement of the absorbance of selected, diluted standard (Trolox, gallic acid) to the standard's concentration. On the basis of the equation obtained, the antioxidant activity of tested sample is calculated (Sochor et al., 2010; Shirazi, Khattak, Shukri & Nasyriq, 2014). Usually, for comparison the degree of the $ABTS^+$ and DPPH radical scavenge activity the Trolox Equivalent Antioxidant Capacity (TEAC or TE) is applied. Trolox is a commercial vitamin E analogue; results are expressed as amount of Trolox equivalents per volume (Badarinath, Mallikarjuna, Chetty, Ramkanth, Rajan & Gnanaprakash, 2010; Sochor et al., 2010; Moon & Shibamoto, 2009**).** For the TPC assay the results are usually presented as a gallic acid equivalent (e.g. as a mgGA/g). However, as Shahidi and Zhong (2015) noted, caffeic acid, ferulic acid and catechins are also popular (Shahidi & Zhong, 2015).

For the statistical analysis also various computer programs can be applied – in this trial it was decided to used KyPlot – a simple program created by KyensLab Incorporated. For this experiment important is to find differences or similarities between samples (various herbs) and between type of extraction solvents (water, ethanol, and aqueous ethanol). The T-Tukey test is a perfect tool for this analysis. An important factor is also a comparison between herbs and extraction solvents and assays. The tests listed above are all spectrophotometric methods based on a absorbance measurement during a discolored reactions between reagent and potential antioxidant substances. Those assays based on the different chemical reaction – the radical scavenging, the ion metal reduction etc. and therefore the response to antioxidant substances can be slightly different. But together, those methods allows to determine how strong the antioxidant properties has the tested compound and whether depends on the selected solvent.

4. RESULTS AND DISCUSSION

In this chapter the results of the experiment will be discussed. Due to the application of four different herbs (couch grass, dandelion, milk thistle, fireweed) each of them will be briefly characterized. Also, the applied methods for the antioxidant properties measurement will be characterized from the chemical point of view.

The antioxidant activity of extracts from herb materials are presented in Table 1 and the radical scavenging activity for $ABTS^{+}$ and DPPH methods are presented in Table 2. As it can be seen in general, the total polyphenolic content (TPC) was significantly ($P<0.05$) higher in all samples treated with aqueous ethanol as a solvent (Table 1). As it was mentioned earlier, the amount of extracted polyphenol compounds depends on solvent selection. Grujic et al. (2012) noted that mono-component solvent is not as much effective as their mixture. Addition of water to ethanol facilitates extraction through polar medium creation.

Tab. 1. Antioxidant and radical scavenging activity of herb extracts

PARAMETER	SAMPLE	SOLVENT		
		WATER	AQUOEUS ETHANOL (50:50)	ETHANOL
TPC [μg GA/ml]	C	0.005 ± 0.00^{Cb}	0.019 ± 0.01^{Da}	0.00 ± 0.01^{Cb}
	D	0.09 ± 0.00^{Bb}	0.244 ± 0.04^{Ba}	0.023 ± 0.00^{Bc}
	M	0.009 ± 0.01^{Cc}	0.083 ± 0.00^{Ca}	0.05 ± 0.00^{Ab}
	F	0.468 ± 0.01^{Ab}	0.625 ± 0.01^{Aa}	0.057 ± 0.00^{Ac}

C – couch grass extract (*A. repens*), D – dandelion (*T. officinale*) extract, M – milk thistle (*S. marianum*) extract, F – fireweed (*E. angustifolium*) herb. Means with different capital letters are significantly different ($p<0.05$) in the same column. Means with different small letters are significantly different ($p<0.05$) in the same row. Means ± standard error.

Agropyron repens (couch grass or quack grass) is a plant with highly branched, long yellowish-white rhizomes. In folk medicine it is usually used for treating various symptoms of urinary disease – prostatic disease, urinary infections, it's also used for calming spasms and pains in the urinary tract and as a soothing diuretic remedy. Couch grass contains phenol substances, carbohydrates, pectins, saponins and essential oils. Moreover, *A. repens* herb is a rich source of minerals especially silica (Al-Snafi, 2015). Furthermore, couch grass is an aggressive herb, presenting allelopathic effects on higher plants (due to the tricin presence) (Friebe, Schultz, Kock & Schnabl, 1995). Due to the lack of information

about polyphenol substances in couch grass extract, it could be suspected that *A. repens* is not a rich source of biologically active compounds. It was noted that couch grass shows the lowest values of antioxidant properties compared to the other herbs (Tab. 1).

Silybum marianum (milk thistle) is a herb used as a medicine for various liver diseases (e.g. removing excess bile from gallbladder, protecting the organ from poisoning, as intoxication after *Amanita* spices consumption). The main antioxidant substance in milk thistle is silymarin. Silymarin is a mixture of flavonolignans (silibinin, silybin A and B, isosilybin A and B) which ensure strong hepatoprotective effects (Soleimani, Delghandi, Moallem & Karimi, 2019). Elwekeel, Elfishawy & AbouZid (2013) found that part of the plant can determine the polyphenol amount – mature fruit and fruit heads of milk thistle contain the highest concentration of silymarin. Moreover, as Chambers et al. (2017) observed sylimarin is usually extracted from seeds (seeds cake or whole pulverized seeds) by hexagene, petroleum, ethyl acetate, acetone or methanol extraction (in conventional methods). However, in the experiment carried out by Barreto, Wallace, Carrier & Clausen (2003) hot water as solvent in extraction process was applied. Hot water presented the solubility characteristic similar to methanol and ethanol, which increase along with the temperature. Authors noted, that depending on temperature of the extraction process (85°C and 100°C), different amounts of polyphenol compounds were identified (taxifolin, silychristin and silybin A and B respectively). This situation was explained by the various polarity of those substances. Authors also pointed out that extraction at 50°C was not so effective compared to the other temperatures. In our experiment, herb extractions were carried out in 40°C so it could be concluded, that this ratio of temperature was too low for milk thistle flavonolignan extractions.

Taraxacum officinale L. (dandelion) was used in folk medicine as antidiabetic, diuretic medicine and as a substance enhancing the immune response. Extract from dandelion also presents hepatoprotective and antioxidant effects. The main component of dandelion is chicorcic acid (dicaffeyltartaric acid) but the plant also contains other polyphenol compounds (e.g. saponins, phenols, flavonoids). For the extraction of antioxidant substances from dandelion aqueous ethanol mixture was the best solvent. The obtained data agree with the results presented by Ivanov (2014), which carried out an extraction process at 80°C, although the extract from *T. officinale* did not present the strongest antioxidant properties. As Ghaima, Hashim & Ali (2013) noted, most of the polyphenol substances are concentrated in flower, root and steam. This could explain lower antioxidant activity of dandelion extract in comparison to the extract from fireweed. However, Sengul et al. (2009) pointed out that the Folin-Ciocalteu method for TPC measurement is not an unlimited, perfect method. Differences between samples and results from other assays could be an effect of various chemical structure of phenolic compounds which affected their antioxidant activity. There is always a risk that antioxidant substances could react between

themselves, so for these reasons, several methods for antioxidant property measurements should always be applied. Moreover, solvent type and the method used boh have a significant influence on the values of antioxidant activity of tested samples (Skotti, Anastasaki, Kanellou, Polissiou & Tarantilis, 2014), which are in agreement with our study.

The antioxidant activity is the ability to inhibit the oxidation process (Shalaby & Shanab, 2013). The amount of the polyphenol substances is usually correlated with antioxidant properties of the extract and with the radical scavenging ability (Ivanov, 2014; Sengul et al., 2009). This dependence was observed for samples measured by using $ABTS^+$ method (Table 2). For samples tested by DPPH assay, it was noted that lower amount of the polyphenol substances in extract from couch grass did not result in lower percentage of radical inhibition values. Extract from milk thistle and fireweed also presented similar data. This dependence was found for ethanolic extract.

Tab. 2. The radical scavenging activity of herbs extracts

PARAMETER	SAMPLE	SOLVENT		
		WATER	AQOUEOUS ETHANOL (50:50)	ETHANOL
$ABTS^+$ [%]	C	$1,23\pm0,17^{Cc}$	$4,85\pm0,22^{Da}$	$3,6\pm0,15^{Db}$
	D	$11,08\pm1,2^{Bb}$	$23,95\pm2,52^{Ba}$	$12,15\pm2,72^{Cb}$
	M	$9,11\pm3,87^{Bb}$	$17,8\pm2,12^{Ca}$	$18,5\pm2,75^{Ba}$
	F	$92,22\pm0,63^{Aa}$	$93,37\pm0,08^{Aa}$	$27,73\pm2,55^{Ab}$
DPPH [%]	C	$1,77\pm0,41^{Dc}$	$2,92\pm0,18^{Db}$	$4,42\pm0,3^{Ca}$
	D	$7,62\pm1,3^{Bc}$	$37,45\pm2,76^{Ba}$	$21,41\pm1,74^{Bb}$
	M	$2,95\pm0,38^{Cc}$	$5,82\pm0,21^{Cb}$	$7,29\pm1,03^{Ca}$
	F	$84,85\pm0,12^{Ab}$	$85,36\pm0,1^{Ab}$	$118,6\pm3,56^{Aa}$

C – couch grass extract (*A. repens*), D – dandelion (*T. officinale*) extract, M – milk thistle (*S. marianum*) extract, F – fireweed (*E. angustifolium*). Means with different capital letters are significantly different ($p<0.05$) in the same column. Means with different small letters are significantly different ($p<0.05$) in the same row. Means ± standard error.

Epilobium angustifolium (fireweed) is a herb commonly used in an alternative medicine for gastrointestinal disorder, rectal bleeding, sleeping disorders, bladder, prostate and kidney diseases. The herb can be applied as an extract or as a tea. According to literature data, *E. angustifolium* is a very rich source of various polyphenol substances (e.g. tannins, phenolic acids, flavonoids, steroids) (Onar, Yusufoglu, Turker & Yanardag, 2012; Granica, Piwowarski, Czerwińska & Kiss, 2014; Schepetkin et al., 2016). Dudonne, Vitrac, Coutiere, Woillez

& Merillon (2009) pointed out that usually strong correlations between ABTS$^+$ and DPPH methods and TPC values are observed. Also, Onar et al. (2012) have found a positive relationship between the amount of polyphenol substances in fireweed water extract and radical scavenging capacity. Moreover, Schepetkin et al. (2016) pointed out that an aqueous extract has a stronger anti-proliferative activity than ethanol extracts. Also, in the experiment carried out by Ostrovska et al. (2017) it was found that ethanol extract from fireweed contains very high amount of polyphenols (26.95 g as gallic acid equivalent per 100 g of dry weight). In our experiment the highest amount of substances of antioxidant activity was found for aqueous ethanol as a solvent. This could be connected with the increase of the solvent polarity which increases the amount of extracted polyphenols (Grujic et al., 2012). According to our experiment, both water and aqueous ethanol extracts from *E. angustifolium* presented strong antioxidant activity. In that case, higher activity percentage (118.6%) of ethanol extract could be connected with the conditions of the method applied. Pérez-Jiménez & Saura-Calixto (2006) observed, that the values of tested samples depended on the polarity of the solvent. For the ABTS$^+$ method, along with the increase of the polarity of the solvent, values of ABTS$^+$ increased, which is in the agreement with our results.

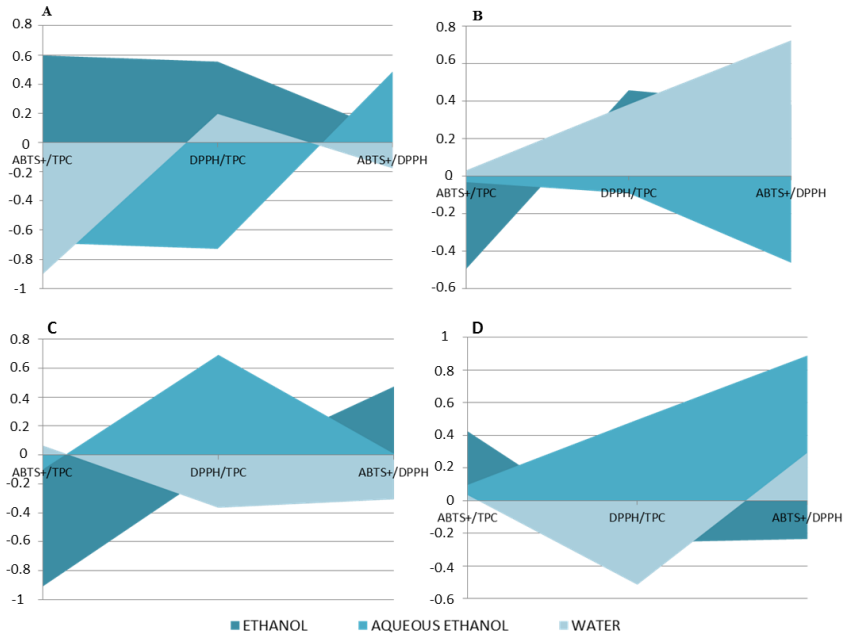

Fig. 3. Correlation between of type of solvent and antioxidant and radical scavenging parameters: A – Dandelion extracts, B – Fireweed extracts, C – Milk thistle extracts, D – Couch grass extracts

Both, $ABTS^+$ and DPPH methods are based on a degree of the color change which is correlated with antiradical scavenging ability of the sample (Dudonne et al., 2009). DPPH solution has a deep, violet color; $ABTS^+$ is characterized by blue-green color. During reactions with an antioxidant substances, radicals in the solutions are reduced and the color is loss, which allows to evaluate the antioxidant potential of tested samples (Alam, Bristi & Rafiquzzaman, 2013; Sahalaby & Shanab, 2012). Neverthless, some differences between methods occur. Preparing of the $ABTS^+$ takes more over 12 hours due to the chemical or enzymatic reactions which allow to free radical generations. For this reasons, the $ABTS^+$ working solution can give slightly different results along with the time. $ABTS^+$ is soluble in organic media and in water which allows to measure lipophilic and hydrophilic substances in tested sample. In the DPPH method, the 2,2-diphenyl-1-picrylhydrazyl is ready to dissolve and can be used directly after preparation. However, this radical is soluble only in organic, alcoholic media (Sahalaby & Shanab, 2012; Arnao, 2000; Thaipong, Boonprakob, Crosby, Cisneros-Zevallos & Bryne, 2006). Furthermore, how Singh & Singh (2008) notes for $ABTS^+$ and DPPH methods various wavelength can be applied (415, 660, 734, 820 and 515 to 528 nm respectively). In the experiment carried out by Arnao (2000) fresh fruits juices and wines were examined by using two antioxidant measurements methods: DPPH (515 nm) and $ABTS^+$ (for $ABTS^+$ two wavelengths – 730 and 414 nm – were applied). For the juices, wavelength 730 nm for $ABTS^+$ method and wavelength 515 nm for DPPH method gives convergent results. Author pointed out that the biggest differences between antioxidant capacities concerned red wines samples. The differences depending not only on the methods, but also depending on wavelength – in general, samples measured at 720 nm presented higher antioxidant activity (216.20 and 234.06 mg TEAC/100 ml) than measured at 414 nm (183.65 and 195.40 mg TEAC /100 ml). Fig. 3 shows the correlations between solvents and between method applied. The strongest, positive correlations were found for dependence between DPPH and TPC. No correlation was observe between ABTS+ and DPPH for dandelion (ethanol solvent) and milk thistle (aqueous ethanol solvent).

5. CONCLUSIONS

The data collected through the Vision-Pro™ UV-VIS software by different spectrophotometric method applied allows to state that aqueous ethanol as a solvent gives best results for phenolic compound extractions for all examined herbs. Extract from fireweed was characterized by the highest antioxidant and radical scavenging activities compared to the other herbs. However, it must be added, that extract from dandelion also showed strong antioxidant properties. As it was also noted, extracts from couch grass presented the lowest antioxidant potential compared to the analyzed extracts. In general, the antioxidant activity of the herb

extracts decreased in the following order W>D>M>C. Results of the research indicate that herb extracts can be applied in the food industry to extend the durability of the product, which allows to make product more attractive for the potential customers. Due to the strong antioxidant properties, extracts from *E. angustifolium, T. officinale* and *S. marianum* should be applied.

