REVISTIA

5th International Conference on Engineering and Formal Sciences

Brussels, 24-25 January 2020

Venue

Universite Libres de Bruxelles

CONFERENCE PROCEEDINGS

ISBN 978-164786089-9

REVISTIA

PUBLISHING AND RESEARCH

5th International Conference on Engineering and Formal Sciences

24-25 January 2020

Bruxelles

Proceedings Book

ISBN 978-1-64786-089-9

Venue

Universite Libres de Bruxelles

Every reasonable effort has been made to ensure that the material in this book is true, correct, complete, and appropriate at the time of writing. Nevertheless, the publishers, the editors and the authors do not accept responsibility for any omission or error, or for any injury, damage, loss, or financial consequences arising from the use of the book. The views expressed by contributors do not necessarily reflect those of Revistia.

Typeset by Revistia

Printed in Brussels

Copyright © Revistia. All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the publisher or author, except in the case of a reviewer, who may quote brief passages embodied in critical articles or in a review.

Address: 11, Portland Road, London, SE25 4UF, United Kingdom

Tel: +44 2080680407

E-Mail: office@revistia.com

International Editorial and Scientific Advisory Board

Javier Cachón Zagalaz, PhD - Universidad de Jaén, Spain
Sevim Yilmaz, PhD - Pamukkale University, Denizli Turkey
Bartosz Kaźmierczak, PhD - Poznań University of Technology, Poland
Warda Sada Gerges, PhD - Kaye College of Education, Israel
Enkhtuya Dandar - University of Science and Technology, Mongolia
Balazs Hohmann, PhD - University of Pécs, Hungary
Gani Pllana, PhD - Faculty of Mechanical Engineering, University of "Hasan Prishtina", Kosovo
Irina Golitsyna, PhD - Kazan (Volga) Federal University, Russia
Jean d'Amour - Åbo Akademi University, Finland
Ornela Bilali, PhD - "Aleksander Xhuvani" University, Albania
Felice Corona, PhD - University of Salerno, Italy
Ahmet Ecirli, PhD, Institute of Sociology, Romanian Academy
Carmen Cecilia Espinoza Melo, PhD - Universidad Católica de la Santísima Concepción in Chile
Célia Taborda Silva, PhD - Universidade Lusófona do Porto, Portugal
Khaled Salah, PhD - Faculty of Education - Alexandria University, Egypt
Panduranga Charanbailu Bhatta , PhD - Samanvaya Academy for Excellence, India
Kristinka Ovesni, PhD - University of Belgrade, Serbia
Sajitha Lakmali Hewapathirana, PhD - University of Colombo, Sri Lanka
Amel Alić, PhD - University of Zenica, Bosnia and Herzegovina
Mohammed O. Ibrahim, PhD - University of Ilorin, Nigeria
Nadia Jaber - Palestinian Ministry of Education & Higher Education
Rodica Sirbu, PhD - Ovidius University, Constanta, Romania
Vania Ivanova, PhD - University of National and World Economy, Bulgaria
Davit Narmania, PhD - Tbilisi State University
Ana Paula Marques, PhD - University of Minho, Portugal
Smaragda Papadopoulou, PhD - Universityof Ioannina - Greece
Syed Zafar Abbas , PhD - Aliz Educational Institutions, Pakistan
Helena Neves Almeida, PhD - University of Coimbra, Portugal

Mihaela Voinea, PhD - Transilvania University of Brasov, Romania

Vereno Brugiatelli, PhD - University of Verona, Italy

Tonia De Giuseppe, PhD - University of Salerno, Italy

Dominika Pazder, PhD - Poznań University of Technology, Poland

TABLE OF CONTENTS

Determining the Probability of Cyberattacks

Pavel Yermalovich

Ph.D. student, Faculté des Sciences et de Génie, Université Laval, Quebec City, Canada

Abstract

The use of information is inextricably linked with its security. The presence of vulnerabilities enables a third party to breach the security of information. Threat modeling helps to identify those infrastructures, which would be most likely exposed to cyberattacks. In some cases, however, threat modeling can not be classified as sufficient method of protection. This paper entitled “Determining the probability of cyberattacks” presents an analysis of different techniques with an attempt to identify the most informative parameters and cyberattack prediction markers, which would lay the foundation for the development of cyberattack probability functions. Next, it would be relevant to design such cyberattack probability functions, which would be used upon the initial identification of a cyberattack. The findings of this research could be applied during the future assessment of risk levels of information systems to ensure more effective information security management.

Keywords: Probability of cyberattacks, information security, cyber security, risk management, risk prediction.

1. Introduction

1.1. Motivation

The information systems are currently monitored by various systems. Such checks (audits) help to obtain various system characteristics within the perimeter of the information system security. On the basis of such verification or data monitoring it is possible to assess a risk level of cyberattack within the protected perimeter (infrastructure, system). The risk assessment is limited to the calculation of its level of risk at a certain time point, such as a freeze frame in a movie. The full risk analysis encompasses several stages. The implementation of each stage takes time. One or several indicators might be radically modified in the process of calculation of values in the end of one or more stages of a risk level analysis. Hypothetically, the risk indicator can be altered significantly as a result of this process and it may even exceed the maximum allowed level. This deviation can not be promptly tracked, while the analysis takes much time. For example, the conduct of risk assessment with MEHARI (MEthod for Harmonized Analysis of RIsk) Expert tool [18] may take more than six months [12]. From this it follows that in the context of such a risk assessment model, there are periods that remain uncontrolled.

The system threat modelling allows to obtain a probabilistic image of a cyberattack plan. At the same time, we can not predict the period of a cyberattack initiation. The periodic scanning of information system for known vulnerabilities helps to identify a list of the system vulnerabilities. However, this list can not ensure an accurate risk assessment in case of all these established vulnerabilities. Thus, for SIEM (Security Information and Event Management), it is important to ensure an obtainment of such a list arranged according to the importance of primary actions and reactions. Additionally, it is important to ensure proper classification of primary responses based on the analysis of the data from this list. First, it would be necessary to use the results of vulnerability assessment covering the most important assets in order to ensure their protection against identified critical vulnerabilities.

To date, there are training developments for Artificial Intelligence (AI) that are formed through the analysis of traffic logs to identify outliers. With this approach, it is possible to identify a cyberattack with a certain probability in a real time. The difference between Intrusion Detection System (IDS) and AI is that the AI learns without analyzing deeply the cyberattack signature.

This research aims at developing a cyberattack prediction system based on various system parameters. Today it is impossible to determine precisely the time point at which the planned cyberattack will be committed and which vector will be chosen. This confirms the relevance of "prediction of cyberattacks" to be able to identify the levels prone to risks at every moment. Thus, it is proposed to extend risk prediction to all the existing data (risk indicators history).

1.2. Our Contributions

Identification of probability of an attack on information system S between t and t+Δt;

Development of attack forecasting function entitled "Oracle".

For example, the system configuration change, system modification allocation of funds for the system protection or creation of a working schedule.

Prior to applying X-parameters, it is suggested to check:

External threats:

Threats in social networks messages or sent emails;

Text in backlinks;

Text in backlinks to the attack target.

Trapping:

Fail2ban[1], IPS, etc.;

Pages of nonexistent administrators (as reconnaissance component);

One has also to consider the creation of markers that can influence the attack probability.

Experience, knowledge, tactics of the attacker side.

2. Literature review

A number of systems, such as Intrusion Detection System, can detect attacks [3] in real time. The detection of attacks is based on the application of rules (attack signatures) and work statistics under normal conditions (without attacks). The attack detection platforms applying machine learning methodology are based on the following three components:

Statistical analysis;

Attack signature analysis, covering the existing or new signatures created by an information security analyst and based on various sources. This analysis enables the identification of known behaviors or known attacks;

Machine learning to identify outliers.

The PatternEX (Threat Prediction Platform) [19] is an example of a system based on the automatic learning method [4]. The artificial Intelligence combines analysts' intuition with machine learning to mimic a security analyst to predict real-time and large-scale threats. To ensure an application of Artifiial Intelligence in InfoSec, PatternEx has developed a patent-pending technology entitled Active Contextual Modeling[2], or ACM. This technology continually identifies new and evolving (active) threats with the help of (contextual) analyst, and, once identified, synthesizes new models (modeling) that can distinguish between malicious and benign models. Another way to recognize threats is heuristic scanning[3], which is a method used by antivirus software to detect new viruses, as well as new variants of the already known viruses. Heuristic analysis is a method based on the supposed behavior of a program to determine whether the program is a virus or not. This method differs from statistical analysis, which is based on comparisons of the program with known viruses referenced in an anti-virus software library. The heuristic method can be used to detect DDoS attacks [11], Predictive Blacklisting, Phishing Attacks [9] and Malicious Web Pages [10]. As can be seen from the examination of attack prediction systems, this direction of research is a new one, and as it usually happens with new systems, it should be improved and refined to ensure good results.

3. Formal consideration

3.1. Problem

This research is an attempt to propose a risk level forecasting method. The Formula for establishing a risk value is

[1] http://bit.ly/2XblTyu
[2] http://bit.ly/2P6JOwd
[3] http://bit.ly/2X9nJ2U

$$R = P \cdot I \quad (1)$$

where R is the risk value, P is the probability of an attack event and I is the impact (a likely consequence) of an attack event.

To determine the risk, we can predict only the probability P, while the impact I is considered by us as a constant value. However, the valuation of assets also plays a significant role in determining the future impact, due to changes in asset prices.

Given:

parameters of information system S:

set of business processes;

set of assets;

set of protection techniques applied to ensure the safety of assets;

security policy;

set of of this system log files;

risk assessment methodology;

attack forecasting time frame (t;t+Δt).