ACKNOWLEDGEMENTS

The research was financed under the program of the Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in 2019–2022 project number 029/RID/2018/19 funding amount 11 927 330.00 PLN".

REFERENCES

Alam, Md. N., Bristi, N. J., & Rafiquzzaman, M. (2013). Review on in vivo and in vitro methods evaluation of antioxidant activity. *Saudi Pharmaceutical Journal*, *21*, 143–152. doi:10.1016/j.jsps.2012.05.002

Al-Snafi, A. E. (2015). Chemical constituents and pharmacological importance of Agropyron repens – A review. *Research Journal of Pharmacology and Toxicology*, *01*(02), 37–41.

Arnao, M. B. (2000). Some methodological problems in the determination of antioxidant activity using chromogen radicals: a practical case. *Trends in Food Science & Technology*, *11*, 419–421. doi:10.1016/S0924-2244(01)00027-9

Badarinath, A. V., Mallikarjuna Rao, K., Madhu Sudhana Chetty, C., Ramkanth, S., Rajan, T. V. S., & Gnanaprakash, K. (2010). A Review on In-vitro antioxidant Methods: Comparisions, Correlations and Considerations. *International Journal of PharmTech Research*, *2*(2), 1276–1285.

Barreto, J. F. A., Wallace, S. N., Carrier, D. J., & Clausen, E. C. (2003). Extraction of Nutraceuticals from Milk Thistle I. Hot Water Extraction. *Applied Biochemistry and Biotechnology*, *108*(1-3), 881-889. doi:10.1385/ABAB:108:1-3:881

Biswas, A. K., Sahoo, J., & Chatli, M. K. (2011). A simple UV-Vis spectrophotometric method for determination of b-carotene content in raw carrot, sweet potato and supplemented chicken meat nuggets. *LWT – Food Science and Technology*, 44, 1809–1813. doi:10.1016/j.lwt.2011.03.017

Chambers, Ch. S., Holečková, V., Petrásková, L., Biedermann, D., Valentová, K., Buchta, M., & Křen, V. (2017). The silymarin composition and why does it matter??? *Food Research International*, *100*, 339–353. doi:10.1016/j.foodres.2017.07.017

Dudonne, S., Vitrac, X., Coutiere, P., Woillez, M., & Merillon, J.-M. (2009). Comparative Study of Antioxidant Properties and Total Phenolic Content of 30 Plant Extracts of Industrial Interest Using DPPH, ABTS, FRAP, SOD, and ORAC. Assays. *J. Agric. Food Chem.*, *57*, 1768–1774. doi:10.1021/jf803011r

Elwekeel, A., Elfishawy, A., & AbouZid, S. (2013). Silymarin content in *Silybum marianum* fruits at different maturity stages. *Journal of Medicinal Plants Research*, *7*(23), 1665–1669. doi:10.5897/JMPR12.0743

Falowo, A. B., Fayemi, P. O., & Muchenje, V. (2014). Natural antioxidants against lipid–protein oxidative deterioration in meat and meat products: A review. *Food Research International*, *64*, 171–181. doi:10.1016/j.foodres.2014.06.022

Friebe, A., Schulz, M., Kock, P., & Schnabl, H. (1995). Phytotoxins from shoot extracts and root exudates of agropyron repens seedlings. *Phytochemistry*, *38*(5), 1157–1159. doi:10.1016/0031-9422(94)00795-U

Ghaima, K. K., Hashim, N. M., & Ali, S. A. (2013). Antibacterial and antioxidant activities of ethyl acetate extract of nettle (Urtica dioica) and dandelion (Taraxacum officinale). *Journal of Applied Pharmaceutical Science*, *3*(05), 096-099. doi:10.7324/JAPS.2013.3518

Granica, S., Piwowarski, J. P., Czerwińska, M. E., & Kiss, A. K. (2014). Phytochemistry, pharmacology and traditional uses of different Epilobium species (Onagraceae): A review. *Journal of Ethnopharmacology*, *156*, 316–346. doi:10.1016/j.jep.2014.08.036

Grujic, N., Lepojevic, Z., Srdjenovic, B., Vladic, J., & Sudji, J. (2012). Effects of Different Extraction Methods and Conditions on the Phenolic Composition of Mate Tea Extracts. *Molecules*, *17*(3), 2518-1528. doi:10.3390/molecules17032518

Gupta, A., Naraniwal, M., & Kothari, V. (2012). Modern extraction methods for preparation of bioactive plant extracts. *International Journal of Applied and Natural Science*, *1*(1), 8–26.

Hinneburg, I., Damien Dorman, H. J., & Hiltunen, R. (2006). Antioxidant activities of extracts from selected culinary herbs and spices. *Food Chemistry*, *97*, 122–129. doi:10.1016/j.foodchem.2005.03.028

Hung, Y., de Kok, T. M., & Verbeke, W. (2016). Consumer attitude and purchase intention towards processed meat products with natural compounds and a reduced level of nitrite. *Meat Science Volume*, *121*, 119–126. doi:10.1016/j.meatsci.2016.06.002

Ignat, I., Volf, I., & Popa, V. I. (2011). A critical review of methods for characterisation of polyphenolic compounds in fruits and vegetables. *Food Chemistry*, *126*(4), 1821–1835. doi:10.1016/j.foodchem.2010.12.026

Ivanov, I. G. (2014). Polyphenols Content and Antioxidant Activities of Taraxacum officinale F.H. Wigg (Dandelion) Leaves. *International Journal of Pharmacognosy and Phytochemical Research 2014-15*, *6*(4), 889–893.

Jung, S., Choe, J. Ch., Kim, B., Yun, H., Kruk, Z. A., & Jo, Ch. (2010). Effect of dietary mixture of gallic acid and linoleic acid on antioxidative potential and quality of breast meat from broilers. *Meat Science*, *86*(2), 520–526. doi:10.1016/j.meatsci.2010.06.007

Karadag, A., Ozcelik, B., & Saner, S. (2009). Review of methods to determine antioxidant capacities. *Food Analytical Methods*, *2*, 41–60. doi:10.1007/s12161-008-9067-7

Kedare, S. B., & Singh, R. P. (2011). Genesis and development of DPPH method of antioxidant assay. *Journal of Food Science and Technology*, *48*(4), 412–422. doi:10.1007/s13197-011-0251-1

Martillanes, S., Rocha-Pimienta, J., Cabrera-Bañegil, M., Martín-Vertedor, D., & Delgado-Adámez, J. (2017). Phenolic compounds. Biological Activity. Application of Phenolic Compounds for Food Preservation: Food Additive and Active Packaging, In M. Soto-Hernandez, M. Palma-Tenango & M. del Rosario Garcia-Mateos (Eds.), *Phenolic Compounds – Biological Activity* (pp. 39–58). IntechOpen. doi:10.5772/66885

Moniruzzaman, M., Khalil, M. I., Sulaiman, S. A., & Gan, S. H. (2012). Advances in the analytical methods for determining the antioxidant properties of honey: a review. *African Journal of Traditional, Complementary and Alternative Medicines*, *9*(1), 36–42. doi:10.4314/ajtcam.v9i1.5

Moon, J.-K., & Shibamoto, T. (2009). Antioxidant Assays for Plant and Food Components. *Journal of Agricultural and Food Chemistry*, *57*, 1655–1666. doi:10.1021/jf803537k

Naczka, M., & Shahidi, F. (2004). Extraction and analysis of phenolics in food. *Journal of Chromatography*, *1054*(1–2), 95–111. doi:10.1016/j.chroma.2004.08.059

Onar, H. C., Yusufoglu, A., Turker, G., & Yanardag, R. (2012). Elastase, tyrosinase and lipoxygenase inhibition and antioxidant activity of an aqueous extract from Epilobium angustifolium L. leaves). *Journal of Medicinal Plants Research*, *6*(5), 716–726. doi:10.5897/JMPR11.1127

Oroian, M., & Escriche, I. (2015) Antioxidants: Characterization, natural sources, extraction and analysis. *Food Research International*, *74*, 10–36. doi:10.1016/j.foodres.2015.04.018

Ostrovska, H., Oleshchuk, O., Vannini, S., Cataldi, S., Albi, E., Codini, M., Moulas, A., Marchyshyn, S., Beccari, T., & Ceccarini, M. R. (2017). Epilobium angustifolium L.: A medicinal plant with therapeutic properties. *European Biotechnology Thematic Network Association*, *1*(2), 126–131. doi:10.24190/ISSN2564-615X/2017/02.03

Pérez-Jiménez, J., & Saura-Calixto, F. (2006). Effect of solvent and certain food constituents on different antioxidant capacity assays. *Food Research International*, *39*, 791–800. doi:10.1016/j.foodres.2006.02.003

Rodríguez-Rojo, S., Visentin, A., Maestri, D., & Cocero, M. J. (2012). Assisted extraction of rosemary antioxidants with green solvents. *Journal of Food Engineering*, *109*(1), 98–103. doi:10.1016/j.jfoodeng.2011.09.029

Schepetkin, I. A., Ramstead, A. G., Kirpotina, L. N., Voyich, J. M., Jutila, M. A., & Quinn, M. T. (2016). Therapeutic Potential of Polyphenols from Epilobium Angustifolium (Fireweed). *Phytotherapy Research*, *30*, 1287–1297. doi:10.1002/ptr.5648

Sengul, M., Yildiz, H., Gungor, N., Cetin, B., Eser, Z., & Ercisli, S. (2009). Total phenolic content, antioxidant and antimicrobial activities of some medicinal plants. *Pakistan Journal of Pharmaceutical Science*, *22*(1),102–106.

Shahidi, F., & Ambigaipalan, P. (2015). Phenolics and polyphenolics in foods, beverages and spices: Antioxidant activity and health effects – A review. *Journal of functional foods*, *18*, 820 – 897. doi:10.1016/j.jff.2015.06.018

Shahidi, F., & Zhong, Y. (2015). Measurement of antioxidant activity. *Journal of Functional foods*, *18*, 757–781.

Shalaby, E. A., & Shanab, S. M. M. (2013). Compraison of DPPH and ABTS assays for determining antioxidant potential of water and methanol of Spirulina platensis. *Indian Journal of Geo-Marine Sciences*, *42*(5), 556-564.

Shirazi, O. U., Khattak, M. M. A. K., Shukri, N. A. M., & Nasyriq, M. N. (2014). Determination of total phenolic, flavonoid content and free radical scavenging activities of common herbs and spices. *Journal of Pharmacognosy and Phytochemistry*, *3*(3), 104–108.

Singh, S., & Singh, R. P. (2008). In Vitro Methods of Assay of Antioxidants: An Overview. *Food Reviews International*, *24*(4), 392-415. doi:10.1080/87559120802304269

Singleton, V. L., & Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic–phosphotungstic acid reagents. *American Journal of Enology and Viticulture*, *16*, 144–158.

Skotti, E., Anastasaki, E., Kanellou, G., Polissiou, M., & Tarantilis, P. A. (2014). Total phenolic content, antioxidant activity and toxicity of aqueous extracts from selected Greek medicinal and aromatic plants. *Industrial Crops and Products*, *53*, 46–54. doi:10.1016/j.indcrop.2013.12.013

Sochor, J., Ryvolova, M., Krystofova, O., Salas, P., Hubalek, J., Adam, V., Trnkova, L., Havel, L., Beklova, M., Zehnalek, J., Provaznik, I., & Kizek, R. (2010). Fully Automated Spectrometric Protocols for Determination of Antioxidant Activity: Advantages and Disadvantages. *Molecules*, *15*, 8618–8640. doi:10.3390/molecules15128618

Soleimani, V., Delghandi, P. S., Moallem, S. A., & Karimi, G. (2019). Safety and toxicity of silymarin, the major constituent of milk thistle extract: An updated review. *Phytotherapy Research*, *33*, 1627–1638. doi:10.1002/ptr.6361

Thaipong, K., Boonprakob, U., Crosbyb, K., Cisneros-Zevallosc, L., & Byrne, D. H. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. *Journal of Food Composition and Analysis*, *19*, 669–675. doi:10.1016/j.jfca.2006.01.003

Wendakoon, Ch., Calderon, P., & Gagnon, D. (2012). Evaluation of Selected Medicinal Plants Extracted in Different Ethanol Concentrations for Antibacterial Activity against Human Pathogens. *Journal of Medicinally Active Plants*, *1*(2),60–68. doi:10.7275/R5GH9FV2

Wojdyło, A., Oszmiański, J., & Czemerys, R. (2007). Antioxidant activity and phenolic compounds in 32 selected herbs. *Food Chemistry*, *105*, 940–949. doi:10.1016/j.foodchem.2007.04.038

Yu, J., Wang, H., Zhan, J., & Huang, W. (2018). Review of recent UV–Vis and infrared spectroscopy researches on wine detection and discrimination. *Applied Spectroscopy Reviews*, *53*(1), 65–86. doi:10.1080/05704928.2017.1352511

Applied Computer Science, vol. 15, no. 4, pp. 63–74
doi:10.23743/acs-2019-30

Submitted: 2019-11-04
Revised: 2019-12-07
Accepted: 2019-12-24

BigData, Hadoop, RSA, Paillier, Cryptography

Shadan Mohammed Jihad ABDALWAHID[*], *Raghad Zuhair YOUSIF*[**] , *Shahab Wahhab KAREEM*[*]

ENHANCING APPROACH USING HYBRID PAILLER AND RSA FOR INFORMATION SECURITY IN BIGDATA

Abstract

The amount of data processed and stored in the cloud is growing dramatically. The traditional storage devices at both hardware and software levels cannot meet the requirement of the cloud. This fact motivates the need for a platform which can handle this problem. Hadoop is a deployed platform proposed to overcome this big data problem which often uses MapReduce architecture to process vast amounts of data of the cloud system. Hadoop has no strategy to assure the safety and confidentiality of the files saved inside the Hadoop distributed File system (HDFS). In the cloud, the protection of sensitive data is a critical issue in which data encryption schemes plays avital rule. This research proposes a hybrid system between two well-known asymmetric key cryptosystems (RSA, and Paillier) to encrypt the files stored in HDFS. Thus before saving data in HDFS, the proposed cryptosystem is utilized for encrypting the data. Each user of the cloud might upload files in two ways, non-safe or secure. The hybrid system shows higher computational complexity and less latency in comparison to the RSA cryptosystem alone.