Find:

probability that an attack take place on information system S between t and t+Δt;

future risk level allowing to assess and identify the budget required to maintain an acceptable risk level.

The idea of the research aimed at the development of this field could be summarized as follows:

Development of attack forecasting function entitled "Oracle". Determination of parameters for attack forecasting function 2 "Oracle":

$$Oracle: X \rightarrow [0,1] \quad (2)$$

Let us consider the research project's general scheme in Figure 1.

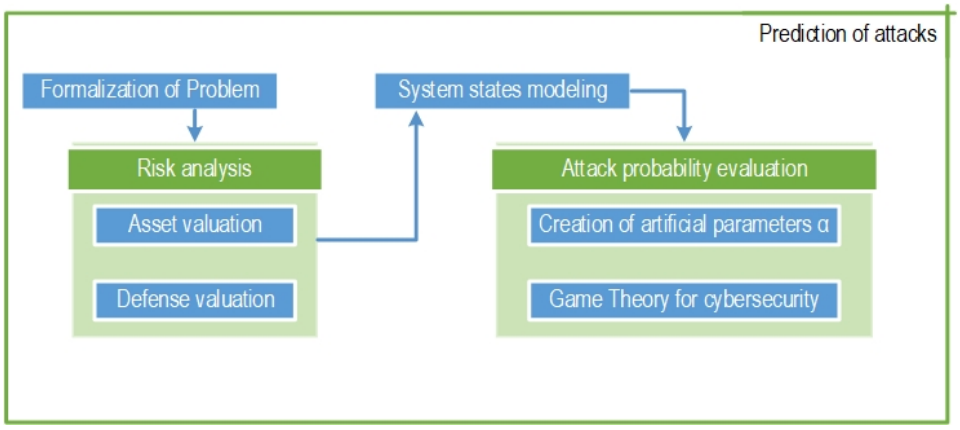


Figure 1: General scheme of the research project.

3.2. Formalization

The following should be noted for proper consideration of this formalization:

Calligraphic letter stands for a domain, for example A;

Capital letter - denotes a subset, for example A;

Small letter stands for - a subset element, for example a.

Based on the above system input parameters, we can represent the initial state of the system S=(A,D). In the output stage we get the modified system state S'=(A,D'), which satisfies the following requirements:

RM(A,D') represents acceptable risk level for our system.

$D' = \min_{x \varepsilon D} Cost(D, Conf, x)$ stands for the selection of a configuration of a minimum cost that meets the requirements for an acceptable risk level.

Based on the input data, we are proceeding to formalizing an attack forecasting function in "Oracle" 3.

$$Oracle: BP, A, D, LOG, SP_r, RM, (t; t + \Delta t) \rightarrow [0,1] \quad (3)$$

Where

Attack - denotes an actual occurrence of an adverse event.

$[0,1]$ is the probability value of the event (attack).

BP: represents a set of business processes. The sequence of $bp_0, bp_1, \ldots$ comprises a set of meta-variables applied by us through the entire BP. Each business process encompasses a lot of assets. For example, content management (bp_0) using assets A comprises a set of assets: Application data (data bases) (a_0), Electronic mail (E-mail) (a_1), Local Area Network services (LAN services) (a_{10}), Web editing Service (a_{27}): $bp_0 = \{a_0, a_1, a_{10}, a_{27}\}$.

A: represents a set of assets. The sequence of $a_0, a_1, \ldots$ describes a set of meta-variables used by us, which are ultimately comprising A. The examples of assets include the Application data (data bases), Electronic mail (E-mail), Local Area Network services (LAN services), Web editing Service, Digital accounting control, etc.

D: describes the defense techniques applied to ensure the safety of assets A. The sequence of $d_0, d_1, \ldots$ describes meta-variables applied through the range D. In our example, the following techniques were used to ensure protection of the operating computer: a firewall, an antivirus and logging. Each security measure could be classified based on its configuration ($Conf$) - the configuration affecting the system system functioning and performance.

LOG is classified as a sequence of lines, where each line represents an event with unique information.

SP: stands for the security policy. The security policy comprises a set of instructions (the sequence of $sp_0, sp_1, \ldots$ describes meta-variables used by us through the range SP) for implementing business processes. The security policy can be accepted and formalized “on paper” as a set of safety rules. However, there are times when the actual security policy is different from what is on paper. Therefore, we will consider two types of security policies, Real (SP_r) and Theoretical (SP_t). For example, the proper use of security measures D to ensure the safety of an acceptable level for collection of assets A. or setting minimum requirements for the configuration of the protection system. The security policy describes which ports should be open to the firewall, antivirus updates frequency, timeframes for the system antivirus scan, logging detailing and the determination of location of the undertaken security measures (firewall at the entrance, antivirus and logging inside the system).

The adopted risk assessment methodology (MEHARI, CobiT, etc.) RM is represented as a function of $RM: A \times D \times SP \rightarrow R$, that returns the level of risk R of loss of assets A when using protection components D and, consequently, disruption of business processes BP.

$(t; t + \Delta t)$ time frame for attack forecasting.

To clarify this formalization we propose to refer to the following article: *Formalization of attack prediction problem* [13].

4. Risk analysis

In order to ensure an adequate protection level, each time we need to identify what should be protected and from whom. For this, it is necessary to identify the assets and obtain the information about them. The above mentioned statement could be illustrated by the following example:

Analysis of business activities or processes:

Determining the ownership of an asset in a business process.

Asset analysis (assessing the degree of importance, the level of loss when the asset is lost, of each value).

Given: List of business activities.

Find: Intrinsic Impact table.

Audit of the protection techniques applied to ensure the safety of assets:

Defense system analysis.

Risk analysis.

Given: List of protection techniques applied to ensure the safety of assetss.

Find: Attack scenarios, risk per asset type, risk per event type.

4.1. Analysis of business activities or processes

Proceeding from the above-mentioned, it is possible to conclude that the obtainment of information about them plays a crucial role. Asset valuation is a very important step in ensuring a proper system security. Thus, we need to know which assets should be protected to maintain their confidentiality, integrity and availability. *MEHARI Expert* [17] uses the following classification of the data, classification of services and classification of compliance with laws and regulations relating.

Sometimes, it is difficult to assign monetary value to assets. This is why each of these classification elements evaluates the classification level in terms of the maximum damage.

There are different methods of risk analysis. Each approach uses its own technique for the interpretation of asset values. It is necessary to represent the valuation of assets to be able to connect and use the results of asset analysis obtained applying different methodology. For that, we could quantify the categories. This will allow us to treat them as generalized attributes.

For example, MEHARI makes assessments based on 4 points scale (from 1 (Weak) to 4 (Unbearable)). As initial approach for this research project we could apply MEHARI only. In this case, the classification level must be determined before initiating risk analysis by MEHARI Expert. This should correspond to the maximum negative consequences of malfunctioning affecting this criterion of the asset evaluation process. It's really easier to determine the assets value in the event of their loss assessing each of the following security principles: availability, integrity, confidentiality. These are three principles for evaluating the Impact of loss asset.

We can represent the Formula 1 for establishing the risk value as follows:

$$R = \sum_{i=0}^{m} \left(P_{availability}(asset_i) \cdot I_{availability}(asset_i) + \right.$$
$$+P_{integrity}(asset_i) \cdot I_{integrity}(asset_i) +$$
$$\left.+P_{confidentiality}(asset_i) \cdot I_{confidentiality}(asset_i)\right) \qquad (4)$$

where R stands for the risk value, $P_{availability}(a_i)$ is the probability of loss availability of asset a_i due to an attack event and $I_{availability}(a_i)$ stands for the impact (a likely consequence) of an attack event due to loss availability of asset a_i, and m stands for the amount of assets. the same for integrity and confidentiality for all the asset. The same relates to the integrity and confidentiality of all assets.

To determine a risk, we can predict only the probability P, while the impact I is considered by us a constant value for the time window $(t; t + \Delta t)$.

The probability of losing an asset is the probability of attacks on a given asset. Similarly, for security purposes, the likelihood of loss of availability / integrity / confidentiality for an asset P_{asset} corresponds to the likelihood of attacks $P(A)$ on the availability / integrity / confidentiality of the asset : $P_{asset} = P(A)$

In Section 5.1. "Cyberattack probability evaluation in each of the states", we take a closer look at the decomposition of the Formula for the probability of an asset attack.

The principle of reasonable sufficiency [8] states that it is fundamentally impossible [7] to create an absolutely insurmountable security system. It is important to choose the appropriate protection level for which the costs, risks and possible damages would be acceptable. Each of these steps takes time. The intervals between modifications can reach several months.

Formula 1 is still applied in MEHARI [17]. It allows to verify an average risk, but does not reflect the risk level in real time. There is a need to know the level of a probabilistic risk to track the risk levels over the specific time period.

The risk level can be calculated separately for each of the assets. The security context (security system) is created for each asset separately. Each security system consists of many components. In this way, we can design a chain of all security components (security barriers) for each asset separately. Each asset in the information system is valued differently in the event of total loss. We can rely on the indicator I (impact) which remains constant throughout time.

The elements of this table need to be expanded to take into account the different levels of probabilities of attacks and impacts. Lets assume that the average risk level for the whole company is ranked as 2 in compliance with MEHARI. This is a "Tolerated" level. However, another department of this company can be assigned an "Unbearable" risk level (4 of 4). Reasoning with the average value does not make it possible to spend the security budget properly. Thus, we want to have a dynamic system that adapts all the time, which can cost on average 25k$ per year, even if the forecast for next month is only around 500$. This can significantly change the value of a risk indicator and even allow to get out of the maximum allowed level (by MEHARI it is level 3 "Unacceptable"). This difference can not anticipated within very strict deadlines, while the analysis takes time. By doing this, as part of this risk assessment model, we will explore the so-called unchecked periods. An example of such case is shown in Figure 2. In this graph, the red line represents the limit of acceptable risk. The blue bars represent the true value of the risk and the black dots stand for the calculated risk values (calculations are carried out every four time units).

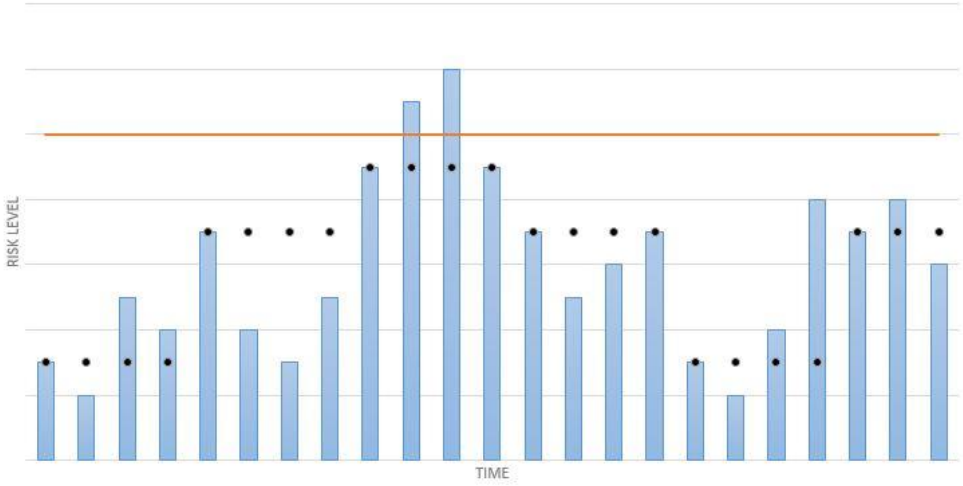


Figure 2: Graphical representation of the periods when risks take the passed calculated value.

In order to be able to represent the risk level at any time, it is suggested to analyze the already existing data sets and associated risks. For that we need to have an estimate of the appropriate level as a function of time, that is, a probability.

4.2. Audit of the protection techniques applied to ensure the assets safety

The system safety assessment approach through a security system audit is structured as questionnaire (yes / no) designed to analyze each security category (according to MEHARI Expert).

It is worth to undertake an analysis of protection measures D for assets A involved in business processes BP for the initial configuration of the system $Conf_h$, where $h = 0$ stands for the initial configuration.

$$D = (Conf, v), \tag{5}$$

where v - vulnerabilities and exposures.

The following Formula 6 is used by us to determine the level of risk for business processes:

$$RM: BP \times A \times D \rightarrow [1..4] \tag{6}$$

MEHARI uses charts containing audit results collected after the provision of "yes" or "no" answers to the questions used to collect the information enabling an analysis of the existing security systems. This analysis takes place within the time interval Δt. The risk level assessment in the time period $t_0 + \Delta t$ takes place with an application of the chosen risk assessment methodology RM. This analysis assesses the existing measures to protect the assets entering the business processes. After completing all the charts relating to the organisation's audit, the MEHARI risk analysis methodology forms a panorama of risks for each asset type (information, services, management processes). The asset protection system needs to be improved to change the risk level. It could be strengthened after the system configuration change i. e. $Conf_{(h-1)} \rightarrow Conf_h$ (where $Conf_h$ stands for the system configuration ensuring lesser risk for business processes). Each of $h \in R^+$ configurations has its own price by function $Cost: (d_{firewall}, Conf_{firewall}) \rightarrow p$, where $p \in R^+$. The task of the defense side is to minimize the risk level R to an acceptable level, with minimal costs $Cost(D, Conf_d)$.

$$\begin{cases} RM(BP \times A \times D') \rightarrow \text{acceptable level of risk} \\ D' = \min\limits_{x \varepsilon D} Cost(D, Conf, x) \end{cases} \tag{7}$$

It is suggested to use a cycle consisting of 1) analysis, 2) modeling, 3) selection and 4) application of the configuration in case of the system configuration change. The cycle is presented in Figure 3.

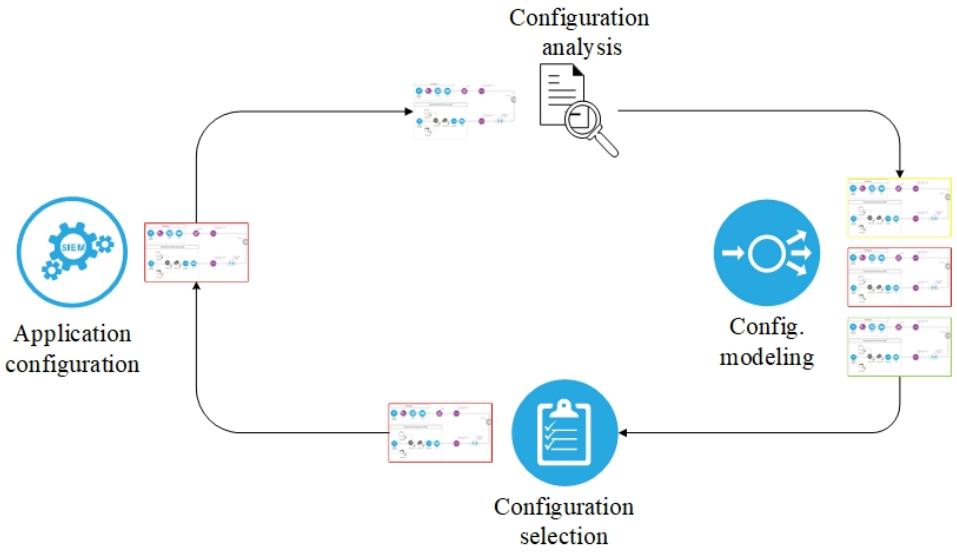


Figure 3: System configuration change cycle.

5. System states modeling

It is necessary to determine and model states of the system applying a formal method, meaning that we should decompose the system into simple components or variables (host, port, service, service version, availability, etc.). Each variable should

be determined for the initial time instant. Later, system will affect the change of variables, i.e. removal and addition of variables describing the state of the system at each time point.

The system state change takes place upon the change of its configuration, switching or connection of any component aimed at ensuring its protection.

security configuration change $D[Conf_h/Conf_{h+1}]$;

disabling the system protection component D^-;

connecting system security component D^+.

The system is modelled by a probabilistic attack graph G, which is tuple,

$G = (S; \tau; \pi; L)$ consisting of

initial or start state $s_0 \in S$;

all of the states S;

transition relationship $\tau \subseteq S \times S$;

probabilistic transition $\pi: S \to S$;

labelling of states $L: S \times S$.

The function π specifies probabilities of transitions from probabilistic states, that applies to all transitions, meaning $s_1 \to s_2 \in \tau$ such that $s_1 \in S$, thus we have $P(s_1 \to s_2) = \pi(s_1)(s_2) > 0$. In that context $\pi(s)$ can be viewed as probability distribution on next states. Intuitively, when the system is in a deterministic state s_0, we have information about the relative probabilistic state s_1. Next it will choose the next state according to probability distribution $\pi(s)$.

In the context of this work it was suggested to apply MEHARI methodology. Consequently, we have four risk levels (identified by different colors, higher intensity corresponds to the maximum risk level). An approximate draft of the system is displayed below in Figure 4.

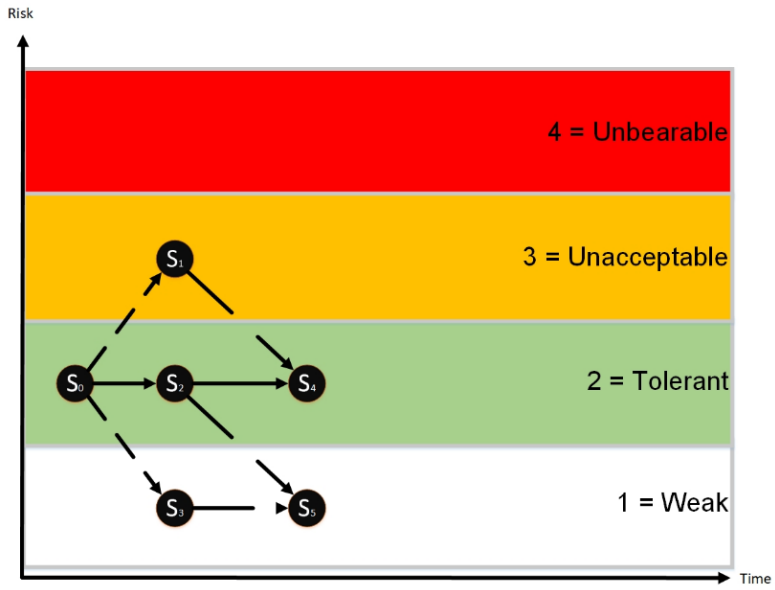


Figure 4: Graphical representation of states and risk levels.