1. INTRODUCTION

Cloud computing has attracted increasing attention since the last few years. Cloud computing provides users with a wide range of resources, such as computing platforms, storage, computing power, and internet applications. Amazon, Google,

[*] Erbil Polytechnic University, Erbil Technical Engineering College, Department of Information System Engineering, Erbil, Iraq, shadanbaban@yahoo.com, shahab.karim@epu.edu.iq
[**] Catholic University in Erbil, Information Technology Department, Erbil, Iraq, raghad.yousif@cue.edu.krd

IBM, Microsoft, etc. are the biggest cloud available in the markets now. With a growing number of companies utilizing resources in the cloud, data from different users need to be protected. Cloud computing is presently used in a tremendous amount in various fields. In daily life, huge amounts of data produced. Consumers use cloud computing services to store this huge amount of data. Some of the major challenges cloud computing faces are to secure, protect and process the data that is the user's property (Merla & Liang, 2017; Kareem, 2009). Big data refers to the processing and retrieval of massive data collection. Big data must also be concerned with the collection of essential and sensitive data from social sites and issues of government and hence, security. This collected data has to encrypt by using appropriate algorithms to secure them. The features of Big Data can be identified in term of four V's (Hilbert, 2016): Volume, Velocity, Variety and Veracity. Every subject holds its job of remaining in Big data. Thus, volume: the amount of data produced and might be stored it could be in the level of different size terabytes rather Petabytes. Variety: which are the data forms and its kinds, structure, unstructured and semi-structured. Velocity: which indicates an input and the output rates of data streams generated and stored in the system. In this context, an abstraction provided in a way that the systems within big data can eventually, independently collect data from the outgoing or incoming clip. Veracity: It's a term of data quality; this context is also Refers to data confidentiality, data privacy, integrity, and availability. Establishments must be grantee that the data and the analyses conducted on the data are precise. Big data processing has become almost pivotal for many governments and business applications with an incredible rate of data generated, collected and analyzed by computer systems (Amrulla, Mourya, Sanikommu & Afroz, 2018). Thus, many factors have participated in data huge increment like the emerge of IoT, object localization and tracking, besides the growing adoption of healthcare devices which gather personal statistics. This prevalence of big data has some disadvantages. The data collected usually involves some personal information about persons, or it is including secrets that would be problematic if the opponent discovers them. Criminal groups create underground markets for the possession and purchase of stolen personal information (Motoyama, McCoy, Levchenko, Savage & Voelker, 2011). Government intelligence services rely on personal, corporate and adverse government eavesdropping and competitive advantage systems. Most recent, highly publicized cyber-attacks against commercial attacks demonstrate this potential for damage, and government targets, it pays millions of dollars to these organizations and causes severe damage to the affected individuals and organizations (Kareem & Hussein, 2017). Furthermore, protection across cloud services is under its developing stage; a huge quantity about safety vulnerabilities would risk data in the cloud. The cloud administrators have no clue as to where and in what format the data is stored. Thus, adequate security measures must be modified to preserve the data, essentially of information leakage plus manipulation. Also, processing/analyzing enormous data in the data center is a dangerous problem in the cloud. Different

spread structures like HADOOP have recently been available (Li, Wang, Zhao, Pu, Zhu & Song, 2015; Ahamad, Akhtar, Hameed, 2019), like Google File System (Yang, Lin & Liu, 2013), which is developed to store and process Big Data. Still, the spread HADOOP structure is common with manufacturing and investigation centres. HADOOP holds pair organizations of functionalities, (i) For storage of large and unstructured data sets (HDFS), has been employed, and (ii) Map-Reduce framework for hug data manipulation. HADOOP usually serves among applying that have huge of data links also petabytes. As a literature survey Chao YANG et al. (Yang, Lin & Liu, 2013). Suggest a triple encryption scheme for enhancing the security of Hadoop. Thus the encryption of HDFS files is achieved by using DEA (Data Encryption Algorithm), whereas RSA has been used in the encryption of data key. Eventually, the RSA private key is secured using the IDEA (International Data Encryption Algorithm). Huixiang Zhou et al. (Zhou & Wen, 2014) They present CP-ABE (Ciphertext policy attribute-based encryption) scheme for access control instead of the traditional schemes like PKI, which requires all relevant customer data to be sent to the resource provider, thus destroying the privacy of the user, and takes more bandwidth and overhead processing. Masoumeh Rezaei Jam et al. (Jam, Khanli, Akbari & Javan, 2014) point out that currently, the core technology of cloud computing are services security and data privacy. A security mechanism based on Kerberos protocol for authentication firewalls of perimeter level security was presented (Ismael, Youail & Kareem, 2014). Security leak was handled by implementing the Apache sentry for access control, triple encryption of data using RSA, DES, IDEA algorithms, was proposed in protecting file system based on fully homomorphic encryption. R. PARMAR1 et al. (Parmar, Roy, Bhattacharyya, Bandyopadhyay & Kim, 2017), proposed a novel method which can be used to secure Hadoop, a cost-effective technique works in Hadoop cluster to give it 3-D security. Muhammad Usama, et al. (Usama & Zakaria, 2017), proposed Data compression and encryption for Hadoop. Hence a combined compression and encryption scheme was presented based on Tent Map and Piece-wise Linear Chaotic Map (PWLM), the proposed approach implements a masking pseudorandom keystream that strengthens the encryption process. The proposed algorithm, providing robust encryption and compression schemes.

HADOOP does not incorporate security mechanisms. The Application of ciphering algorithms in HADOOP data encryption, then storing them at HDFS has reported in several works. Ciphering schemes perform different replacements and do some manipulation on the clear message to transforms it into ciphertext, which must be random and incomprehensible. Different ciphered schemes were developed and employed for the sake of information security. Hence the two main categories are: (i) Symmetric-key (secret key) cryptosystems (Chandra, Bhattacharyya, Paira & Alam, 2014) like Advanced Encryption Standard (AES), Data Encryption Standard (DES), and Triple DES (ii) Asymmetric-key (public key) algorithms (Chandra, , Alam, Paira & Sanyal, 2014) like Elliptic Curve Diffie-Hellman (ECDH) and RSA. The proposed approach is considered as an

attempt to improve what was presented by the paper (Usama & Zakaria, 2017) at both of encipherment /decipherment procedures for securing files of Big Data-based Hadoop-integrated AES and OTP algorithms (Mahmoud, Hegazy & Khafagy, 2018). An architecture to secure Hadoop was examined in paper (Park & Lee, 2013). Thus for data encryption and decryption, AES encryption/decryption classes are added. Implement two HDFS pairing integrations and HDFS-RSA (Shetty & Manjaiah, 2016) applied since various amazing kinds of extensions from HDFS. Analyses demonstrated adequate expenses for understanding processes also significant overhead for recording actions (Yang, Lin & Liu, 2013). Three encryption scheme (Inukollu, Arsi & Ravuri, 2014) integrated with cloud data storage system depending on Hadoop to encrypt files in HDFS based on DES and RSA then refer to IDEA for securing the RSA private key for the users. The encryption of the HDFS files implemented when they stored in a buffer after uploading data to HDFS. In this work, a modified asymmetric key cryptosystem is being presented to secure Big data. The following is the organization of this paper: Section II outlines the security framework. Section III, based on HDFS and MapReduce, presents the Big Data at HADOOP. Section IV discusses the proposed optimized hybrid encipherment algorithm and compare it with the classical public-key cryptosystems before applying it to secure Big Data at HADOOP. Section V presents the discussion of the simulation results. Finally, section VI list the conclusions.

2. SECURITY ISSUES

Big data is about data storage, data processing, data recovery. Many technologies, such as memory management, transaction management, visualization and networking, are used for these purposes. These technologies security issues are also applicable to big data. Big data's four major security issues are authentication, data level, network level and generic matters (Bhandarkar, 2010; Raghad, Kareem & Hasan, 2016).

2.1. Authentication Level Issues

A lot of clusters and nodes are present. Each node has priorities or rights that are different. Administrative nodes can access any data. But sometimes it will steal or manipulate the critical user data if any malicious node has organizational priority. Many nodes are joining clusters for faster execution with parallel processing. Any malicious node can disturb the group in the event of no authentication. Logging in big data plays an important role. If logging not provided, no activity that modifies or deletes data will record. If the new node joins the cluster, the absence of logging will not recognize it. Users may also sometimes use malicious data unless the log provided.

2.2. Data level issues

Data is an essential part of big data and also plays a vital role. Data is nothing but some of the government or social networking sites necessary and personal information about us. The main issues that could be handled by the data level are integrity and availability of data like protection and distribution of data. Big data environments such as Hadoop store the data as it is without encryption to improve efficiency. If the hacker accesses the machines, he/she cannot be stopped. Information stored in a distributed data store for quick access in many nodes with replicas. But if hacker deletes or manipulates any reproduction or information from another node, then it will be difficult to recover that data.

2.3. Network-level issues

There are many nodes in clusters, and these nodes are used to compute or process data. This data processing can be done anywhere between the cluster nodes. It is, therefore, difficult to determine which node data is processed. It will be complicated because of this difficulty on which node safety should be provided. Two or more nodes can communicate or share their data/resources via the network. RPC (Remote Procedure Call) often used for network communication. But until and unless it is encrypted, RPC will not be secure.

2.4. General level issues

Many technologies are also used in the big data environment to process the data for some traditional security tools for security purposes. Over the years, traditional tools have been developed. Thus with the new distributed form of big data, these tools may not be performed well. As big data uses many data storage, data processing and data recovery technologies, there may be some complexities due to these different technologies.

3. BIG DATA AND HADOOP

Hadoop architecture consists mainly of two primary components which are: (HDFS) to store Big Data and MapReduce to analyze Big Data (Bhardwaj, Singh, Vanraj & Narayan, 2015). HDFS is a file management system used for the distributed storage of massive datasets on the Hadoop cluster in with a default block size of 64 MB (Dubey, Jain & Mittal, 2015). After storing the input files in HDFS, then it manipulated with MapReduce software. Eventually, the results moved to the output folder of HDFS (Dean & Ghemawat, 2008). MapReduce in Hadoop is an application software designed for processing huge volumes of data sets over machine set (Zhou & Wen, 2014). MapReduce is the core scheme used by the Hadoop system

for spreading a bunch of work. Each input data, which inhabits throughout the cluster on a distributed file system, is divided into groups of equal size to facilitate and simplify in a suitable, and almost error-free manner the enormous volumes from processing the data under parallel at huge organizations regarding tools. As specified by the name, MapReduce involves two –stages like data calculation within Hadoop, the initial stage is the map, and the other stage signifies reducing, i.e. a huge amount from data sets is transformed inside structured key-value pairs and provided since inputs (Dean & Ghemawat, 2008).

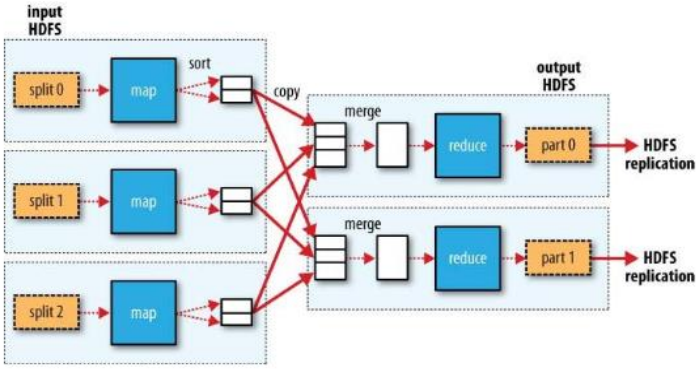

Fig. 1. MapReduce Data Stream

Figure 1 shows the MapReduce computation data flow. The mapper doesn't write directly on disk but uses the benefit from buffering some writings. Every mapper becomes a round buffer of memory among default size is 100 MB which can do modified through improving each property of (io. sort. mb). That makes a rapid flush. If the buffer is loaded up before specific inception, it initiates the transfer to the disk the content of the barrier. Before each spill appears on the drive, each thread separations these data based on the reducers that require ongoing background thread performs any sort of in-memory within the key-based partition before the spill takes place to the disk. If a mixer is started, it applies the output of the in-memory kind (Dean & Ghemawat, 2008).

4. PROPOSED ALGORITHMS

Hadoop is the primary provider of large-scale cloud data processing and storage, and is, therefore, uses some techniques of encryption to ensure security. This paper introduces new technology – this technique based on cascading two public-key cryptosystems (RSA and Paillier) (Kareem, 2009). Hybridization's a way to over-come the limitations of using each cryptosystem individually and to improve security. It is considered that all the files written to HDFS must be previously encrypted.

The HDFS client is responsible for keys generations (public and private keys). Then the proposed hybrid system is employed While the file caching in HDFS encrypted it utilizing the unstructured data for the file. The HDFS starts sending an encrypted file on the data nodes. These stages shown in Figure 2.

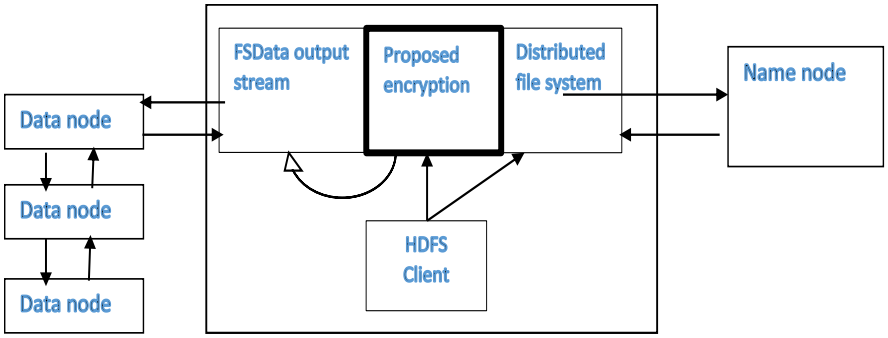

Fig. 2. Encryption procedure in HDFS

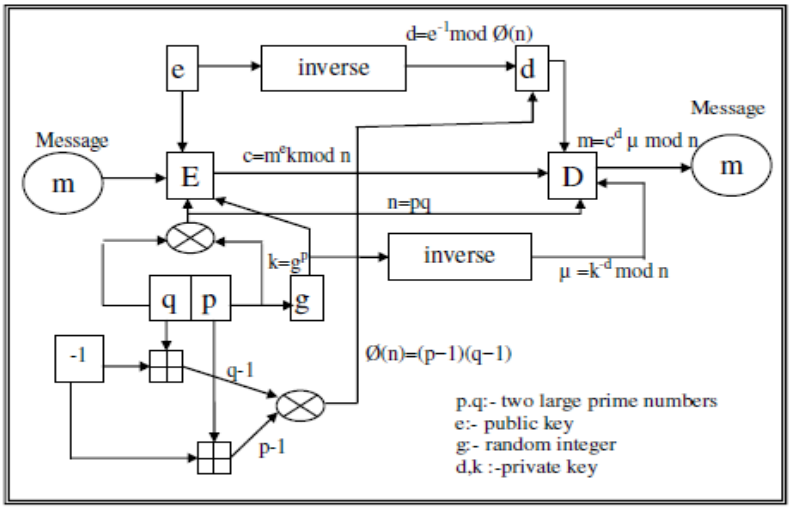

Fig. 3. Process of hybrid public key algorithm

HDFS consists of a Name Node that stores Metadata which manages the namespace the file system and monitor clients obtain for the files that encrypted. The files that encrypted is made up regarding one or higher blocks collected within a collection of data nodes. This proposed hybrid system described in Figure 3.

From Figure 3. the keys (public and private) generation procedure is based on the mechanism used by the RSA cryptosystem and its depicted by Algorithm1 below:

Algorithm 1*: – Key Generation of the proposed algorithm*
INPUT*: Choose big prime random numbers p and q*
OUTPUT: *A private key (p,q,d) and a public key, (n; e),*
User B received message from user A.

1. *Select two large random (and distinct) primes p and q, each roughly the same size.*
2. *Compute n = p*q and Ø(n) = (p − 1)(q − 1).*
3. *Select a random integer e, 1 < e < Ø , such that gcd(e; Ø) = 1.*
4. *Select a random integer g and compute* $k=g^p \bmod n$
5. *Use the extended Euclidean algorithm to compute the unique integer d, 1< d < Ø, such that ed ≡ 1 (mod Ø).*
6. *User's public key is (n; e); the user's private key is (d,k).*

The encryption process takes the scenario of Paillier algorithm to encrypt the message (m); thus, the cypher text computed as (C), with an augmented parameter $k = g^p$.

$$C = m^{e*k} \bmod n \tag{1}$$

Algorithm2 below shows in detail the encryption procedures:
Algorithm2*: – Encryption process of the proposed algorithm*
INPUT*: Plaintext to encrypt, and receiving the user's public key (n; e).*
OUTPUT*: Encrypted ciphertext.*
User A sends the message to user B.
To encrypt B should do the following:

(a) Obtain A's authentic public key (n; e).
(b) Represent the message as an integer m in the interval [0; n − 1].
(c) Compute $c = (m^e*k) \bmod n$.
(d) Send the ciphertext c to A.

This parameter transmitted along with the ciphertext, this parameter is used in decryption process to recover (m) to compute k then k^{-1}, While. This method explained in the following algorithms:

$$m = c^d * k^{-1} \bmod n \tag{2}$$

To recover the message, m four messages generated m. So the correct plain text is one of them. This procedure explained in (Algorithm3) below:
Algorithm3*:-Decryption process of the proposed algorithm*
INPUT*: Received encrypted ciphertext and the receiver's private key a.*
OUTPUT*: Original plaintext.*
To recover plaintext m from c, B should do the following:

(a) Compute $\mu = k^{-d} \bmod n$.
(b)Compute $m = c^d * \mu \bmod n$.

After applying the proposed encryption scheme, data stored in the cloud. Thus via HADOOP File System (HDFS), data will be stored in a cluster. Whenever the user requests data, the server will introduce the encrypted data to the decryption procedure. The user then uses the private key to retrieve the decrypted data using a hybrid system which is the proposal of this paper.