5.1. Cyberattack probability evaluation in each of the states

We introduced Formula 1 for establishing the risk value of an asset. Let us rewrite this Formula considering the probability of losing an asset as a result of possible attack on one of the assets:

$$R_{asset}(Attacks) = \sum_{i=0}^{k} (P_{availability}(attack_i) \cdot I_{availability}(attack_i) + \\ + P_{integrity}(attack_i) \cdot I_{integrity}(attack_i) + \\ + P_{confidentiality}(attack_i) \cdot I_{confidentiality}(attack_i)) \quad (8)$$

where $R_{asset}(Attacks)$ is a risk value due to attack events, $P_{availability}(attack_i)$ is the probability of loss availability of asset $attack_i$ due to attack events and $I_{availability}(attack_i)$ is the impact (a likely consequence) of an attack event of loss availability of asset $attack_i$, and the same for integrity and confidentiality for the asset due to attack events. The number of attacks is expressed by k.

Each attack is based on exploiting the existing vulnerability using an attack vector. Attack vector [1] - is a path or route used by the intruder to gain access to the target (asset). Vulnerability [5] - is a weakness in design, implementation, operation or internal control of a process that could expose the system to adverse threats at the time of threat events.

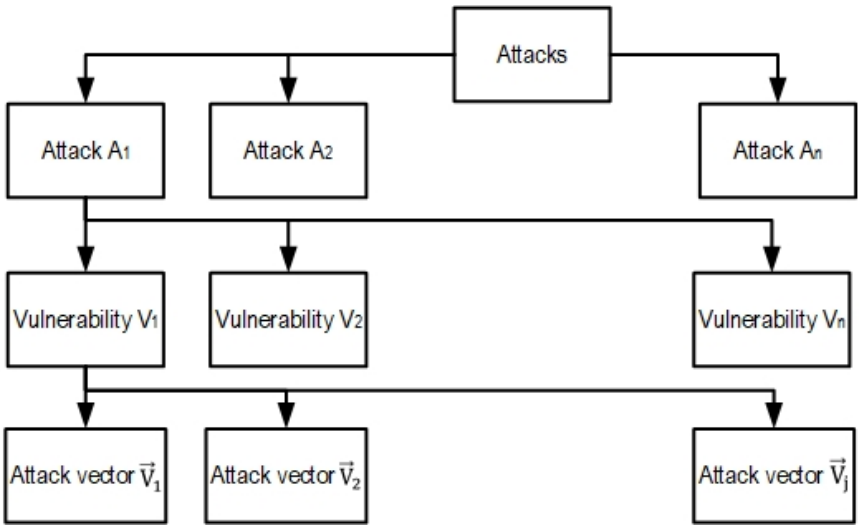


Figure 5: Attack consists of exploiting an existing vulnerability using an attack vector.

We can rewrite the likelihood of an attack (a_i) of all the attacks A on one of asset for each of security principle (availability, integrity or confidentiality) as follows: $P_{availability|integrity|confidentiality}(a_i) = P_{(c,[t;t+\Delta t])}(a_i|s)$.

Where

c - situation context within the time window $[t; t + \Delta t]$;

s - system state in the time window $[t; t + \Delta t]$.

Visual representation[1] of an attack on an asset is displayed in Figure 6.

[1] Full size picture of decomposition of an attack on an asset http://bit.ly/2PnSI8I

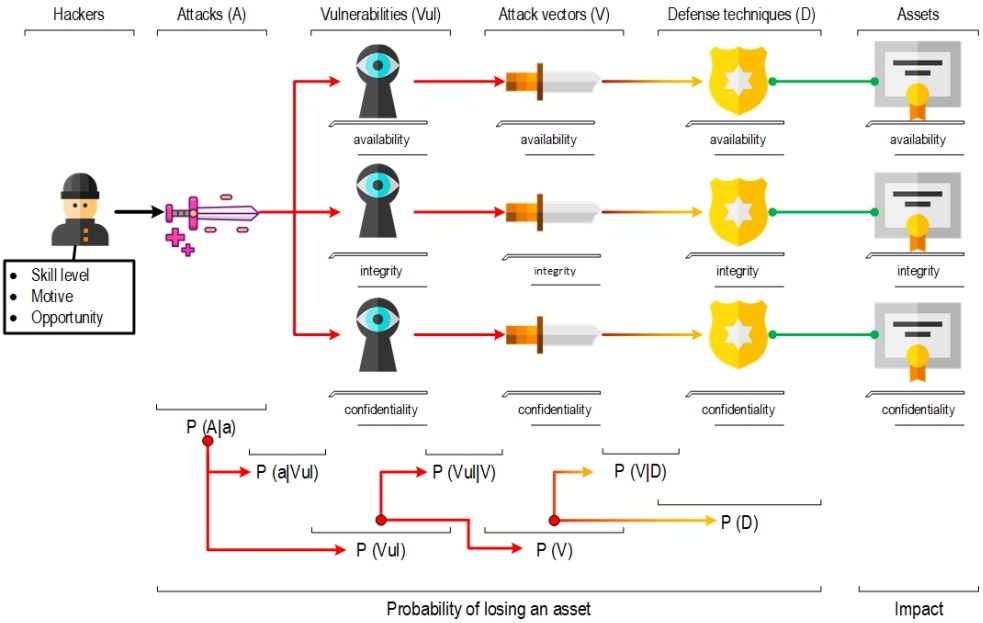


Figure 6: Visualized decomposition of an attack on an asset.

Since we cannot rely on average risk values, it is necessary to establish the probability of an attack a_i within the time window $[t; t + \Delta t]$ inside the context c (system parameters and LOGs), denoted by $P_{(c,[t;t+\Delta t])}(a_i|s)$:

$$P_{(c,[t;t+\Delta t])}(a_i|s) \quad = \sum_{i=0}^{k} \left(P_{(c,[t;t+\Delta t])}(A|a_i) \cdot P_{(c,[t;t+\Delta t])}(a_i)\right) \tag{9}$$

Where

k - stands for a number of all system states S;

$(P_{(c,[t;t+\Delta t])}(A|a_i)$ - denotes probability of attack a_i from various attacks (frequency of using this attack a_i), statistical data;

$P_{(c,[t;t+\Delta t])}(a_i)$ - stands for probability of occurrence of an attack a_i. This event is predetermined by total probability of exploiting all vulnerabilities $|Vul|$ during the time interval $[t; t + \Delta t]$. Let us consider this decomposition in more detail in the Formula 10.

$$P_{(c,[t;t+\Delta t])}(a_i) = \sum_{j=0}^{|Vul|} P_{(c,[t;t+\Delta t])}(a_i|vul_j) \cdot P_{(c,[t;t+\Delta t])}(vul_j) \tag{10}$$

Where

$|Vul|$ - number of vulnerabilities enabling commitment of an attack a_i;

$P_{(c,[t;t+\Delta t])}(a_i|vul_j)$ - probability that precisely this vulnerability vul_j among among others vulnerabilities of attack a_i (frequency of using this vulnerabilities vul_j), statistical data.

The probability of occurrence of a vulnerability vul_j event depends on total probability of exploitation of attacks vector $\overleftarrow{V}$ during the time interval $[t; t + \Delta t]$ presented in Formula 11.

$$P_{(c,[t;t+\Delta t])}(vul_j) = \sum_{i=0}^{|\overleftarrow{V}|} P_{(c,[t;t+\Delta t])}(vul_j|\overleftarrow{v_i}) \cdot P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}) \tag{11}$$

Where

$|\overleftarrow{V}|$ - identifies the number of vulnerability attack vul_j;

$P_{(c,[t;t+\Delta t])}(vul_j|\overleftarrow{v_i})$ - probability of using attack vector $\overleftarrow{v_i}$ from among various attacks vectors for the vulnerability vul_j (frequency of using this attack vector $\overleftarrow{v_i}$), statistical data.

More detailed representation of the attack vector probability can be tracked by analyzing Formula 12.

$$P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}) = \sum_{j=0}^{|D|} P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}|d_j) \cdot P_{(c,[t;t+\Delta t])}(d_j) \quad (12)$$

Where

$|D|$ - number of protective measures against the attack vector v_i;

$P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}|d_j)$ - harm caused by an attack vector $\overleftarrow{v_i}$ with a valid protection measures d_j (return value of quality of defense against attack vector), statistical data;

$P_{(c,[t;t+\Delta t])}(d_j)$ - probability of using this protection measure (yes or no).

When calculating values in one or more stages of the risk level analysis, some parameters of P (Formula 13) might be changed. Each attack vector has different parameters α_1 (information flow on the computer network), α_2 (number of backlinks), …, α_n, which affect the probability of an attack.

The detection of anomalies for parameter α_i serves as one of the ways to determine the intrusion through IDS. For example, IDS / IPS analyze the data to detect the following intrusions:

Anomaly detection;

Signatures and Heuristic Detections.

For each probability $P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}|d_j)$ of attack within context c, attack vector $\overleftarrow{v_i}$ and protection measure d_j we can define the function $f(\alpha_1, \dots, \alpha_n)$:

$$P_{(c,[t;t+\Delta t])}(\overleftarrow{v_i}|d_j) = f(\alpha_1, \dots, \alpha_n) \quad (13)$$

The improved prediction of probability of attacks could be achieved through the creation of artificial parameters. Consider such countermeasures to increase the prediction chances. Some countermeasures assume the application of parameters α that increase the predicting probability. The examples of attacks and countermeasures are presented below:

Copy a site using HTTrack software or similar programs [16]. In this case, it would be necessary to perform a real-time speed analysis and check the order of web pages asking. Consider an option of script-markers adding upon finding a complete copy of the website (JavaScript for the website's pages). The downloaded pages will automatically contain these built-in scripts. This will allow to determine whether these pages were opened from a different address or not. When opened, the address of the opened page will not match the original (canonical) address. The script for these markers sends information to the attack prediction system with the note "view the saved pages of the site" ($\alpha_{view-saved-pages-site}$) in case you view the downloaded pages on your website. In this case, it is desirable to send the maximum complete information from the computer which was used to open the registered copy of a website. This will help to obtain the information about the potential attacker, which will change the system risk level.