5. EXPERIMENTAL RESULTS AND ANALYSIS

HDFS and MapReduce functions have used for performance evaluation of encrypted HDFS. Each node has i3 core, four processors, 4 GB of memory, and 750 GB of the hard disk. Encryption Time: The time is taken by the RAS alone or the hybrid algorithm to encrypt the Hadoop divided dataset files into ciphertext using a key. It is calculated in milliseconds. Decryption Time: The time taken by the RAS alone or the hybrid system to decrypt the Hadoop split dataset files back into the plaintext using the private key. It calculated in milliseconds. Thus the Encryption Time is equivalent to system current time before encryption subtracted from it the system current time after encryption. Whereas the Decryption Time is equal to the system current time before decryption subtracted from it the system current time after decryption.

Figure 4 depicts the results of the comparison between encryption schemes, the RSA alone and the Hybrid system with different file sizes. It's clear that the proposed method showed efficient time consumption compared to the RSA for files size stars from 100 MB and ends with 1 GB with a step size of 100 MB.

And hence, the proposed method (Hybrid system) in the encryption stage is faster than the default RSA. Figure 5 shows the running time for RSA and the proposed method in the decryption stage. The encrypted files applied to this stage are of different sizes. By utilizing both of RSA and the hybrid system (the proposed method), it's evident that decryption time needed by the hybrid ciphered method is shorter than that required by RSA. Table I. Shows the computational complexity of the Hybrid cipher (proposed method) with, RSA and Paillier cryptosystems from which it's clear that the proposed method has doubled the computational complexity as compared to the individual systems (RSA or Paillier).

Tab. 1 Computational Complexity of the Proposed Method, RSA and Paillier

Method	Encryption	Decryption
RSA	$T(c) = O (\log n)^3$	$T(M) =O (\log n)^3$
Paillier	$T (c) = 2O (\log n)^2$	$T(M) =O (\log n)^3$
Hybrid system	$T(c) =2O (\log n)^3 +O (\log n)$	$T(M)= 2 O (\log n)^3 +3+O (\log n)$

Tab. 2. Time of encryption Process the file size in MB and time in second

File size in MB	RSA Encryption	Paillier Encryption	Hybrid method
100	220.5882	444.85287	224.26467
200	235.9926	550.6494	247.2303
300	265.7913	911.7259	324.5126
400	285.9564	1106.6512	304.7076
500	300.3254	1162.2592	320.0188
600	310.3456	1201.0374	330.6961
700	327.2813	1266.5786	348.7423
800	345.0357	1335.2881	367.6609
900	359.7902	1392.3880	383.3838
1000	368.5446	1426.2676	392.7114

Tab. 3. Time of encryption Process the file size in MB and time in second

File size in MB	RSA	Paillier	Hybrid Method
100	47.6470	78.5294	61.7646
200	133.6091	234.9738	181.2826
300	686.2018	788.6856	788.6856
400	955.9391	1493.3747	1031.6014
500	1391.3672	2186.2474	1526.0225

6. CONCLUSION

While Hadoop allows overcoming the difficulties confronted by big data in businesses and organisations, it has no security mechanism. An attacker or eavesdropper may compromise the data stored in Hadoop. The authenticity of data is always at stake, while Hadoop takes not implement any protection tool. Before storing it in HDFS, the proposed Hybrid asymmetric key algorithm encrypts the file content by obtaining that of the various network attacks. The file or data can therefore now collected under Hadoop without troubling on protection problems through utilizing the encryption methods to the records before it saved in Hadoop. The proposed Hybrid system supports most cloud computing system service models such as Service Software (SaaS), Service Infrastructure (IaaS), and Service Platform (PaaS). It also supports data management and security issues (Authentication, Integrity, Availability, and Confidentiality) in security and key management for data transfer.

The proposed method showed excellent time consumption with different file sizes in the encryption and decryption stages with higher complexity (double the computational complexity in decryption stages). The future work would be integrating both of ElGamal and RSA asymmetric key cryptosystem. The limitation of the proposed hybrid system is the time taken by the decryption procedure to discover the correct plaintext form the four alternatives messages resulted by Paillier method decryption.

REFERENCES

Bhardwaj, A., Singh, V. K., Vanraj, & Narayan, Y. (2015). Analyzing BigData with Hadoop Cluster in HDInsight Azure Cloud. *Annual IEEE India Conference (INDICON).* India: IEEE. doi:10.1109/INDICON.2015.7443472

Ahamad, D., Akhtar, M., & Hameed, S. A. (2019). A Review and Analysis of Big Data and MapReduce. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(1), 1–3.

Amrulla, G., Mourya, M., Sanikommu, R. R., & Afroz, A. A. (2018). A Survey of : Securing Cloud Data under Key Exposure. *International Journal of Advanced Trends in Computer Science and Engineering, 7*(3), 30–33.

Bhandarkar, M. (2010). MapReduce programming with apache Hadoop. *International Symposium on Parallel & Distributed Processing (IPDPS)* (pp. 1-2). Atlanta: IEEE .

Chandra, S., Alam, S. S., Paira, S., & Sanyal, G. (2014). A comparative survey of symmetric and asymmetric key cryptography. *International Conference on Electronics, Communication and Computational Engineering (ICECCE)* (pp. 83–93). IEEE.

Chandra, S., Bhattacharyya, S., Paira, S., & Alam, S. S. (2014). A Study and Analysis on Symmetric Cryptography. *ICSEMR* (pp. 1–8). IEEE.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *6th Symposium on Operating Systems Design and Implementation* (pp. 107–113). ACM.

Dubey, A. K., Jain, V., & Mittal, A. P. (2015). Stock Market Prediction using Hadoop Map-Reduce Ecosystem. *2nd International Conference on Computing for Sustainable Global Development* (pp. 616–621). IEEE.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, *34*(1), 135–174.

Inukollu, V. N., Arsi, S., & Ravuri, S. R. (2014). Security Issues Associted With big Data In Cloud Computing. *International Journal of Network Security & Its Applications (IJNSA)*, *6*(3), 45–56.

Ismael, R. S., Youail, R. S., & Kareem, S. W. (2014). Image Encryption by Using RC4 Algorithm. *European Academic Research, II*(4), 5833–5839.

Jam, M. R., Khanli, L. M., Akbari, M. K., & Javan, M. S. (2014). A Survey on Security of Hadoop. *4th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 716–721). IEEE.

Kareem, S. W. (2009). *Hybrid Public Key Encryption Algorithms For E-Commerce.* Erbil: University of Salahaddin–Hawler.

Kareem, S. W., & Hussein, Y. T. (2017). Survey and New Security methodology of Routing Protocol in AD-Hoc Network. *The 1st International Conference on Information Technology* (pp. 452–464). Erbil.

Li, B., Wang, M., Zhao, Y., Pu, G., Zhu, H., & Song, F. (2015). Modeling and Verifying Google File System Modeling and Verifying Google File System. *16th International Symposium on High Assurance Systems Engineering* (pp. 207–214). IEEE.

Mahmoud, H., Hegazy, A., & Khafagy, M. H. (2018). An approach for Big Data Security based on Hadoop Distributed File system. *International Conference on Innovative Trends in Computer Engineering (ITCE 2018).* Aswan: Aswan University.

Merla, P., & Liang, Y. (2017). Data analysis using hadoop MapReduce environment. *IEEE International Conference on Big Data (Big Data)* (pp. 4783–4785). Boston: IEEE.

Motoyama, M., McCoy, D., Levchenko, K., Savage, S., & Voelker, G. M. (2011). An analysis of underground forums. *ACM SIG- COMM Conference on Internet Measurement Conference IMC '11* (pp. 71–80). New York: ACM.

Park, S., & Lee, Y. (2013). Secure Hadoop with Encrypted HDFS. *International Conference on Grid and Pervasive Computing* (pp. 134–141). Springer.

Parmar, R. R., Roy, S., Bhattacharyya, D., Bandyopadhyay, S. K., & Kim, T.-H. (2017). Large-scale encryption in the Hadoop environment: Challenges and solutions. *IEEE Access*, *5*, 7156–7163.

Raghad, Z. Y., Kareem, S. W., & Hasan, A. O. (2016). *Design Security System Based on AES and MD5 for Smart Card*. Sulaimanyia: Charmo university.

Shetty, M. M., & Manjaiah, D. H. (2016). Data security in Hadoop distributed file system. *IEEE Int. Conf. Emerg. Technol. Trends Comput. Commun. Electr. Eng. ICETT 2016* (pp. 939–944). IEEE.

Usama, M., & Zakaria, N. (2017). *Chaos-Based Simultaneous Compression and Encryption for Hadoop*. PLoS One.

Yang, Ch., Lin, W., & Liu, M. (2013). A Novel Triple Encryption Scheme for Hadoop-based Cloud Data Security. *Fourth International Conference on Emerging Intelligent Data and Web Technologies* (pp. 437–442). IEEE.

Zhou, H., & Wen, Q. (2014). A new solution of data security accessing for Hadoop based on CP-ABE. *5th International Conference on Software Engineering and Service Science* (pp. 525–528). IEEE.

Applied Computer Science, vol. 15, no. 4, pp. 75–92
doi:10.23743/acs-2019-31

Submitted: 2019-10-29
Revised: 2019-11-25
Accepted: 2019-12-01

Short-term Load Forecasting, Deep Learning Architectures, RNN, LSTM, CNN, SAE

Saheed ADEWUYI[*], Segun AINA[**], Moses UZUNUIGBE[***], Aderonke LAWAL[**], Adeniran OLUWARANTI[**]

AN OVERVIEW OF DEEP LEARNING TECHNIQUES FOR SHORT-TERM ELECTRICITY LOAD FORECASTING

Abstract

This paper presents an overview of some Deep Learning (DL) techniques applicable to forecasting electricity consumptions, especially in the short-term horizon. The paper introduced key parts of four DL architectures including the RNN, LSTM, CNN and SAE, which are recently adopted in implementing Short-term (electricity) Load Forecasting problems. It further presented a model approach for solving such problems. The eventual implication of the study is to present an insightful direction about concepts of the DL methods for forecasting electricity loads in the short-term period, especially to a potential researcher in quest of solving similar problems.

1. INTRODUCTION

The power systems structure is characterised by complex infrastructures that are necessary for the sourcing and delivery of electricity to end-users. In order to deliver electricity to end-users, the power Generation Company (GenCo) will transport power through networks of power transmission lines, which is controlled by the Transmission Companies (TransCo). The Distribution Companies (DisCo), also known as the Utilities, receive power from the TranCo and ensure its safe delivery to consumers.

[*] Osun State University, Department of Information and Communication Technology, Osogbo, Osun State, Nigeria, saheed.adewuyi@uniosun.edu.ng
[**] Obafemi Awolowo University, Department of Computer Science and Engineering, Ile-Ife, Osun State, Nigeria, s.aina@oauife.edu.ng,
[***] Transmission Company of Nigeria, 132/33 kV, Ajebandele, Ile Ife, Osun State, Nigeria, moses_uzunuigbe@yahoo.com

The utilities have the responsibility to meet the electricity demand of their customers. Electricity demand is the load on the electrical system which the system must satisfactorily bear and service for customers. This load or demand increases as population increases. To manage this, Utilities need to carry out load forecasts of electricity ahead of need. Electricity load forecasting is germane to the GenCo, DisCo, and TransCo stakeholders, especially in a deregulated economy. The electricity market deregulation and unbundling of the power industry has engendered this even more. Each of the resulting companies that is, the GenCo, TransCo and DisCo have responsibilities to meet the demand of her customers ahead of electricity supply. Therefore, electricity load forecasting is their essential routine.

Load forecasting in power systems is the prediction of users' demands on the grid prior to actual consumption. Load forecasting will, therefore, help the power players across classes to manage the power system's load effectively and efficiently. With load forecasting, the Utilities will especially make essential decisions critical to its operation and planning. This includes purchasing decision and power generation decision. Also, the decisions can be one of the following: Load switching; infrastructure development; capacity planning; maintenance schedules; energy demand; production adjustment; and contract evaluation (Ghullam & Angelos, 2017; Kuo & Huang, 2018; Seunghoung, Hongseok & Jaekoo, 2017).

Therefore, tackling the problem of electricity consumption forecasting using deep learning techniques involve simplifying it based on Fig. 1. Fig. 1 shows that load forecasting problem can be simplified based on five categories: Model, Horizon, Aim, Variables and Area (Luis *et. al.,* 2012, 2013, 2014). Load forecasting is simplified based on the type of the model to develop, which results in linear or non-linear model categorisation. Fig. 1 further classifies the problem into four categories based on the horizon (Feinberg & Genethliou, 2005; Luis *et. al.,* 2012). This includes Very Short Term Load Forecasting (VSTLF) which falls within seconds or minutes, Short Term Load Forecasting (STLF); spanning a week from an hour, Medium-Term Load Forecasting (MTLF); which covers a period from a week up to a few months and Long Term Load Forecasting (LTLF); which is usually more than a year (Luis *et. al.,* 2012, 2013, 2014). Furthermore, the electricity load forecasting is simplified based on the aim of the forecast that is, the values to be predicted, which can be a single value or multiple values (Luis *et. al.,* 2014). When considering certain factors influencing load consumption, which are ingredients of forecast estimation, it is essential to include time, weather and customer class (Feinberg & Genethliou, 2005; Swalin, 2019). These factors are classified as variables in Fig. 1. They, therefore, include variables such as load data and calendar data; or a combination of load data, calendar data and weather parameters; or another combination of load data, calendar data and other data variables which may be demographic, economic and social in nature, usually prevalent in the residential class of electricity end-users' class.

The Fig. 1 also categorises area of forecast estimation as including country, region/city and community/microgrids. In any situations, load forecasting can be implemented for one of these identified categories.

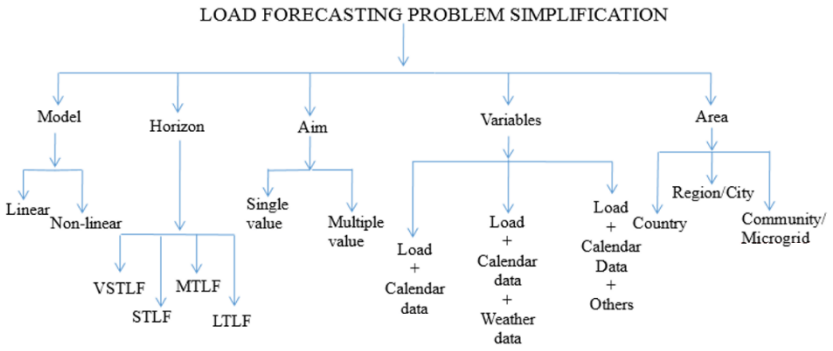

Fig. 1. Load forecasting problem categorisation

In this paper, the scope of the study is limited to four Deep Learning (DL) algorithms already established in the literature (Bengio, 2009; Brownlee, 2018; Chengdong, Zixiang, Dongbin, Jianqiang & Guiqing, 2017; Deng, 2013; Ghullam & Angelos, 2017; Hamedmoghadam, Joorabloo & Jalili,2018; Hussein & Hussein, 2017, Kuo & Huang, 2018; Nor, Rahaini & Siti, 2018; Swalin, 2019) for modelling electricity load forecasting. The techniques are the standard Recurrent Neural Network (RNN), Long Short Term Memory (LSTM) network, Convolutional Neural Network (CNN), and Stacked Autoencoder (SAE). Also, the concepts of deep learning were introduced. These cover the motivating influences for the technology; its classification, as well as a few terminologies of deep architectures. Also, the problems with primitive methods for modelling sequential electricity load were highlighted. As a glimpse, we present a concise approach to developing load forecasting problems. This will be to define a robust or scalable model, preprocess data and carry out performance evaluation of the model.

The rest of the paper is arranged as follow. Section 2 presents background knowledge and an overview of modelling electricity load forecasting in the short-term. Section 3 extends the discussion on deep architectures relevant for STLF problems. Here, the four DL techniques are discussed in detail. In Section 4, an approach towards implementing the STLF problem is presented. Finally, in Section 5, conclusions from the work are drawn.

2. BACKGROUND

In order to establish a background into the study of electricity load forecasting, there is a need to introduce the concept from the perspective of how a model can be developed for it. We categorized the approaches into two: the primitive or classical approach and the DL approach. These two are briefly discussed in the subsections that follow.

2.1. Primitive approach

Classical approaches have been researched and discussed in academia and are applied in the industry with varying success. The techniques are well reported. Linear methods like Auto-Regressive Integrated Moving Average (ARIMA) model have been widely chosen because they are easily understood and well effective on some problems (Brownlee, 2018). This classical approach has three variants including itself, ARIMA and two others namely: Auto-Regressive Moving Average (ARMA) and Autoregressive (AR) models (Nor, Rahaini & Siti, 2018). ARIMA models are regression type of models that adopt lagged values of the dependent variable and or random disturbance term as explanatory variable (Sarabjit & Rupinderjit, 2013). The explanatory variables have in-built dependence relationship (Sarabjit & Rupinderjit, 2013). The model is an integration of two autoregression and moving average models. A Seasonal ARIMA can be hybridised with a Back Propagation Algorithm Neural Network to achieve a more accurate load forecasting (Yi, Jie, Yanhua & Caihong, 2013). It will, however, be interesting to highlight that these traditional methods have suffered from various limitations:

1. Complete data: This means it sees data as wholesome and cannot manage issues of missing or corrupt data automatically.
2. Linear relationships: This implies that it addresses only linearities and leaves out complexities in the data distributions.
3. Fixed temporal dependence: This implies that the relationship between observations at each time-step as well as the number of lagged features in the input must be scrutinised and explicitly stated.
4. Univariate data: Usually, real-life problems are characterised by multiple variables as input, but classical approaches are mostly able to handle univariates.
5. One-step forecast: Practical problems will require multiple-step forecasts than the single-step forecast characterized by primitive models.

The problems highlighted, therefore, leave users with the requirement to hand-engineer features which are expensive to create (Gamboa, 2017). At the wake of these issues, Deep learning (DL) techniques emerged. Their inventions have tremendously advanced the sphere of artificial intelligence capabilities to solving vast human problems. In fact, DL has helped in providing dependable solutions to those problems sustained by primitive methods while handling more sophisticated problems.

2.2. Concept of deep learning

Deep learning is a machine learning technique that learns features and tasks directly from data (Brownlee, 2018). Data can be images, text, or sound. Any of these data types can be used as input to deep learning models for problem-solving. Deep learning also refers to a class of machine learning techniques, where many layers of information-processing stages in hierarchical architectures are exploited for pattern classification and representation or feature learning (Deng, 2013). Further to this, deep learning involves neural networks that are able to naturally learn arbitrary complex mappings from inputs to outputs. It also has support for multiple inputs and outputs. Interestingly, some of these features offer a great promise for electricity load forecasting (Brownlee, 2018), particularly on problems characterised by complex/nonlinear dependencies, or multivalent inputs, and multi-step forecasting. These features and other neural network capabilities offer great promises, such as the automatic feature learning characteristic of convolutional neural networks and the natural support for sequential data in recurrent neural networks (Brownlee, 2018). Sequential data are datasets whose features are constrained by time, making it a little difficult. Electricity load data is an example. Electricity load profile of a customer is measured as a function of time (per hour). This is evident in customers' electricity bills, which is computed in terms of energy used. A unit of energy used is measured in Kilowatt Hour (kWh). Furthermore, electricity loads, unlike other machine learning problems solved by classification of labels or even regression analysis of quantities, add time complexity. This inherently makes them have certain temporal dependencies among data features (Brownlee, 2018). The temporal dependencies, therefore, introduce difficulties in handling data for the purpose of model's fitting and evaluation. Conversely, the temporal structure characterising the electricity load can equally enhance modelling by providing added structures such as trends and seasonality, which when leveraged improve model performance on problems (Brownlee, 2018).

2.2.1. Motivation for deep learning

Although deep learning techniques have been in use for some time, in recent times it gained a lot of popularity due to certain developments. These, according to (Deng, 2013) are: First, due to DL techniques' increased accuracy at performing several human-related tasks. This is as a result of recent advances in machine learning and signal/image processing. Second, as a result of the increased chip processing ability, such as the Graphics Processing Units, engendered by a high reduction in the cost of computing hardware. Third, there are larger volumes of labelled data available.

The above highlighted motivating points that are key to DL algorithms' application on a particular problem. The same motivations are the reasons for applying DL techniques on electricity loads forecasting problems.

2.2.2. Classifying deep learning architectures

Most deep learning architectures use neural network-based methods. This is why some deep learning techniques such as Deep Belief Networks (DBN) are interchangeably referred to as Deep Neural Network (DNN) in literature (Deng & Yu, 2013). The term deep in deep learning refers to the number of hidden layers in the neural network. The hidden layer count can be unlimited. In order to classify deep learning architecture, a three-way classification scheme is summarized from various work done by researchers and industry experts. These classification schemes are grouped into generative, discriminative, and hybrid algorithms (Deng, 2013; Deng & Yu, 2013).

The generative deep architectures are learning architectures intended to characterize high-order correlation properties of the observed or visible data for pattern analysis or synthesis purposes, and/or characterize the joint statistical distributions of the visible data and their associated classes. They are unsupervised learning algorithms. Examples include autoencoders, Boltzmann machine and sum-product network (Deng, 2013). In relation to this, we can assert that some electricity load forecasting problems belong to this category. This is because the problem has been tackled with a few DL algorithms, such as the auto-encoder, in this class. This is documented in (Hussein & Hussein, 2017).

The discriminative type of deep architecture focuses on direct provision of discriminative power for pattern classification instances, often achieved by characterising the posterior distribution of classes conditioned on the visible data. These architectures are supervised learning in nature (Deng & Yu, 2013). Examples of such models include some learning algorithms like stacked networks, recurrent neural network and convolutional neural network. We also found in (Hussein & Hussein, 2017) that electricity load forecasting problem is a member of this class.

The hybrid deep architectures are either comprising or making use of both generative and discriminative model components. This architecture type has the goal of discrimination, assisted at times by the outcomes of generative deep networks. This can be accomplished by better optimization and or regularization of the deep networks in discriminative models. The goal can also be accomplished when discriminative criteria for supervised learning are used to estimate the parameters in any of the deep generative or unsupervised deep networks in generative models (Deng, 2013; Deng & Yu, 2013). Examples of such architectures include DNN-DBN model, DNN-Conditional Random Field (CRF) among several others. Similarly, electricity load data have also been solved by hybridised algorithms. The work of (Hussein & Hussein, 2017) is good case. The study combined DNN with SAEs and CNN with LSTM.

2.2.3. Application of deep learning techniques to STLF problem

Deep learning architectures have been applied to acoustics, images and signal processing studies, with tremendous successes (Merkel, Povinelli & Brown, 2017). For this reason, its applicability to electricity load forecasting problems is also a possibility because load profiles are characterised by some non-linear factors. For instance, in (Hussein & Hussein, 2017), some deep learning techniques were analysed. The DL architecture analysed include the Feed Forward Neural Network (FFNN), that is characterised by influx of input signal from the input layer to the output layer, one layer at a time, without looping back. Other model architectures also analysed in (Hussein & Hussein, 2017) are the Recurrent Neural Network (RNN), which allows data to flow in any direction, and the Convolutional Neural Network (CNN), which is applied to computer vision problems and acoustic modelling as well. Similarly, the other architectures are the Stacked Autoencoders (SAE) and Long Short Term Memory (LSTM). Seunghyoung *et al.* (2017), the need to investigate important aspect of Demand-Side Management (DMS) was studied on forecasting electricity loads. The study forecasts individual customer's daily load using deep neural network based approach. Ghullam and Angelos (2017), developed Feed-Forward DNN and Recurrent DNN models to predict short term electricity loads. The study analysed time and frequency as features influencing electricity load demand. Wan (2014), presented Restricted Boltmann Machine (RBM) as deep learning pre-training method for STLF problem. Kuo and Huang (2018), studied an introduction of accurate deep neural network algorithm for short-term load forecasting (STLF). Rahul *et al*. (2018), developed a novel approach for long-term load forecasting; although, this was with the aim to forecasting electricity loads at hourly horizon. In Hussein (2018), an investigation into application of DNN to forecasting electricity loads was done for a DisCo. The study also proposed a multi-layered DNN's system for the problem. The next section discusses more in perspective some of these deep learning techniques used for predicting electricity loads especially in the short term horizon.

3. SPECIFIC TECHNIQUES AND METHODS FOR ELECTRICITY LOAD FORECASTING

In this section we introduce four of the deep learning techniques with their structures, which have been adopted in the literature for solving short term electricity load forecasting problems.

3.1. Recurrent neural network

Recurrent Neural Network (RNN) is a deep learning technique with a long history, but only become popular as result today is traceable to the work published by Schmidhuber and Sepp (1997) and a few other researchers. An RNN can be understood as copies of a single network but each one transferring its signal to another as in Fig. 2. So, to recognise the need for an RNN, patterns in signals must be observed to change with time, just as in a typical electricity loads data. In such scenarios the best model is an RNN or its advanced variant, the LSTM.

This deep learning model has a simple structure with a built-in feedback loop, see Fig. 2; which allows the network to transfer electricity loads from previous time-step to next time-step. This capability thereby results in a situation referred to as persistent flow of information, recognised as RNNs' memory capabilities. The architecture of an RNN consists of units interacting in discrete time via weighted and directed connections with weights *wij*, linking unit *j* to *i*, with *i* being the first unit and *j* the last unit of the network (Hussein & Hussein, 2017).

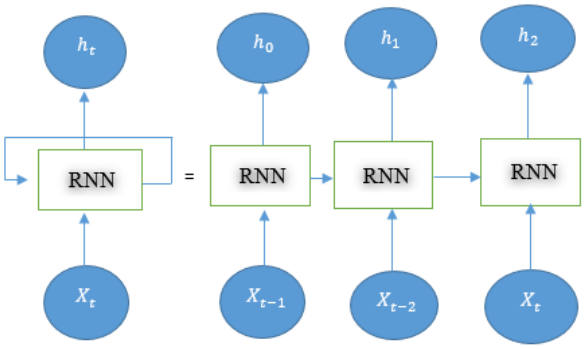

Fig. 2. Standard RNN structure

Furthermore, every unit has an activation function $\hat{m}(t)$ which is adjusted at every time-step, $t = (1,2,3,..)$ for each electricity load exposed to it. An activation, $\hat{m}^i$ of unit i is updated by computing its network input sum N^i where:

$$N^i = \sum_j w_{ij}\hat{m}(t-1) \quad (1)$$

and squashing it with a differentiable function like sigmoid function σ results in:

$$\hat{m}(t) = \sigma(N^i(t)) \quad (2)$$

3.2. Convolutional neural network

A Convolutional Neural Network (CNN) is one of the most popular algorithms for deep learning with images and videos. Like other algorithms, a CNN is composed of an input layer, an output layer, and many hidden layers sandwiched. The CNN provides better accuracy in highly non-linear problems. The CNN uses the idea of weight sharing whose sets are treated as kernels (Merkel, Povinelli, & Brown, 2017). Fig. 3. is a one dimensional convolution and pooling layer. After the convolution process, the inputs $x1, x2, x3, x4, x5, x6$ (in this case are the electricity load consumed by power users) are transformed to the feature maps $c1, c2, c3, c4$.

Pooling follows, wherein the feature map of convolution layer is sampled and dimension is reduced. The feature dimension before pooling is 4 but after the process the dimension is reduced to 2, as shown in Fig. 3.

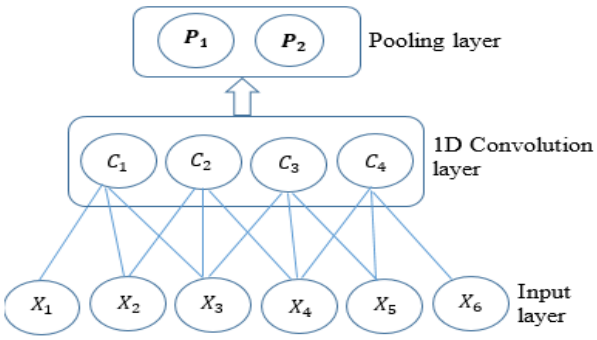

Fig. 3. One dimensional convolution and pooling

The pooling process is a vital procedure of the CNN architecture, for extraction of essential convolution features. A feature map is derived by repeated use of a function across sub-regions of the entire image, that is, by convolution of the input image with a linear filter adding a bias term and then applying a non-linear function. So by denoting the mth feature map at a given layer as h^m, for which filters are determined by the weights W^m and bias b_m, the feature map h^m is derived as in Eq. (3), for hyperbolic non-linearities:

$$h_{ij}^m = \tanh(W^m * x)_{ij} + b_m \tag{3}$$

In order to obtain a richer representation of the electricity load data, the hidden layer can be stacked, that is structured to compose multiple feature maps as in (Hussein & Hussein, 2017).

3.3. Stacked autoencoders

Stacked Autoencoders (SAEs) are autoencoders characterised by some multiple building blocks to construct its deep structure (Chengdong *et al.,* 2017). The SAEs utilise stacked architecture, with an autoencoder in each layer (Hussein and Hussein, 2018). Autoencoders are neural networks that have the power to encode its input data, such as electricity loads consumption, into a new representation using unsupervised type of learning. They are hidden layer of neurons that are trained to encode raw input data into a new representation and decode them to reconstruct the original input with minimal deformation possible (Hamedmoghadam, Joorabloo & Jalili, 2018). The target, that is the output, is equal to the input of the model (Chengdong *et al.,* 2017; Hussein, 2018).

The following are, therefore, three important things to take note about autoencoder:

1. Autoencoders are data-specific: This means that the load forecasting deep learning technique will only be able to encode those data similar to what had seen before.
2. Autoencoders are lossy: This will mean that the decoded outputs that is, the electricity load reconstruction, in this case, will be degraded and compared to its original inputs.
3. Autoencoders are learned automatically from data examples: This means it is easy to train specialised instances of the algorithm that will perform well on specific type of input. Therefore, no new data feature engineering is required, but data training.

As an illustration however, Fig. 4. is a simple autoencoder with *H* hidden layer nodes. As a matter of fact, autoencoder has two main parts, the encoder and decoder. The encoding process seeks to exploit and then reveal a hidden representation $\sigma_1(\mathrm{x})$, of typical electricity load profiles, which can be computed as:

$$\sigma_1(\mathrm{x}) = \mathrm{f}(w_1 x + b_1) \tag{4}$$

where: $w_1 -$ is an encoding matrix,
$b_1 -$ is an encoding bias vector,
$f(\cdot) -$ is the activation function.

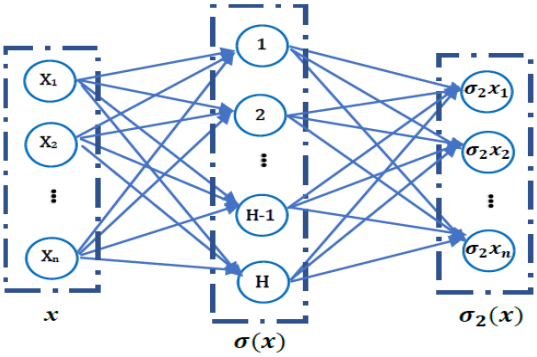

Fig. 4. A simple autoencoder

As a matter of fact, autoencoder has two main parts, the encoder and decoder. The encoding process seeks to exploit and then reveal a hidden representation $\sigma_1(x)$, of typical electricity load profiles, which can be computed as:

$$\sigma_1(x) = f(w_1 x + b_1) \tag{4}$$

where, w_1 is an encoding matrix, $b1$ is an encoding bias vector, and $f(\cdot)$ is the activation function. The activation function can be any of the sigmoid, Rectified Linear unit (ReLu) or tanh functions. Conversely, the decoding process requires that a decoding matrix be determined in order to recover the reconstructed hidden representation $\sigma_1(x)$, back into its original form that is, $\sigma_2(x)$. This therefore leads to the computation of the decoded output as in Eq. (5):

$$\sigma_2(x) = g(w_2 \sigma_1(x) + b_2) \tag{5}$$

where: $w_2 -$ is the decoding matrix,
$b_2 -$ is the decoding bias vector,
$g(\cdot) -$ is the encoding activation function.

It is expected that the error between the input x and the reconstruction $\sigma_2(x)$ is as minimal as possible. To ascertain this, it is imperative to minimise the loss function Eq. (6) below:

$$L(\sigma_2(x)) = \frac{1}{2}\sum_{m=1}^{N} \left\| x^{(m)} - x_2(x^{(m)}) \right\|^2 \tag{6}$$

Moreover, the optimal parameter set of the autoencoder can as well be known by solving the following optimization problem:

$$\Psi = \{w_1, w_2\} = \underset{w_1, w_2}{arg} \min L\,(x, \sigma_2(x)\,) \tag{7}$$

where: $w_1 -$ is the encoding,
$w_2 -$ is the decoding matrices,
$L(x, \sigma 2(x)) -$ is loss function, minimised for optimisation purposes,
$\Psi -$ is a notation defined for the optimisation problem.