Create false or incorrect metadata for txt, docx, xlcx, pptx, etc. In this case, it is necessary to consider the creation of fake characters (first and last name). The search is done in the search engines that provide erroneous information to the potential attacker. Relying on this collection of information, it is necessary to transmit to the attacker the fake web page where it is necessary to obtain the most complete information on the potential attacks during the transmission and, consequently, a notification is sent to the attack forecasting system with a note "display false pages with the transmission of the search system" ($\alpha_{display-false-pages}$).

Fill in the robots.txt file with additional false information and try to confuse the attacker who is interested in this file. Add links to this file, which would provoke the attacker's interest, such as links that include words like "admin", "login", etc. Establish link tracking transitions in the same way as in step 2. All calls to the robots.txt file by agents that are not search engine robots must be logged to inform the system prediction attacks through the note "display of the robots.txt file is from intruder" ($\alpha_{display-robots.txt-by-intruder}$).

In the html code of the page, it is necessary to provide comments where fake addresses of system administrators and/or developers are found. Applying scanning programs, like The Harvester [15], the attacker will get "an interesting target for an attack" that will be an excellent marker for predicting attack system ($\alpha_{display-target-for-an-attack}$).

In case of our example, shown in Figure 7, it is possible to embed script-markers for the attack prediction system on Cloud VPS. With built-in scripts-markers, the $S_{scripts-markers}$ system parameter $\alpha_{scripts-markers}$ will be added to the list of indirect parameters affecting the probability of attack.

It is possible to add false metadata, including a fake author of some files (docx, pdf, etc.). One has to create a page with contacts of this fake author who created these files (docx, pdf, etc.) on the website. Relying on the site's viewing statistics, we can track visits to the pages of the fake author's files. Viewing this page is a consequence of studying method documents for the purpose of *footprinting and reconnaissance*. It means that someone is interested in the created fake pages, then the probability of the parameter $\alpha_{metadata-fake-search}$, will change the state of the system to the state $S_{metadata-faike-search}$.

For a variety of attack vectors one must identify parameters $\alpha_1, \dots, \alpha_n$ to produce the probability function of various attacks. These parameters, indicating a certain type of attack, must be "sieved" using the BigData method. It is necessary to eliminate noise from the results of the study. Following the BigData methodology, the most informative parameters will be highlighted.

To estimate the probabilistic law of our reference parameters $\alpha_1, \dots, \alpha_n$ we should choose different learning techniques, for example, the artificial neural network. This could help us find the exact function of $f(\alpha_1, \dots, \alpha_n)$ for each probability $P_{c_{[t;t+\Delta t]}}(a_i)$ of attack.

6. Example of simplified website administration on a cloud dedicated server

In this section we will attempt to clarify the application of our approach by referring to a simplified example. Figure 7 illustrates an example of simplified website administration on a cloud dedicated server (system S).

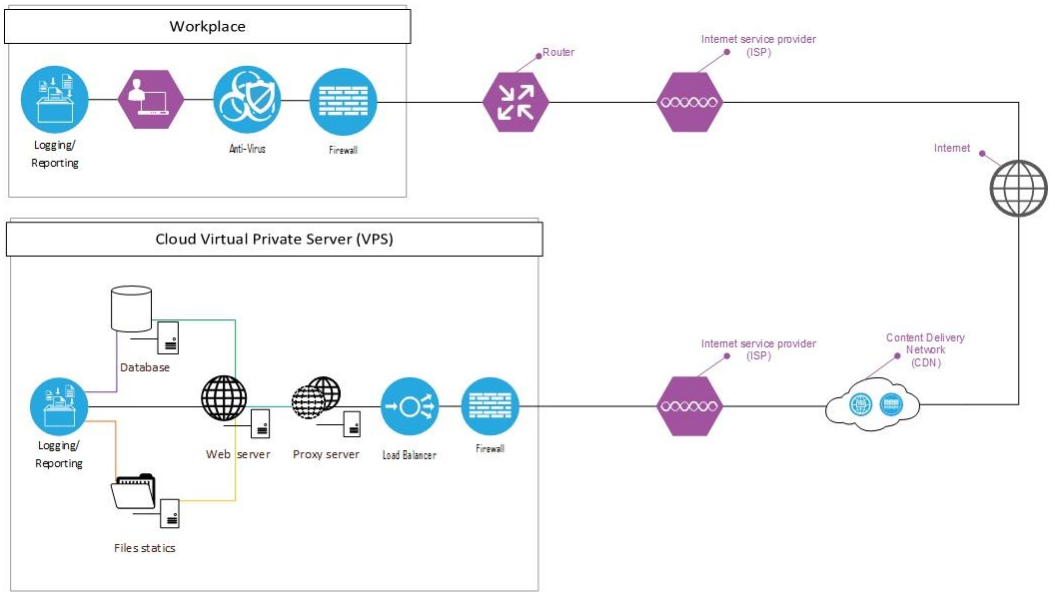


Figure 7: Website administration on a cloud dedicated server.

The list below contains a general set of website administration components on a dedicated server:

Workplace:

Firewall;

Anti-virus;

Logging/reporting by operating system (OS) and by anti-virus.

Router (for connection to the Internet Service Provider (ISP));

Content Delivery Network (CDN);

Cloud Virtual Private Server (VPS):

Firewall;

Load balancer (e.g. Gobetween, Nginx, etc.);

Proxy server (e.g. Squid, Varnish, etc.);

Web server (e.g. Apache, IIS, etc.);

Statics files (e.g. images, CSS, JS, etc.);

Database (e.g. MySQL, MSSQL, etc.);

Logging/reporting by OS.

The administrator has root access via SSH to the server through one-step authentication via a Certificate for Authentication, while a hacker wants to obtain a database dump. This is a simplified model consisting of two sides: administrator and hacker. We will refer to this example later to clarify different steps within the framework of our approach.

6.1. Example of assets analysis

Let us analyze the system description in more detail. The initial stage encompasses the consideration of assets. They are falling under definition of business processes that use assets (application data, data bases, personal office data, local area network services, common services, working environment, digital accounting control, etc.) A: $a_0 \dots a_n$, where $n \in N$.

Each asset class comprises the following components avl - Availability, int - Integrity, cnf - Confidentiality:

$$a = (Asset_0, avl, int, cnf) \tag{14}$$

An estimate is made of their value $(Val_{avl}, Val_{int}, Val_{cnf})$. This evaluation occurs by assigning each asset a certain criticality level when the asset is completely lost. Thus, the damage is taken into account after the asset's complete loss. The asset valuation is carried out in compliance with the selected methodology (MEHARI, CobiT, etc.). The system S performs an analysis in compliance with the chosen risk assessment methodology RM. This means that the value of the used assets is determined in case of each business process.

A: represents a set of assets. The sequence of $a_0, a_1, \dots$ describes a set of meta-variables used by us, which are ultimately comprising A. For this example of assets: a_0 =Application data (data bases).

A^N: describes a set of attribute names of A. Availability (avl), Integrity (int), Confidentiality (cnf), Efficiency (eff) are attribute names applied by us to specify a security class of an asset a. For example, efficiency of the management process in order to comply to the legal, regulatory or contractual requirements, domain laws and regulations.

A^V: represents a set of attribute values of N_{att}. We are proposing to use 4 point scale (1 = Low, 2 = Acceptable, 3 = Inadmissible, 4 = Intolerable) for any asset's attribute name (avl, int, cnf, eff) while assessing the degree of importance (the level of loss when the asset is lost) of each value.

Val: represents a function 15 of assets valuation.

$$Val: A \times A^N \to A^V \tag{15}$$

In the example, provided by us, it is defined as $Val_{a_0} = (3,3,4)$ (Application data) and established for the following three attribute's components (Availability $a_0[0] = 3$, Integrity $a_0[1] = 3$, Confidentiality $a_0[2] = 4$).

For a_0 = Application data (data bases)

$Val_{a_0[0]} = 3$ (Availability = Intolerable);

$Val_{a_0[1]} = 3$ (Integrity = Inadmissible);

$Val_{a_0[2]} = 4$ (Confidentiality = Inadmissible);

6.2. Example of protection measures analysis

It is worth to undertake an analysis of protection measures D for assets A involved in business processes BP for the initial configuration of the system $Conf_h$, where $h = 0$ stands for the initial configuration.

$$D = (Conf, v), \tag{16}$$

where v stands for vulnerabilities and exposures.

The following protection measures are ensured in case of our system:

Cloud Virtual Private Server (VPS);

Firewall;

Logging/reporting by OS.

The following Formula 17 is used by us to determine the risk for business processes:

$$RM: BP \times A \times D \to 2\text{"Tolerant"} \tag{17}$$

The risk level is assessed for the applied system configuration $Conf_i$, and it corresponds to the undertaken security measures. The network audit is carried out applying the chosen risk assessment methodology RM. At configuration modeling stage, the system configuration variants are created to match the acceptable risk levels.