In the autoencoder, this optimization problem, Ψ is often solved using one of the variants of the backpropagation algorithms, such as the conjugate gradient method or the steepest descent method (Chengdong *et al.*, 2017).

In summary, the technique affords extraction of useful features essential to forecasting electricity loads.

3.4. Long Short-Term Memory

This is a special kind of RNN. It was developed to overcome the lingering problem of long-term dependency suffered by the standard RNN architecture. This, therefore, provides dependable solutions to a lot of sequence problems. The LSTM networks structure consist of many connected LSTM cells as simply depicted in Fig. 5. The Fig. 5 is a single cell of an LSTM network showing arrows pointing towards the structure and another exiting from it. These will mean that there exists connected LSTM cells before and after it.

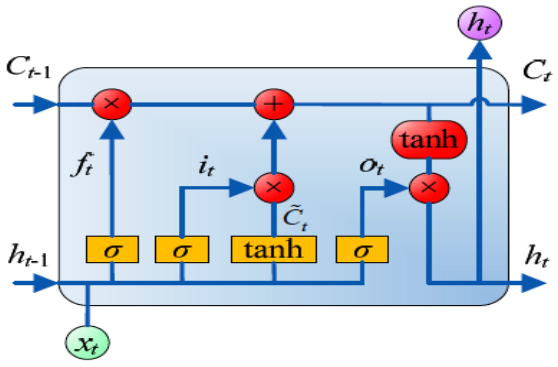

Fig. 5. LSTM Structure (Kuo & Huang, 2018)

The LSTM internal structure is characterised by four interacting neural network layers, also known as the repeating modules. The main idea about the LSTM network operating mechanism is its cell state and the gates layers. The cell state can be termed as a conveyor belt and runs straight down the modules, with only minor interactions, making data to flow seamlessly and uncorrupted down the chain. The LSTM is capable to remove or add information to the cell state.

This is carefully regulated by the gates. The gates are a way to optionally let data input, such as an electricity load features, through the cell state. They are composed of a sigmoid neural net layer and a pointwise multiplication operation as in Fig. 5. The sigmoid layer output is binary, describing how much of each component should pass through the gate or not. A '0' means "let nothing through," while a '1' will mean "let everything through!" An LSTM has three of these (forget, input and output) gates, to protect and control the cell state.

The following are the system of equations that were established in the literature as responsible for the satisfactorily operating performance of the deep learning LSTM technique.

$$f_t = \sigma\big(W_f.[h_{t-1}, x_t] + b_f\big) \tag{8}$$

$$i_t = \sigma(W_i.[h_t - 1, x_t] + b_i) \tag{9}$$

$$\widehat{C_t} = tanh(W_C.[h_t - 1, x_t] + b_c) \tag{10}$$

$$C_t = f_t \times C_{t-1} + i_t \times \widehat{C_t} \tag{11}$$

$$o_t = \sigma(W_o \cdot [h_t - 1, x_t] + b_o) \tag{12}$$

$$h_t = o_t \times tanh(C_t) \tag{13}$$

Eq. (8) decides which of the previous information to be discarded of the cell state. This decision is made via the sigmoid layer called forget gate layer. In Eq. (9) and Eq. (10), decision on new information to be included to the cell state is made. There are two parts. The first is the sigmoid layer termed input gate layer, which decides which value is to be updated and the second is the hyperbolic tangent layer, which creates vector of new candidate values that could be added to the cell to the state. Eq. (11) updates the old cell state, C_{t-1} into the new cell state, C_t. Notice that the previous steps already decided what to do, so update is only done here. Eq. (12) and Eq. (13) decide what is to be generated from the network, which is the prediction outcome. The output is based on the cell state, in filtered state. First, a sigmoid layer is run, which decides what parts of the cell state is to be output. Then, the cell state is passed through a *tanh* layer and is multiplied by the output of the sigmoid gate, so that only the decided parts are predicted. This is the final stage of the LSTM operating mechanism.

From Eqs. (8) to (13), the notations used are defined as follows: x_t, is the load features input at time t; x_{t-1}, is the previous hidden layer computations. The w_f, w_i, w_c, and w_o are respective weight matrices of the forget, input, cell state and output gate layers of the structure, that are regulating data input inflow. The b_f, b_i, b_c, and b_o, are the respective bias vector for each gate layer and c_{t-1}, is the

previous cell state while c_t, is the new cell state. Furthermore, whereas $\hat{c}_t$, is a vector of new candidate values for the cell state; σ and $tanh$, are respectively sigmoid and hyperbolic tangent activation functions. The notation o_t, is the output of the sigmoid gate and o_h, is the output of the current hidden layer.

In summary, LSTM is an exciting model for forecasting power loads because of its ability to effectively handle datasets characterised by an order of time such as the electricity consumption data.

3.5. Modelling Approach

In order to forecasting electricity loads in the short- term horizon, there is need to synthesise the problem into its constituents. These implies the defining, compiling, fitting, evaluating and predicting the model. These technical constituents of model development are adoptable by adapting the Fig. 6.

The subsections that follow further detail a typical roadmap for applying electricity load data on any of the deep learning techniques discussed so far, in order to forecast the next hour or day-ahead consumption profile.

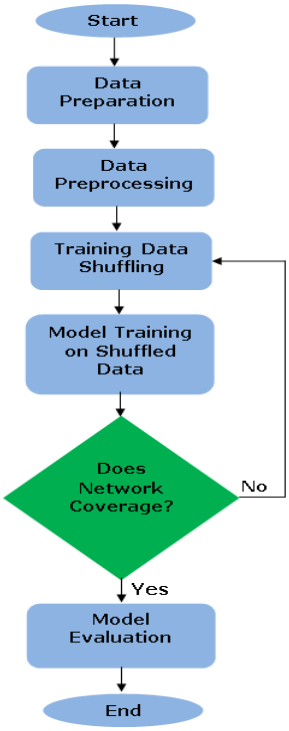

Fig. 6. A simple flowchart for load forecasting

3.5.1. Model training data source

Data is an important aspect of STLF forecasting problem. Therefore, in order to tackle the problem, it is imperative to known the source of the model training dataset. The model data source is how an electricity forecast engine obtains its training set, for the network prediction goal. So, a two-way classification was developed and is delineated in Fig. 7. This is the engineering approach and Artificial Intelligence (AI) or data-driven approach. The engineering method obtains model training data from the context features of the building structure. It also gets the data via the system information of the Heating, Ventilating, and Air Conditioning (HVAC) appliances of that structure and other home appliances, which load forecasting task is estimated. Conversely, the AI or data-driven approach gets training data from the historic electricity consumption data of the study area. So, depending on the load forecasting task in view a researcher would have to make decisive choice regarding model training set.

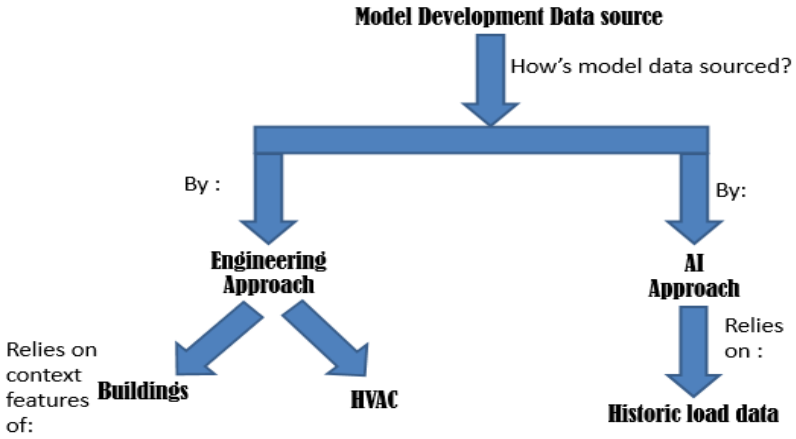

Fig. 7. How model obtains training data

3.5.2. Data preprocessing

Having obtained the model's training data, the next task will be to carry out appropriate preprocessing procedure on the electricity load forecasting problem. This approach will usually involve data cleansing and preparation. Data cleansing would mean that the data be devoid of any incompleteness or missing value, noise, and inconsistency. Data cleaning routines work to "clean" the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. A typical framework for handling missing data is presented in (Swalin, 2019). Preparing data would require that it is preprocessed by scaling numeric data and transforming categorical data. Numeric data scaling will improve

network stability and modelling performance. This can be achieved through normalization and standardization methods. Similarly, data transformation is converting labels data to quantities, this is achieved in two standard steps namely: integer encoding and one-hot encoding (Brownlee, 2018).

4. EVALUATION METRICS

The evaluation of model's performance of load forecasting model can be assessed by obeying the objective function of the model. The applicability and suitability of a deep learning model on typical electricity load forecasting problem is measured by some evaluation metrics such as the root mean square error and mean absolute percentage error. In general, performance evaluation of a regression type of deep learning model, as in this case, can be measured by one of the following metrics, among others:

4.1. Mean Square Error and Root Mean Square Error

The Root Mean Square Error (RMSE) is a frequently used measure of the differences between samples predicted by a model and the values actually observed. RMSE is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are. RMSE therefore measures how spread out these residuals are. In other words, it speaks about how concentrated the data is around the line of best fit. The RMSE for a training and test sets should be very similar if a good model is built. Formally, the RMSE is given as below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{pred} - Y_{true}\right)^2} \tag{14}$$

Where: $RMSE$ – is evaluation metric of interest,
N – is number of observations,
Y_{pred} – ordinates of the actual loads,
Y_{true} – ordinates of the predicted loads.

4.2. Mean Absolute Error

Mean Absolute Error (MAE) is a term used in determining absolute difference between two variables. Assume, Y_{pred} and Y_{true} are variables of paired observations expressing the same conditions, the MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_{pred} - Y_{true}\right| \tag{15}$$

4.3. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is yet another approach to evaluating model prediction accuracy. It often expresses accuracy in terms of percentage. This is expressed mathematically as:

$$MAPE = \frac{100\%}{N}\left|\sum_{i=1}^{N}\frac{Y_{pred}-Y_{true}}{Y_{pred}}\right| \quad (16)$$

5. CONCLUSIONS

The foregoing study discusses some of the deep learning techniques that have been applied to electricity load data when modeling typical load forecasting problem. Load forecasting was also categorised as including the type of model, aim of forecast, horizon of forecast, variables of interest in modeling load forecasting problem and the area where forecast is to cover. More so, the interested model data source is categorised as harvestable from an engineering source or AI source. Finally, the DL techniques for modeling a typical electricity load forecasting problem is discussed.

REFERENCES

Bengio, Y. (2009). Learning deep architectures for AI. *Foundation and Trends in Machine Learning*, *2*(1), 1–127.

Brownlee, J. (Ed.) (2018). *Deep learning for time series forecasting: Predicting the future with MLPs, CNNs and LSTMs in Python*. Machine learning mastery.

Chengdong, L., Zixiang, D., Dongbin, Z., Jianqiang, Y., & Guiqing, Z. (2017). Building energy consumption prediction: An extreme deep learning approach. *Energies*, *10*(10), 1525–1545.

Deng, L. (2013). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, 3(2). doi:10.1017/ATSIP

Deng, L., & Yu, D. (2013). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, *7*(3-4), 197–387.

Feinberg, E. A., & Genethliou, D. (2005). Load forecasting. In J. H. Chow, F.F. Wu, & J. Momoh (Eds.), *Applied Mathematics for Restructured Electric Power Systems. Power Electronics and Power Systems*. Springer, Boston, MA.

Gamboa, J. (2017). Deep learning for time-series analysis. arXiv: 1701.01887.

Ghullam, M. U., & Angelos, K. M. (2017). Short term power load forecasting using deep neural networks. *ICNC*, *10*(1109), 594–598, 7876196.

Hamedmoghadam, H., Joorabloo, N., & Jalili, M. (2018). *Australia's long-term electricity demand forecasting using deep neural networks*. arXiv: preprint arXiv:1801.02148.

Hussein, A. (2018). *Deep Learning Based Approaches for Imitation Learning* (doctoral dissertation). Robert Gordon University Aberdeen, Scotland.

Hussein, S., & Hussein, P. (2017). Load forecasting using deep neural networks. In *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE. doi:10.1109/ISGT.2017.8085971

Kuo, P., & Huang, C. (2018). A high-precision artificial neural networks model for short-term energy load management. *Energy*, *11*(1), 213–226.

Luis, H., Carlos, B., Javier, M. A., Lorena, C., Belen, C., Antonio, S., Diane, J. C., David, C., & Jorge, G. (2012). A study of relationship between weather variables and electric power demand inside a smart grid/ smart world. *MDPI Sensors*, *22*(9), 11571–11591.

Luis, H., Carlos, B., Javier, M. A., Lorena, C., Belen, C., Antonio, S., Diane, J. C., David, C., & Jorge, G. (2013). Short-term load forecasting for micro-grids based on artificial neural networks, *MDPI Sensors*, *6*(3), 1385–1408.

Luis, H., Carlos, B., Javier, M. A., Lorena, C., Belen, C., Antonio, S., & Jaime, L. (2014). Artificial neural network for short-term load forecasting in distribution systems, *MDPI*, *7*(3), 1576–1598.

Merkel, G. D., Povinelli, R. J., & Brown, R. H. (2017). Deep neural network regression for short-term load forecasting of natural gas. *Report: Marquette University*.

Nor, H. M., Rahaini, M. S., & Siti, H. H. A. (2018). ARIMA with Regression Model in Modelling electricity load demand, *Journal of Telecommunication, Electronic and Computer Engineering*, *8*(12), 113–116.

Rahul, K. A., Frankle, M., & Madan, M. T. (2018). Long term load forecasting with hourly predictions based on long-short-term-memory networks. In *2018 IEEE Texas Power and Energy Conference (TPEC)*. IEEE. doi:10.1109/TPEC.2018.8312088

Sarabjit, S., & Rupinderjit, S. (2013). ARIMA Based Short Term Load Forecasting for Punjab Region. *IJSR*, *4*(6), 1919–1822.

Schmidhuber, J., & Sepp, H. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Seunghoung, R., Hongseok, K., & Jaekoo, N. (2017). Deep neural network based demand side short term load forecasting. *Energies MDPI*, *10*(1), 3–23.

Swalin, A. (2019). How to handle missing data. *Towards Data Science*. Retrieved from https://towardsdatascience.com/how-tohandle-missing-data-8646b18db on 18/01/2019.

Wan, H. (2014). Deep neural network based load forecast. *Computer Modelling and New Technologies*, *18*(3), 258–262.

Yi, Y., Jie, W., Yanhua, C., & Caihong L. (2013). *A new strategy for short-term load forecasting*. Hindawi.

Applied Computer Science, vol. 15, no. 4, pp. 93–108
doi:10.23743/acs-2019-32

Submitted: 2019-09-10
Revised: 2019-09-13
Accepted: 2019-12-06

nduction motor, wavelet transformation, backlash zone, neural networks.

Marcin TOMCZYK[0000-0002-5383-7168]*, *Anna PLICHTA*[0000-0001-6503-308X]**, *Mariusz MIKULSKI****

APPLICATION OF WAVELET – NEURAL METHOD TO DETECT BACKLASH ZONE IN ELECTROMECHANICAL SYSTEMS GENERATING NOISES

Abstract

This paper presents a method of identifying the width of backlash zone in an electromechanical system generating noises. The system load is a series of rectangular pulses of constant amplitude, generated at equal intervals. The need for identification of the backlash zone is associated with a gradual increase of its width during the drive operation. The study uses wavelet analysis of signals and analysis of neural network weights obtained from the processing without supervised learning. The time-frequency signal representations of accelerations of the mechanical load components were investigated.