6.3. Example of system states modeling

Consider the example of a network shown in Figure 7. In our example, a hacker would undertake an attempt to download the entire database dump. Here we are speaking about SQL injection[1]. In that context we would need to use the following Formula 18 as a first step:

$$R_{database}(\text{SQLinjection}) = \sum_{i=0}^{1} (P_{availability}(\text{SQLinj.}) \cdot 3 + P_{integrity}(\text{SQLinj.}) \cdot 3 + \\ + P_{confidentiality}(\text{SQLinj.}) \cdot 4) \tag{18}$$

Relying on OWASP it is possible to conclude that $P_{confidentiality}$ is one of the most widespread types of attack[2].

$$P_{(c,[\text{month}])}(\text{SQLinj.}|s) = \sum_{i=0}^{1} (1 \cdot P_{(c,[\text{month}])}(\text{SQLinj.})) \tag{19}$$

For illustrative purposes let us consider only the following vulnerability CVE-2019-8429 [14] applying Formula 10:

$$P_{(c,[\text{month}])}(\text{SQLinj.}) = \sum_{j=0}^{1} P_{(c,[\text{month}])}(\text{SQLinj.}|\text{CVE} - 2019 - 8429) \cdot \\ \cdot P_{(c,[\text{month}])}(\text{CVE} - 2019 - 8429) \tag{20}$$

Only one attack vector (Network) is identified for this vulnerability, therefore, Formula 11 will look as follows:

$$P_{(c,[\text{month}])}(\text{CVE} - 2019 - 8429) = \sum_{i=0}^{1} 1 \cdot P_{(c,[\text{month}])}(\overbrace{v_i}) \tag{21}$$

Taking into account the applied protection measures (firewall and logging), we may represent Formula 12 as follows:

$$P_{(c,[\text{month}])}(\overbrace{v}) = \sum_{i=0}^{1} (P_{(c,[\text{month}])}(\overbrace{v}|\text{firewall}) \cdot 1) \tag{22}$$

In our case, the firewall does not protect against SQL-injection. The success rate of this attack corresponds to 1. Accordingly, the level of risk remains "Unacceptable" in one month : $S_0 \to S_1$ (Figure 4). To maintain the level of risk, it is

[1] OWASP SQL Injection http://bit.ly/2vmHNTD

[2] Top 10-2017 A1-Injection http://bit.ly/2VVAopZ

necessary to reconfigure the system $S_0 \rightarrow S_2$ (Figure 4). For possible attack scenarios, it is necessary to select such α parameters (markers) that will indicate a planned attack. In our case, a log analysis can show how often hackers are attempting to identify vulnerabilities. For our example, the period of one month is based on the logs analysis. Thus, we can predict the level of risk for the future. This will allow us to prepare for an attack in advance.

7. Future Work

One of the possible directions of the scientific work is the Theory of Games. This theory enables the prediction of behavior patterns of attacking and defending sides. Let us briefly consider this option.

Today, it is possible to draw an analogy between an Intruder (or a group of cybercriminals) and an information security specialist (a group of information security experts). Both could be compared with two gamers (teams) playing against each other in real time. This game environment creates a realistic situation when both teams must take rapid decions, which might have serious consequences.

The teams are lacking time and have rather limited amount of information while making such decisions. In such a game, the defending party may incur maximum financial losses, while the attacking party can be prosecuted. The stakes in such a game are raised much higher, in the case the attacking party is represented by a special governmental department acting nationwide.

The mechanism of this game is shown in Figure 8.

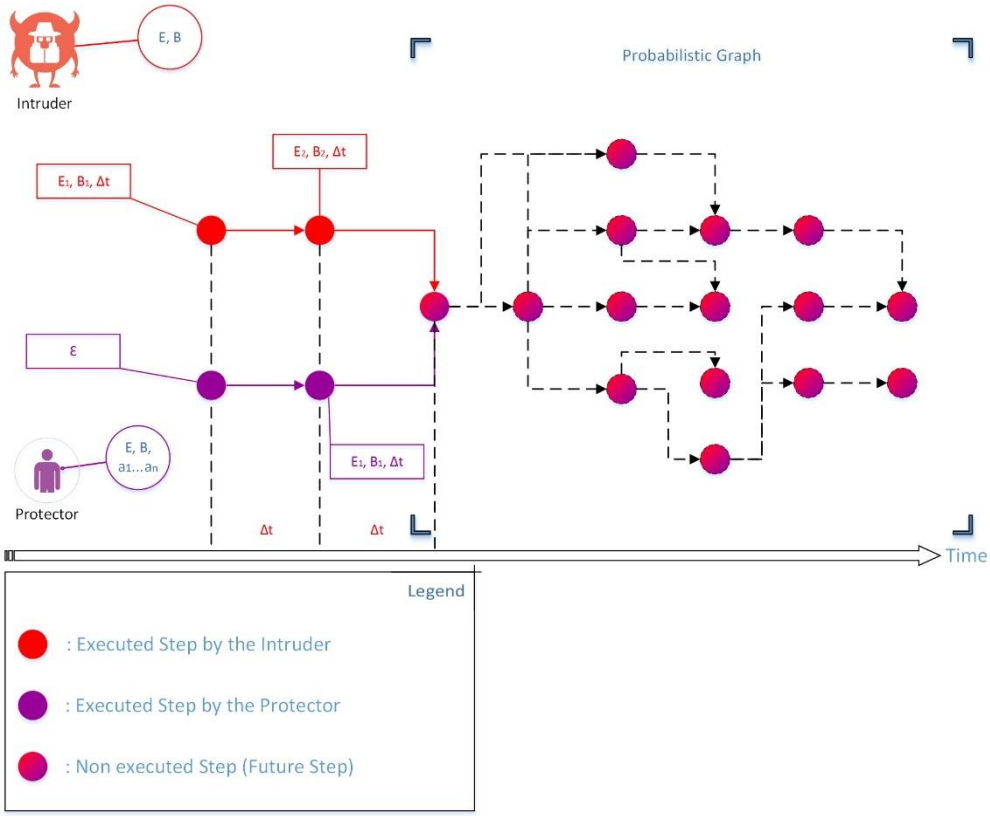


Figure 8: The mechanism of game for cybersecurity.

Consider a formal example of such a game. The attacker - Intruder (I) starts his game by identifying the vulnerabilities in the information security system of the defending party - Protector (P). Protector makes an action plan $a_1 \ldots a_n$ to improve the protection of your system in real time. Intruder, in turn, analyses the presence of the vulnerabilities $v_1 \ldots v_n$. Afterwords,

he/she compiles an attack strategy. Each side has its own characteristics in this case, which include the level of experience E, the budget B, the time limit for the accomplishment of certain actions Δt. Bayesian games could be viewed as another representation of the characteristics describing the game theory. In Bayesian games, the challenge is to identify how a player can assign appropriate initial beliefs to its opponents. Furthermore, it is sometimes interesting to consider a dynamic update of the players' beliefs. The formal definition of the repeated Bayesian game can be expressed applying a 7-tuplet [6]:

$$G = (N, \Theta_k, A_k, H(t_q), \Sigma_k, \mu_k, U_k) \quad (23)$$

where:

N is the set of game players (M stands for a number of players);

Θ_k is the set of possible player types $k \in N$;

A_k is the set of available action types $k \in N$;

$H(t_q)$ is the set of possible $date - t_q$ game histories;

Σ_k is the set of the player's behavior strategies $k \in N$;

μ_k is the player's posterior belief k defined as a conditional probability that its opponents' types are θ_{-k}, given history $h(t_q)\varepsilon H(t_q)$ and type $\theta_k \in \Theta_k$;

U_k is the utility function of a player $k \in N$ until time $t_q, q \geq 0$, given the history $h(t_q)$.

Let us consider as a basis the shortest time for the execution of an attack or countermeasures taken during Δt. Each side has its own budget B and experience corresponding to level E. Anyway, a certain level of experience is required to perform one action, which also results in time expenditures expresses as $n \cdot \Delta t$. In this case, the empty action is taken ε. Upon the accomplishment of several actions, each system state would affect the opponent's subsequent steps. The system states modeling is enabled through the resort to probability graphs.

Conclusion

This research work has practical applications in information security systems. The findings of this work will contribute to the development of prediction of cyberattacks. Thus, it will be possible not only to simulate a threat, but to determine the level of its risk, depending on different configurations of security systems. This will enable more effective information security management. This work will serve as a basis for further research in the area of distribution of funds for the investment in information security [2]. This research work is also an attempt to prove that in the context of system security it will be possible to predict the level of risk of the weakest points based on the analysis of statistical data and hackers' behaviour in different contexts.

References

[1] Common vulnerability scoring system calculator version 3 cve-20198429," http://bit.ly/2XufnCR, accessed: 2019-04-20.
[2] Owasp risk rating calculator, https://bit.ly/2VmPUij, accessed: 201905-07.
[3] Threat prediction platform http://bit.ly/2Zadf4X, accessed: 2018-0331.
[4] Ellis and T. McELwee, "System and method for predicting impending cyber security events using multi channel behavioral analysis in a distributed computing environment," Mar. 21 2017, US Patent 9,602,530.
[5] Seifert, I. Welch, and P. Komisarczuk, "Identification of malicious web pages with static heuristics," in *Telecommunication Networks and Applications Conference, 2008. ATNAC 2008. Australasian*. IEEE, 2008, pp. 91–96.
[6] Siaterlis and B. Maglaris, "Detecting ddos attacks with passive measurement-based heuristics" in *Computers and Communications, 2004. Proceedings. ISCC 2004. Ninth International Symposium on*, vol. 1. IEEE, 2004, pp. 339–344.
[7] Achmadi, Y. Suryanto, and K. Ramli, "On developing information security management system (isms) framework for iso 27001-based data center," in *2018 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2018, pp. 149–157.
[8] Fudenberg, "Andj. Tirole, Game Theory," 1991.

[9] J. Farahani, M. H. A. Kachoee, and M. A. A. Kachoee, "Vulnerability assessment of the critical infrastructure against man-made threats," *Industrial Engineering & Management Systems*, vol. 17, no. 1, pp. 136– 145, 2018.
[10] Cavusoglu, B. Mishra, and S. Raghunathan, "A model for evaluating it security investments," *Communications of the ACM*, vol. 47, no. 7, pp. 87–92, 2004.
[11] Debar and A. Wespi, "Aggregation and correlation of intrusiondetection alerts," in *Recent Advances in Intrusion Detection*. Springer, 2001, pp. 85–103.
[12] Mehari expert (2010) tool." http://bit.ly/2uHDFgh, accessed: 201903-31.
[13] N. C. Wael Kanoun, Frederic Cuppen, in "Evaluation des risques dans une processus de supervision de la securité," in *Ecole Nationale Superieure des Télécommunications.* IFSIC, 2007, pp. 11–12.
[14] N. Minar, K. H. Kramer, and P. Maes, "Cooperating mobile agents for dynamic network routing," in *Software agents for future communication systems*. Springer, 1999, pp. 287–304.
[15] P. Mongsawad *et al.*, "The philosophy of the sufficiency economy: a contribution to the theory of development," *Asia Pacific Development Journal*, vol. 17, no. 1, p. 123, 2010.
[16] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, "Phishnet: predictive blacklisting to detect phishing attacks," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–5.
[17] P. Yermalovich and M. Mejri, "Formalization of attack prediction problem," in *2018 IEEE International Conference" Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*. IEEE, 2018, pp. 280–286.
[18] Mehari expert 2010 fr, http://bit.ly/2Glaeal, accessed: 2018-02-21.
[19] Harvester - osint, http://bit.ly/2KxADGk, accessed: 2018-02-23.
[20] Httrack." http://bit.ly/2VQl6T2, accessed: 2018-02-23.

Students' Consciousness of their Problem Solving Approaches as a Key to Creativity in Design

Marwa Hassan Khalil

Associate Professor of Architecture, Department of Architectural Engineering and Environmental Design, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt.

Abstract

Architectural engineering students are constantly dealing with ill-defined and tangled design problems. Many scholars accentuated the importance of creative thinking in tackling such wicked and complex problems. Accordingly, getting engaged in an ill-defined problem solving process requires specific personality traits that are often critical to creativity and innovation in design. In that sense, architectural engineering curricula need to provide various strategies through which such individual skills can be nurtured and developed. The objective of this study is to empirically identify the different patterns of students' approaches in solving problems and the role of group discussions in such a process. The study adopted a qualitative approach, in a live class setup, through a series of workshops to allow for in-depth exploration of the students' problem solving skills and abilities. The intention is to help students in discovering and in being aware of their own way of solving problems and identifying its strengths and weaknesses. This is considered a core and significant step towards the improvement and development of their design thinking skills. The findings of the study have emphasized the positive impact of the cyclical behavior in the creative problem solving process and highlighted the different key issues and lessons emerging from students' consciousness of the mental processes that occurred during this iterative process. Such awareness and consciousness of those emergent issues is expected to encourage conscious design, increase tolerance for ambiguity and improve self-confidence which are believed to dramatically help students in creatively solving ill-defined architectural design problems.

Keywords: problem solving, individual skills, creativity, design thinking, ill-defined problems, design education, architectural design; students' awareness, conscious design, Geneplore model

1 Introduction

In architecture, designers are constantly dealing with wicked, tangled and ill-defined design problems (Buchanan, 1992; Rittel, 1972; Schumacher, 2012). Addressing such complex problems call for creative thinking in order to be able to solve them (Hocking & Vernon, 2017). Getting involved and immersed in an ill-defined and intricate design problem solving process requires an immense cognitive and mental effort, courageous behavior, high tolerance to ambiguity and self-confidence (McAdam & McClelland, 2002; Paletz & Peng, 2009; Reiter-Palmon & Illies, 2004). In that sense, fostering creativity in engineering and design education to enable students to develop as creative designers is quite inevitable (Baillie, 2002; Lau, 2017). Preparing them to be future professional designers necessitates the implementation of practical steps that make creativity an integral part of the architectural curricula (Kowaltowski, Bianchi, & de Paiva, 2010). However, many scholars highlighted the fact that education programs for development of creativity in design education are quite scarce and not provided sufficiently (Baillie, 2002; Bourgeois-Bougrine, Buisine, Vandendriessche, Glaveanu, & Lubart, 2017; Cho, Hong, & Kwang-Soo, 2016).

Therefore, different educational institutions are expected to provide various means and strategies to couple creativity and design education. In an attempt of improving and developing the students' design thinking and cognitive skills, the study aims to investigate the different students' approaches in solving problems through a series of workshops in a course entitled AR221 Scientific Thinking in the Department of Architectural Engineering and Environment Design at the Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt. This allows for an in-depth exploration of the students' problem solving skills and abilities, identifying their strengths and weaknesses.

The study raises a set of questions that can be summarized as follows:

- Are students aware of their own way of solving problems?
- Can this awareness and consciousness of the mental processes occurring in problem solving help them in the development and improvement of their design thinking skills?

To answer the questions and meet the objective, the study adopted a qualitative approach, in a live class setup, through two different workshops. The study employed a variety of data gathering tactics and methods to allow for data triangulation.

2 Creativity in Problem Solving: A Cognitive Perspective

There are a variety of approaches that tackle creativity from different perspectives. Many scholars addressed this issue and worked on the categorization of the different approaches and paradigms to creativity. Sternberg and Lubart (1999) discussed creativity in terms of six different approaches: mystical, psychoanalytic, pragmatic, psychometric, cognitive and social-personality approaches. They concluded that another confluence approach, a multidisciplinary one, in which different paradigms and components of creativity can converge is required.

Consequently, a categorization that is based mainly on the work of Taylor (1988), Sternberg and Lubart (1999), and Villalba (2008) was proposed (Cachia, Ferrari, Ala-Mutka, Punie, & Institute for Prospective Technological Studies, 2010). Their categorization included five main approaches: psychometric, psychoanalytic, self-expression and mystical, end-product and cognitive approaches. To them, the cognitive approach embraces phase oriented studies, pragmatic methods and thinking theory. More precisely, it is considered as an umbrella term under which several original paradigms such as pragmatic, cognitive and social-personality approaches are combined.

The perspective proposed by Cachia et al. (2010) in understanding the cognitive approach in a multi-disciplinary nature is quite relevant to this study, which is in an educational context, for various reasons: 1) it addresses creativity as a process and in education, emphasis on the process should be always given a top priority (Cachia et al., 2010); 2) it considers creativity as a skill and accordingly can be developed and nurtured especially in educational and learning environments (Edward de Bono, 2007, 2009); 3) this multi-disciplinary approach is in a sense a step towards implementing different ways and strategies of learning appropriate to each person (Gardner, 1999). Therefore tackling creativity from an inclusive cognitive perspective that involves phase oriented studies, pragmatic methods and thinking theories is the main focus of this study.

There are diverse well known phase oriented studies that deal with the different steps and stages of the creativity process such as the model of Wallas (1926) and the Geneplore model (Finke, Ward, & Smith, 1992). This study focuses on the Geneplore model, which is not linear in nature, because of its relevance to the engineering design cycle and this can help in developing the design thinking skills of architectural engineering students. As implied by its name, the Geneplore model divides the process of creativity into a generation phase and exploration phase (Finke et al., 1992) . The model accentuates the dynamic nature of the mental processes that might occur in a back and forth behavior throughout the problem solving process. If the output is non-satisfactory a return to the generative phase is encouraged (Fig. 1). Regarding pragmatic methods, creativity is developed through different techniques and methods, such as those proposed by Edward de Bono, who is a leading figure in this approach. His tools and methods are used to develop lateral thinking skills and to broaden one's perception of a matter (Edward de Bono, 1970, 1991, 2000, 2007). In addition, thinking theories focus on how personality traits and environmental factors are related to creativity (Cachia et al., 2010). Such theories highlight the importance of several personality traits such as self-confidence, attraction to complexity, risk taking, self-efficacy, willingness to overcome obstacles and tolerance for ambiguity to creativity (Gardner, 1999; Sternberg & Lubart, 1999).

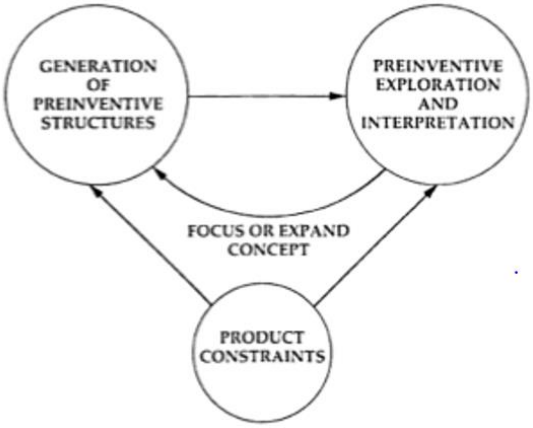


Fig. 1. The basic structure of the Geneplore model (Finke et al., 1992)

Based on the above review of the different schools of thoughts within the inclusive cognitive approach (Cachia et al., 2010), one can conclude the commonalities and overlaps between them. They all focus on the essence of creativity as a process rather than supporting the notion that creativity comes suddenly and unexpectedly. Accordingly, this study tackles creativity in problem solving from this cognitive perspective. It was conducted in a live class setup where students were engaged in different workshops and exercises that are based on pragmatic methods. An in-depth exploration of the students' problem solving skills during the generation, exploration and modification phases was conducted. Reflection and insights of such investigations and its impact on their personality traits were concluded.