* Cracow Univeristy of Technology, Faculty of Electrical and Computer Engineering, Warszawska 24, 31-155 Kraków, Poland, marcin.tomczyk@pk.edu.pl
** Cracow Univeristy of Technology, Faculty of Computer Science and Telecommunications, Chair of Computer Science, Warszawska 24, 31-155 Kraków, Poland, aplichta@pk.edu.pl
*** State Higher Vocational School in Nowy Sącz, Institute of Engineering,Zamenhofa 1a, 33-300 Nowy Sącz, Poland, mmikulsk1@poczta.onet.pl

1. INTRODUCTION

The diagnostics of electromechanical processes deals with the recognition of undesired changes of their states. The states are presented in the form of a sequence of intentional actions conducted in the assumed time by a specific set of machines and devices at a determined amount of available resources. Faults and other destructive events resulting from the wear and tear as well as increased exploitation time can be reasons for changes in these states. If such a change of state exceeds specific value it should be detected by a diagnostic system, recognized as fault, and identified as quickly as possible, in its early phase of formation (Korbicz, Kościelny & Kowalczuk, 2002).

Some information carried by signals appearing in electromechanical systems are essential for diagnostic reasoning, and particular attention should be paid to their extraction and application (Zhang, Zhu, Yang, Yao & Lu, 2007).

Measuring signals generated by sensors or converters using the measuring path are subjected to analysis. However, they often contain irrelevant content, i.e. trends or fast-changing components whose character resembles noise. The attempts to limit such unnecessary characteristics and using digital-to-analogue converters to guarantee appropriate value of signal-to-noise relation can result in side effects having negative impact on the analysis, e.g. frequency masking or stroboscopic effect. Therefore, sampling frequency must be appropriate to the components of the signal that contain relevant information (Zhang, Zhu, Yang, Yao & Lu, 2007).

In mechanical systems containing backlash zones, non-linear elastic-absorbing elements or faulted bearings, diagnostic signals pertaining to time and frequency can be generated using transformation methods that enable simultaneous testing of the spectral properties in both of these fields (Duda, 2007).

One of the increasingly popular and applied methods of time-frequency analysis is wavelet transformation based on multistage signal decomposition at changeable resolution (Duda, 2007; Zając, 2009).

In contrast with Fourier analysis in which analysed functions are expressed by means of polynomials derived from harmonic functions, wavelet transformation describes them using special functions — wavelets derived from a dedicated function called mother wavelet. The created wavelet functions are subjected to repeated transformations. The set of base functions of transformation obtained in this way has a number important scalable properties related to the time and frequency; one can analyze the relationships between the particular function and its transformation coefficients (Doniec, 2010).

Owing to the recent progress in signal processing technology, many diagnostic methods have been presented that concern engine diagnostics by means of wavelet analysis. These are, for instance:

- discussion on the usefulness of wavelet analysis for the initial processing of diagnostic signals to train and test neural damage detectors of induction motors (Kowalski, 2003),

- detection of microcracks on the surface of bearing race (Zając, 2009), on its rolling element (Chebil, Noel, Mesbah & Derihe,2009), as well as in the bearing-connecting elements (Aktas & Turkmenoglu, 2010),
- analysis of electro-energetic signals by means of the high-resolution methods of spectrum analysis (Łobos, Leonowicz, Rezmer & Schegner, 2006),
- detection of rotor crack in induction cage motor by means of the frequency analysis using MCSA analysis as well as continuous wavelet transformation (Granda, Aguilar, Arcos-Aviles & Sotomayor, 2017).

Recently, more and more studies focus on the application of neural networks like, e.g.:

- using of single neural network to detect faults at the various stages of work of nuclear power plant (Wysogląd, 2003),
- detection of faults in chemical plants by means of dynamic networks (Fuente & Saludes, 2000),
- solving problems pertaining to modelling and classification in the object and system diagnostic processes using GMDH networks (*Group Methods and Data Handling*) (Duch, Korbicz, Rutkowski & Tadeusiewicz, 2000). This study contains a lot of neural architectures with dynamic properties characterized by their excellent efficiency during modelling of diagnosed processes.
- identification of mechanical parameters in three-phase induction cage motor by means of model of neural network using gradient decrease method (Balara, Timko, Źilkova & Leśo, 2017).

2. METHODOLOGY AND ANALYSIS OF THE FAULT IDENTIFICATION DIAGNOSTIC ALGORITHM

Simulation tests were conducted for the nominal conditions of the induction motor whose model was located in a stationary coordinate system related to the stator (model α, β, 0). The induction motor was loaded by a working machine similar in character to a dynamic mass-absorbing-elastic element.

Model of induction motor has been created in the MATLAB/Simulink environment. The following parameters of induction motor have been assumed in the conducted tests (parameters of the substitute scheme are expressed in relative units): $r_S = 0.059$, $r_w = 0.048$, $x_S = 1.92$, $x_R = 1.92$, $x_M = 1.82$, $w = x_S * x_w + - x_m * x_m = 0.374$, $T_m = 0.86$ [s].

Fig.1 presents a simplified diagram of connection of a working machine with the induction motor. The diagram includes backlash in the clutch connecting the induction motor drive with a working machine. It also includes the connection between the generator of normally distributed signals and generator of rectangular impulses and the dynamic mass-absorbing-elastic element and load moment.

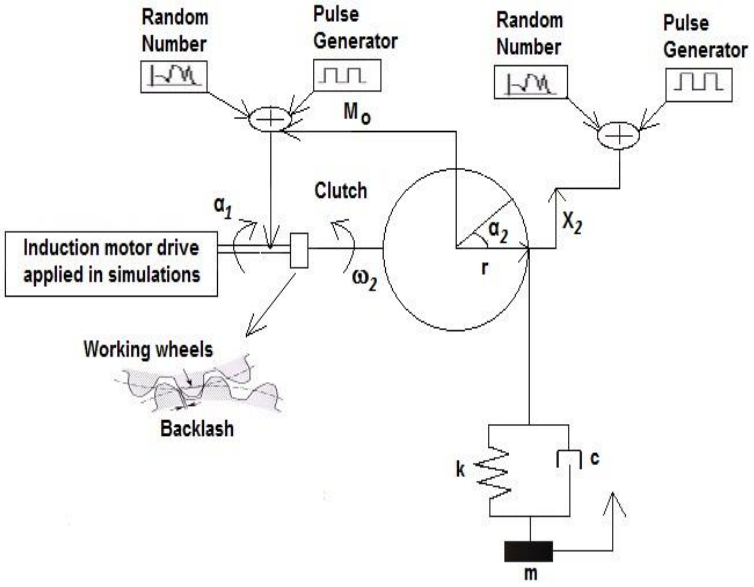

Fig.1. Diagram of a dynamic mass-absorbing-elastic element connected to the induction motor used in the simulation tests.

3. METHODOLOGY OF TESTS DEDICATED TO IDENTIFY CHANGES OF WIDTH OF BACKLASH ZONE IN THE ELECTROMECHANICAL SYSTEM GENERATING NOISES (FRICTION MODEL DESCRIBED USING OSTWALD–WAELE RELATIONSHIP)

Simulation tests of wavelet scalograms for coefficients of wavelet expansions of two physical quantities have been conducted: linear acceleration on the circuit of the drive wheel of motor's rotor a_1 and linear acceleration of lifted mass of dynamic mass–absorbing–elastic element a_2. The results of simulation tests for each physical quantities were written in the matrix M_1.

Additionally, generator of rectangle impulses and generator of normally distributed signals were connected to the electromechanical system

Sampling time of normally distributed random numbers generator was equal to $2*10^{-4}$ [s]. Sampling time of impulses in generator of rectangle impulses was equal $2*10^{-3}$ [s]. For both generators variance has been assumed equal 0.05 and mean value equal 0.

Simulation tests have been executed in the following way.

For all measurements of faults (in four groups of tests) the same values were provided: the elasticity coefficient k =100 [N/m], radius r = 0.15 [m], mass m = 10 [kg] and surface area of absorber's cylinder S_1 = 0.00311565 [m^2]. In the next groups of tests the following values of consistency coefficient

(apparent viscosity) was assumed: η_k = 0.0125 [Pa·s^{n_1}], η_k = 0.025 [Pa·s^{n_1}], η_k = 0.0375 [Pa·s^{n_1}] and η_k = 0.05 [Pa·s^{n_1}]. During the measurements in each group of simulation tests the creep index n_1 for a pseudo-plastic liquid was assumed amounting to: 0.89, 0.91, 0.93, 0.95 and 0.97. The values of the creep index n_1 in the tests were written in matrix K_1 in the ascending order.

In each test in which the value of the creep index n_1 was changing, one conducted simulation tests, the width of backlash zone amounted to: 0.0025, 0.00375, 0.005,0.0075, 0.009 and 0.01, in the above order.

In matrix K_2, widths of backlash zone applied in the conducted simulation tests were written in an ascending order.

In all the executed simulation tests, the principle was followed according to which the process of testing the electromechanical system dynamics within the backlash zone begins in the moment when the expression in the left part of the following inequality (1) is smaller than the expression in its right part:

$$\left|\alpha_1 - \alpha_2\right| < \frac{K_{2(i)}}{r}; \quad i = 1,2..6 \tag{1}$$

where: r – radius of the drive wheel of a working machine [m],
K_2 – value that has been taken sequentially from the matrix and corresponding to the assumed width of backlash zone in mechanical connection,
i – an index number of matrix's column K_2.

Values of angles have been calculated according to the following formulas:

$$\alpha_1 = \int \omega_1 \cdot dt \tag{2}$$

$$\alpha_2 = \int \omega_2 \cdot dt \tag{3}$$

After meeting the condition determined by inequality described by the formula (1) load torque of the dynamic mass-absorbing-elastic element is zeroed.

Matrix M_1 contained 2048 sequentially chosen samples recorded after the occurrence of backlash zone in the tested model. This experiment has been carried out for every measurement of both tested physical quantities. Fig. 2 presents the example of collecting samples in backlash zone against the linear acceleration of the mass of the dynamic mass-absorbing-elastic element.

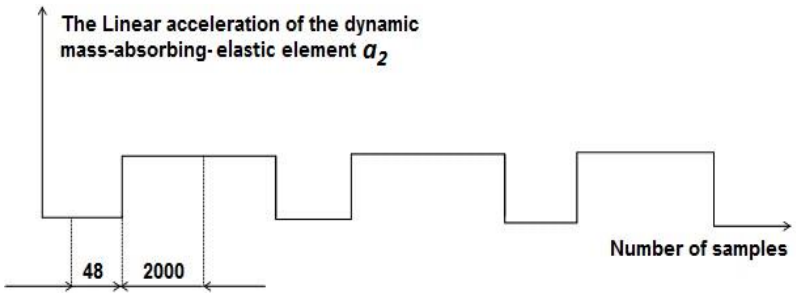

Fig. 2. Testing of dynamics in the backlash zone by means of the choice of time range

Removal of noises from all the variables placed in matrix M_1 has been executed by means of removal of relevant details of a given wavelet for the particular variable. Subsequently, the registered samples have been written in matrix M_2.

A signal obtained after the removal of details has been written sequentially for the appropriate value of the creep index n_1 in matrix M_2 whose dimension amounted to 6x2048.

The wavelet type and order have been selected so that the shape of the basic wavelet would be approximately adequate to the character of the transient course of the physical quantity obrtained in the test for the smallest backlash value.

After conducting the tests for the respective variables, the following wavelets were chosen:

a) a_1 – wavelet function symlet of the order 5,

b) a_2 – wavelet function symlet of the order 5.

On the basis of calculations conducted for the generator of normally distributed random numbers, decomposition (number of detail) level has been determined amounting to 2 whereas for the generator of rectangle impulses decomposition level (number of detail) was equal to 6.

After the tests of hard and soft elimination for the analyzing wave, the noises placed in matrix M_1 were removed, which consisted, e.g. in the removal of details whose frequencies were similar to the frequencies of the disturbing impulses.

By means of the calculated decomposition levels for both generators used in the simulation tests and as a result of observation of frequencies of disrupting impulses, the following numbers of details for variables a_1 and a_2 were deleted: 1, 2, 3, 4, 5, 6.

4. DESCRIPTION OF PROCESSING OF TWO-LAYER NEURAL NETWORK OF TYPE *COUNTER-PROPAGATION*

In the conducted simulation tests, two-layer neural network was learnt without the supervision for the values of matrix M_2.

The first layer of this network is named Kohonen's layer and represents a set of exemplary pairs of input signals of neural network X_1 and values of weights W_1. In the second layer of this network, named Grossberg's layer, X_2 values represent input signals while W_2 values represent the given set of the exemplary weights.

The input signals of neural network in the first layer represent Kohonen's layer were calculated in matrix X_1 according to the formula:

$$X_{1(j)} = M_{2(p)(j)};\ j = 1,2...2048;\ p \in \langle 1,6 \rangle \tag{9}$$

where: $M_{2(p)(j)}$ – values of matrix for the tested width of backlash zone and registered for the applied values of the creep index n_1,
p – number of column in K_1 matrix.

*Ac*cording to the assumptions typical of *Counter-Propagation* networks, values of input signals must be normalised to fulfill the condition (Tadeusiewicz, 1993):

$$X_1^{\ T} * X_1 = 1 \tag{11}$$

Normalization of input signals X_1 is conducted according to the formula:

$$X_{1(i)} = \frac{X_{1(i)}}{\sum_{j=1}^{2048} \left(X_{1(j)}\right)^2};\quad i = 1,2...2048 \tag{12}$$

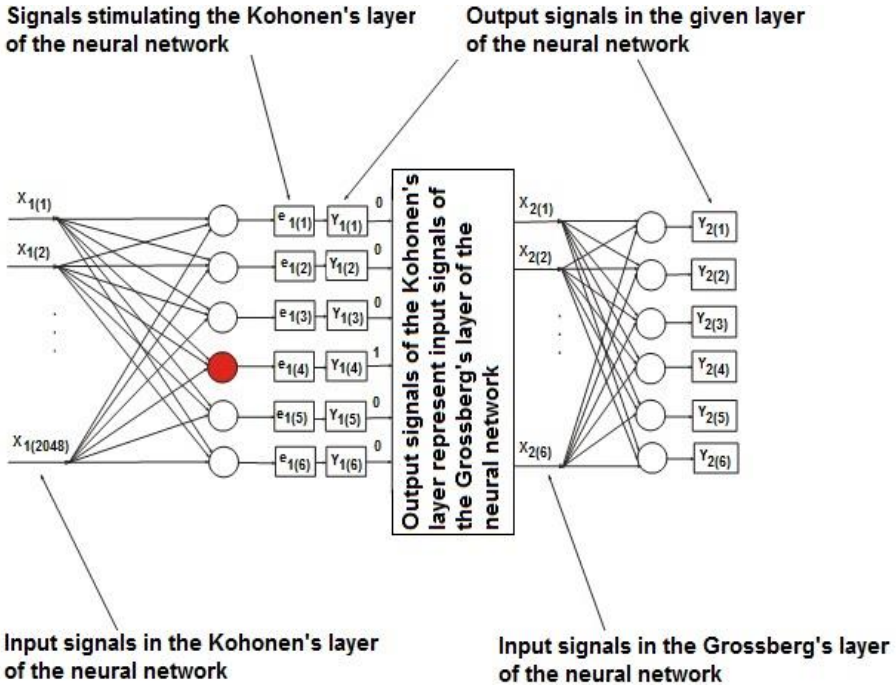

Fig. 3. Diagram of the used neural network dedicated to identification of width of backlash zone in the electromechanical system generating noises with viscous and fluid friction described by means of Ostwald–de Waele power equation (the circles represent neurons in neural network; the red neuron in the Kohonen's layer indicates a random 'winner neuron')

The aim of normalization of input signals carried out in the simulation tests was to ensure the appropriate adaptation of values of weights W_1 during the processing of neural network.

The selection of number of epochs in the discussed model of two-layer neural network was determined experimentally on the basis of observation of the obtained values of W_2 weights.

The initial values of weights in the first layer W_1 were determined according to the formula:

$$W_{1(i,\,j)} = M_{3(i)};\ i = 1,2...6;\ j = 1,2\ldots2048 \tag{13}$$

where: M_3 – calculated arithmetic means of matrix M_2 rows for the value of the creep index n_1 amounting to 0.89 for the group of simulation tests in which the consistency coefficient η_k was equal to 0.025.

Arithmetic means of matrix M_2 rows were calculated in matrix M_3 according to the formula:

$$M_{3(i)} = \frac{\sum_{j=1}^{2048} M_{2(p)(i,j)}}{2048}; \qquad i = 1,2...6;\ p = 1 \tag{14}$$

where: M_2 – values of matrix recorded in the simulation tests at the value of the creep index n_1 equal to 0.89 for the group of tests in which the consistency coefficient η_k was equal to 0.025; p =1 stands for the number of column in the matrix K_1.

Multiplication of input signals X_1 by values of weights W_1 of the particular neurons of the first layer of neural network results in the calculation of signals E_1 according to the formula:

$$E_{1(i)} = \sum_{j=1}^{2048} X_{1(j)} * W_{1(i,j)}; \quad i = 1,2...6 \tag{15}$$

Signals E_1 represent total stimulations of all neurons in the first layer from which the so-called "winner neuron" is selected having maximum value of the stimulation determined on basis of the following formula:

$$E_{1(t)} = \max(E_{1(i)});\ i = 1,2...6;\ t \in <1,6> \tag{16}$$

Therefore, only for the victorious neuron output signal Y_1 has the value equal to 1 according to the equation:

$$Y_{1(i)} = \begin{cases} 1; & i = t \\ 0; & i \neq t \end{cases} \quad i = 1,2...6;\ t \in <1,6> \tag{17}$$

The correction of weights is a result of product of difference between the input value X_1 and weight associated with:

- input value W_1,
- experimentally selected learning coefficient of neural network l_1 whose the range varies from 0 to 1.

The correction of weights for the "winner neuron" in the first layer of neural network W_1 was calculated by means of the well-known WTA (*Winner Takes All*) rule according to the formula (Osowski, 1996):

$$W_{1(t,j)} = W_{1(t,j)} + l_1 \cdot (X_{1(t,j)} - W_{1(t,j)});\ j = 1,2...2048;\ t \in <1,6> \tag{18}$$

Input signals of the second layer of neural network representing Grossberg's layer were calculated in matrix X_2 according to the formula:

$$X_{2(j)} = Y_{1(i)};\ i = 1,2...6;\ j = 1,2...6; \tag{19}$$

The applied initial values of matrix weights of the second layer in matrix W_2 were the same for the particular neurons placed in this layer and determined on the basis of the formula:

$$W_{2(i,j)} = m_1;\ i = 1,2...6;\ j = 1,2...6; \tag{20}$$

where: m_1 – arithmetic mean of matrix M_3 determined on the basis of the formula (14).

The arithmetic mean m_1 was calculated according to the formula:

$$m_1 = \frac{\sum_{i=1}^{6} M_{3(i)}}{6} \tag{21}$$

The advantage of processing of this neural network is the possibility of setting relatively small number of epochs and therefore obtaining the values of weights W_2 in the second layer appropriate for diagnostic purposes.

The output signals of the second layer of neural network Y_2 were calculated according to the formula:

$$Y_{2(i)} = \sum_{j=1}^{6} X_{2(j)} * W_{2(i,j)};\quad i = 1,2...6 \tag{22}$$

Correction of weights results from the product of difference between:

- output value Y_2,
- the set value Z_1 associated with: Y_2, input value X_2, and experimentally selected learning coefficient of neural network l_2 whose range varies from 0 to 1.

Correction weights in this layer of neural network W_2 was calculated on the basis of well-known "outstar" rule (Osowski, 1996) according to the formula:

$$W_{2(i,j)} = W_{2(i,j)} + l_2 \cdot (Y_{2(i)} - Z_{1(i)}) \cdot X_{2(j)};\ i = 1,2...6;\ j = 1,2...6 \tag{23}$$

where: Z_1 – matrix of set values of neural network.

The applied set values Z_1 were the same for all the neurons in Grossberg's layer and calculated according to the formula:

$$Z_{1(i)} = m_2;\ i = 1,2...6 \tag{24}$$

where: m_2 – the arithmetic mean of values of matrix M_2 for the tested width of backlash zone at the given value of creep index n_1.

The arithmetic mean m_2 was calculated according to the formula:

$$m_2 = \frac{\sum_{j=1}^{2048} M_{2(p)(j)}}{2048};\qquad j = 1,2...2048; p \in <1,6> \tag{25}$$

To identify the width of backlash zone for all the tested physical quantities one had to create pattern vectors W_5 as well as tested vectors W_6. These vectors were registered during the simulation tests for the assumed value of creep index n_1. Pattern vectors W_5 were created for the group of simulation tests conducted at the consistency coefficient η_k = 0.025.

To determine both pattern vectors and tested vector, maximum value T_2 was determined on the basis of all the weights placed in the Grossberg's layer W_2 subjected to the following formula:

$$T_2 = \max(W_{2(i,j)});\quad i = 1,2...6;\ j = 1,2...6 \tag{26}$$

The decicion to select maximum value among all weights W_2 instead of minimum weight, necessary in determining pattern vectors and tested vector, proved to be a correct strategy.

On the basis of maximum value T_2 matrix W_3 was determined according to the formula:

$$W_{3(j)} = W_{2(d,j)};\ T_2 \in W_{2(d,j)};\ d \in <1,6>;\ j = 1,2...6 \tag{27}$$

During the simulation tests dedicated to identify width of backlash zone at the changing value of consistency coefficient η_k, different "winner neurons" in the Kohonen's layer were selected. This fact resulted from the use of different values of input signals X_1 for this same value of the creep index n_1, which caused disruption of order of the obtained values weights W_2 in the Grossberg's layer.

Sorting the values of matrix W_3 instead of assuming different numbers of input signals X_1 is a factor responsible for a significant versatility of the presented diagnostic method due to the possibility of using this procedure in the case of more physical quantities to test.

In the presented diagnostic method matrix W_3 was sorted in the descending order. After sorting, the value of this matrix was written in matrix W_4 according to the following formula:

$$W_4 = \left[W_{3(1)} \geq W_{3(2)} \geq W_{3(3)} \geq .... \geq W_{3(6)} \right] \tag{28}$$

Pattern vectors W_5 as well as tested vector W_6, used in identification of width of backlash zone, were determined, respectively, according to the formula:

$$W_{5(i,j)} = W_{4(i,j)};\ i = 1,2...6;\ j = 1,2...6 \tag{29}$$

$$W_{6(j)} = W_{4(j)};\ j = 1,2...6 \tag{30}$$

Identification of the assumed width of backlash zone at the assumed value of the creep index n_1 is possible owing to the calculation of values of matrix B according to the formula:

$$B_{(i)} = \sum_{j=1}^{6} \left| W_{6(j)} - W_{5(i,j)} \right|; \quad i = 1,2...6 \tag{31}$$

Determination of minimal value of matrix B causes determination of index nr_1 according to the following formula:

$$B_{(nr_1)} = \min\left(B_{(d)}\right);\ d = 1,2...6; nr_1 \in <1,6> \tag{32}$$

Calculations concerning the value of index nr_1 are necessary for the correct identification of width of backlash zone determined on the basis of the particular number of column K_2 referring to this index according to the formula:

$$i = nr_1 \tag{33}$$

where: i – number of column im matrix K_2.

5. RESULTS OF SIMULATION FOR DIAGNOSTIC ALGORHYTHM DEDIASTED TO IDENTIFICATION OF WIDTH OF BACKLASH ZONE IN ELECTROMECHANICAL SYSTEM GENERATING NOISES

In the tables below, the column labeled *Test parameters* contains widths of backlash zone identified in the tests. However, the column *Results* comprises the bolded final results of calculations of values of matrix *B*.

Tab. 1. Selected values of matrix *B* for linear acceleration of the induction motor a_1 in electro-mechanical system containing friction described by means of Ostwald – de Waele power equation

Test parameters	Results	Test parameters	Results
backlash zone = 0.009, consistency coefficient η_k = 0.0125, creep index n_1 = 0.91, epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	1.5895 0.9248 0.2620 0.4710 **0.0025** 0.8902	backlash zone = 0.009, consistency coefficient η_k = 0.0125, creep index n_1 = 0.91, epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.01	0.0859 0.0500 0.0142 0.0255 **0.0001** 0.0481
backlash zone = 0.0025, consistency coefficient η_k = 0.05, creep index n_1 = 0.95, epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	**0.2949** 0.9600 1.6225 2.3551 1.8860 0.9922	backlash zone = 0.0025, consistency coefficient η_k = 0.05, creep index n_1 = 0.95, epochs = 30, learning coefficient of neural network l_1 = 0.9, learning coefficient of neural network l_2 = 0.01	**0.0165** 0.0538 0.0910 0.1321 0.1058 0.0557
backlash zone = 0.005, consistency coefficient η_k = 0.0375, creep index n_1 = 0.97, epochs = 20, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	0.5442 0.2718 **0.0003** 0.2998 0.1076 0.2585	backlash zone = 0.005, consistency coefficient η_k = 0.0375, creep index n_1 = 0.97, epochs = 40, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	3.18481 0.5904 **0.0019** 1.7545 0.6297 1.5130

Tab. 2. Selected calculated values of matrix *B* for linear acceleration of a mass a_2 in electro-mechanical system containing friction described by means of Ostwald-de Waele power equation

Test parameters	Results	Test parameters	Results
backlash zone = 0.009, consistency coefficient η_k = 0.0125, creep index n_1 = 0.93, number of epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	0.6245 0.4992 0.3698 0.1376 **0.0002** 0.0868	backlash zone = 0.009, consistency coefficient η_k = 0.0125, creep index n_1 = 0.93, number of epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.01	0.0517 0.0413 0.0306 0.0114 **$1.5565 \cdot 10^{-5}$** 0.0072
backlash zone = 0.00375, consistency coefficient η_k = 0.05, creep index n_1 = 0.89, number of epochs = 30, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	0.1256 **0.0003** 0.1292 0.3613 0.4988 0.5857	backlash zone = 0.00375, consistency coefficient η_k = 0.05, creep index n_1 = 0.89, number of epochs = 30, learning coefficient of neural network l_1 = 0.9, learning coefficient of neural network l_2 = 0.01	0.2931 **0.0006** 0.3016 0.8434 1.1643 1.3672
backlash zone = 0.0025, consistency coefficient η_k = 0.0375, creep index n_1 = 0.91, number of epochs = 20, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	**0.0001** 0.0761 0.1550 0.2963 0.3800 0.4330	backlash zone = 0.0025, consistency coefficient η_k = 0.0375, creep index n_1 = 0.91, number of epochs= 40, learning coefficient of neural network l_1 = 0.1, learning coefficient of neural network l_2 = 0.09	**0.0003** 0.1799 0.3663 0.7003 0.8980 1.0231

Bolded values of matrix *B* in the presented tables are correct results obtained finally in the process of identification of the fault number.

Pattern vectors W_5 were registered in the simulation tests in which one applied the learning coefficient of Kohonen's layer l_1, the learning coefficient of Grossberg's layer l_2, and the number of epochs of processing of neural network consistent to the assumed values of these variables placed in the above presented tables.

For both tested physical quantities pattern vectors W_5 were created in the group of simulation tests for which value of the assumed consistency coefficient η_k was 0.025.

In the conclusion, the number of epochs of processing of neural network was fixed to 30 because it ensured the greatest selection of values of elements of matrix B in comparison with the results of calculations for the number of epochs between 20 and 40.

On the basis of the presented results, one can notice that for both tested physical quantities, in the experiment consting in linear acceleration of the induction motor a_1 and also linear acceleration of mass a_2, values of matrix B gradually deteriorate for the learning coefficient of Kohonen's layer l_1 decresing in the range from 0.1 to 0.9 and for the learning coefficient of Grossberg's layer l_2 increasing in the range from 0.09 to 0.01.

6. CONCLUSIONS

Using of time-frequency methods with multistage signal decomposition and also two-layer neural network processed without supervised learning was applied to monitor electromechanical system being in the backlash zone that included mass-absorbing-elastic load and where noises generated by means of generator of Gauss and generator of rectangle impulses (at the sampling frequency 50 kHz) brought disturbances to the received signals.

Distributions of coefficients of wavelet expansion of state variables that describe the tested physical quantities and the obtained values of weights of the second layer of the processed neural network – used for the linear acceleration on circuit of the drive wheel of motor A_1 and linear acceleration A_2 dynamic mass-absorbing element – enable obtaining of the correct results of identification of width of backlash zone. The simulation tests proved that the blur of spectrum, difficulties in obtaining small deviations from the state considered as correct and its nonlinear deformation may result from the inappropriate selection of the base wavelet.

It should be carefully chosen taking into accunt the character of the tested course on the basis of the:

- selection of the central frequency of signal associated with fault,
- frequency determined on the basis of sampling time of the generator of normally distributed random numbers,
- sampling time of generator of rectangle impulses.

The obtainment of better shapes of de-noised signals and consequent obtainment of correct final results of identification of width of backlash zone for the model of complex electromechanical system (usually described by nonlinear characteristics of elements) was possible owing to:

- the additional removal of some physical quantities of the tested electromechanical system apart from the calculated decomposition levels (details) of the assumed wavelet,
- determination of variance and mean value of random signals introduced by means of Gaussian generator and generator of rectangle impulses.

REFERENCES

Aktas, M., & Turkmenoglu, V. (2010). Wavelet-based switching faults detection in direct torque control induction motor drives. *Science, Measurement & Technology, IET*, *4*(6), 303–310.

Balara, D., Timko, J., Źilkova, J., & Leśo, D. (2017). Neural networks application for mechanical parameters identification of asynchronous motor. *Neural Network World*, *3*, 259–270.

Chebil, J., Noel, G., Mesbah, M., & Derihe, M. (2009). Wavelet Decomposition for the Detection and Diagnosis of Faults in Rolling Element Bearings. *Jordan Journal of Mechanical and Industrial Engineering*, *3*(4), 260–267.

Doniec, R. (2010). *Wykorzystanie metod sztucznej inteligencji do regulacji poziomu insuliny w organiźmie człowieka* (doctoral dissertation). Politechnika Śląska, Gliwice.

Duch, W., Korbicz, J., Rutkowski, L., & Tadeusiewicz, R. (2000). *Biocybernetyka i inżynieria biomedyczna 2000. Sieci neuronowe.* Tom 6. Warszawa: Akademicka Oficyna Wydawnicza EXIT.

Duda, J. T. (2007). *Pozyskiwanie wzorców diagnostycznych w komputerowych analizach sprawności urządzeń, Diagnostyka procesów i systemów* (pp. 1–16). Warszawa: Akademicka Oficyna Wydawnicza EXIT.

Fuente, M. J., & Saludes, S. (2000). Fault detection and isolation in a non-linear plant via neural networks. *IFAC Proceedings Volumes*, *33*(11), 463–468.

Granda, D., Aguilar, W. G., Arcos-Aviles, D., & Sotomayor, D., (2017). Broken bar diagnosis for squirrel cage induction motors using frequency analysis based on MCSA and continous wavelet transform. *Mathematical and Computational Applications*, *22*(2), 30. doi:10.3390/mca22020030

Korbicz, J., Kościelny, J. M., & Kowalczuk, Z. (2002). *Diagnostyka procesów. Modele. Metody sztucznej inteligencji*. *Zastosowania*. Warszawa: WNT.

Kowalski, Cz. (2003). Stan obecny i tendencje rozwojowe metod monitorowania i diagnostyki napędów z silnikami indukcyjnymi. *Wiadomości Elektrotechniczne*, *4*, 160–164.

Łobos, T., Leonowicz, Z., Rezmer, J., & Schegner, P. (2006). High resolution spectrum-estimation methods for signal analysis in power systems. *IEEE Trans. Instrum. Measur*., *55*(1), 219–225.

Osowski, S. (1996). *Sieci neuronowe – w ujęciu algorytmicznym.* Warszawa: WNT.

Tadeusiewicz, R. (1993). *Sieci neuronowe*. Warszawa: Akademicka Oficyna Wydawnicza.

Wysogląd, B. (2003). Metody diagnozowania łożysk tocznych z zastosowaniem transformacji falkowej. *Diagnostyka*, *29*, 47–52.

Zając, M. (2009). *Metody falkowe w monitoringu i diagnostyce układów elektromechanicznych.* Monografia 371. Kraków: Politechnika Krakowska.

Zhang, J. W., Zhu, N., Yang, L., Yao, Q., & Lu, Q. (2007). A fault diagnosis approach for broken rotor bars based on EMD and envelope analysis. *Journal of China University Mining & Technology*, *17*(2), 205–209.