3 Methodology

The objective of this study is to identify the different patterns of students' approaches in solving problems and the role of group discussions in such a process. The participants in this study are students who are enrolled in a course entitled AR221 Scientific Thinking[1] in the Department of Architectural Engineering and Environment Design at AASTMT. The intention is to help students in discovering their own way of solving problems and identifying its strength and weakness. This is considered as a significant step towards the improvement and development of their design thinking skills.

3.1 Research Design

The study adopted a qualitative approach in a live class setup to allow for in-depth exploration of the students' problem solving skills and abilities. Such a live setup helps in understanding the phenomenon under investigation in the typical complex and messy setting of a semester with all the normal pressures associated with it (Taborda, Chandrasegaran, Reid, Ramani, & Kisselburgh, 2012). The study relies on the analysis of two different workshops. The data collection draws from different sources such as oral and written students' reflections and insights, the author's observational field notes, photographs, videos, audio recordings and samples of students' work.

3.2 Description of the Context and Workshops

The study was conducted in fall 2019 and 42 students in two different classes, 21 per each, were involved. It focuses on the analysis of two different workshops in which different exercises were conducted. The exercises are inspired from and based on original problems proposed by Edward de Bono (1991).

The first workshop was held on two consecutive sessions with a total duration of four hours. Most of the students were working in pairs constituting 23 groups numbered from 1 till 23. In this exercise, students were required to create abstract

[1] This course, in its current status, has been designed, developed and taught by the author since fall 2008.

compositions according to a set of constrains that increase in complexity from one step to another using six identical block-shaped objects[1].

The second workshop was held over a two-hour session. Students were working in groups composed of 4 – 5 students, with a total of 10 groups labeled from A to J. In this exercise, students were required, in two different problems[2], to place the bottles of water on a flat surface and the distance between each of them should be slightly more than the length of a linear flat element, in this case wooden tongue depressors were used. Using a maximum of four tongue depressors, they should construct a platform on top of the bottles and it ought to be strong enough to support a full plastic bottle of water.

By analyzing the work conducted and the mental processes involved during the problem solving process in the above workshops, the research is expected to answer the following questions:

- What are the different classifications of solutions and patterns of students' approaches in solving problems?
- What are the key issues and lessons emerging from students' consciousness of the mental processes that occurred during the cyclical problem solving process?

4 Analysis and Discussion

During the different workshops and exercises, which were based on pragmatic methods and techniques, multiple Geneplore cycles occurred and in fact one cycle informed the other. Encouraging students to alternate between generation and exploration has resulted in a significant improvement in the solutions and alternatives proposed by the students. This is quite consistent with the findings of Finke et al. declaring that "This cycling between the phases of generation and exploration typically occurs when people engage in creative thinking" (Finke et al., 1992, p. 18).

In this qualitative study, there was extensive verbal and visual material, in the form of oral and written students' reflections and insights, the author's observational and field notes and samples of students' work. A classification, sorting and categorization task was conducted to analyze this material in order to understand the reasons behind such a significant improvement in results. During this analysis process, close scrutiny helped in identifying the different categories of solutions and patterns of students' approaches in solving problems. Moreover, abstracting out assisted in capturing the common recurring ideas and themes which has led to the extraction of different key issues and lessons emerging from students' consciousness of the mental processes that occurred during this cyclical process.

The following part will discuss those different categories and patterns of problem solving and the emergent lessons highlighted while relating them to the relevant literature and previous studies. The findings of this qualitative study will be supported by quotations, observational notes along with case descriptions selected from the work of students.

4.1 Classification of the Different Students' Solutions and Alternatives

In digging deep trying to understand what happened during this cyclical problem solving process, different patterns and approaches of solving problems were discovered and solutions can be grouped and classified as follows:

4.1.1. Incorrect Trials Leading to Valid Solutions

It was observed that many students arrived at a valid solution through incorrect trials (or with the help of incorrect trials). Through trial and error, they concluded that it is not a must to change the whole idea, sometimes developing and modifying it leads to the required results. One of the students telling his colleague while looking at the alternative "We do not have to start from scratch, we can modify the existing incorrect solution and try to make it fit the given criteria" and finally they were able to rectify it (Fig. 2). Many students comprehended the vital role of incorrect trials in proposing valid and interesting solutions and how failure can inform success.

[1] Problem 1: arrange the six blocks so that each touches two and only two other blocks; problem 2: arrange the six blocks so that each touches three and only three other blocks; problem 3: arrange the six blocks so that each touches four and only four other blocks; problem 4: arrange the six blocks so that each touches five other; problem 5: arrange the six blocks in the following fashion (one block must touch only one other, one block must touch only two others, one block must touch only three others, one block must touch only four others, one block must touch only five others.

[2] First problem in this exercise was to use three bottles where each bottle forms the corner point of a triangle of equal sides; second problem in this exercise was to place four bottles where each bottle is placed at the corner of a square.

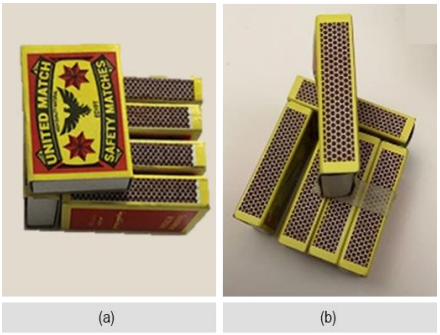


Fig. 2. An example showing (a) an incorrect trial leading to (b) a valid solution by experimenting through subtracting, displacing, flipping and rotating *(workshop one, problem 5, group 13)*

4.1.2. **Simple Valid Solutions Turning to Complex Solutions**

After arriving at valid solutions that fit the criteria of a given problem, some students who were still provoked to propose more creative alternatives started to dig in more. They tried to explore with different directions, orientations, positions, visual inertia, degrees of stability and surface area. For example, one of the groups presented some horizontal and vertical alternatives and they were even more challenged and tried to merge the horizontal and vertical treatments in addition to changing the stability and visual inertia of the compositions generating more complex solutions (Fig. 3). In another exercise, and through play and discovery, a group initially proposed an alternative in which they were able to lift the weight using the four wooden tongue depressors and placing the bottle with a small surface area, then a much smaller area on the platform (Fig. 4a and 4b). They were provoked and extremely engaged and tried to use only three sticks to lift the weight and they succeeded; surprisingly they decreased the surface area of the weight on the platform again until they were able to balance it (Fig. 4c and 4d). In general, this provocative and generative behavior has resulted in restructuring of the existing patterns and accordingly synthesizing it into new ones creating more complex and interesting alternatives.

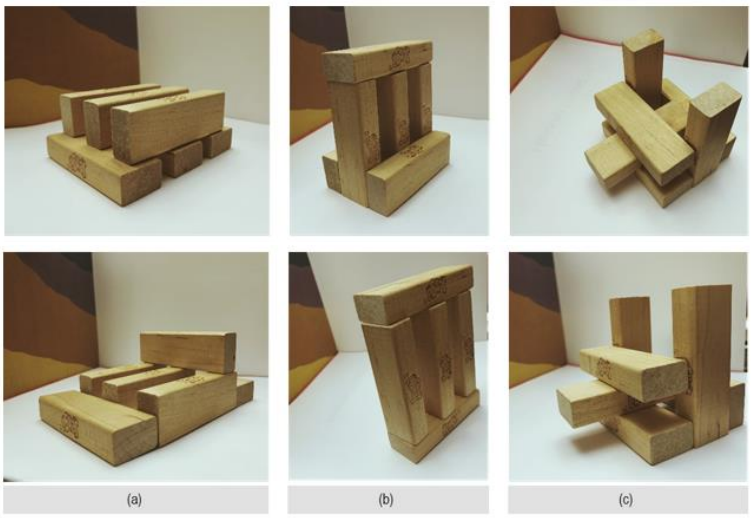


Fig. 3. An example showing simple valid solutions turning to complex solutions through experimenting (a) horizontally, (b) vertically and (c) merging horizontal and vertical solutions to generate new alternatives *(workshop one, problem 2, group 14)*

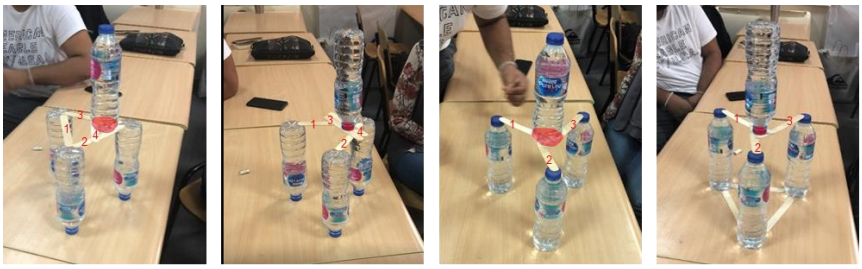


Fig. 4. Attraction to complexity through lifting the weight on the platform (a) using the four tongue depressors on a small surface area, (b) using the four tongue depressors on a much smaller surface area, (c) using only three tongue depressors on a small surface area and (d) using only three tongue depressors on a much smaller surface area *(workshop two, problem 1, group F)*

4.1.3. Solutions that Meet the Criteria Becoming Solutions that Exceed the Criteria

Some students were able to take the challenge even further than just meeting or satisfying the given constrains or rules. For example, they started to imagine, associate and link their abstract compositions to the surrounding environment (Fig. 5a). Others modified their alternatives trying to re-structure it in order to fulfill another self-imposed criteria (Fig. 5b). Others started to specifically relate and tie to different architectural phenomena, concepts and ideas such as cantilevers, voids and orientation in architecture (Fig. 5c). Furthermore, one of the groups not only associated to architecture by mentioning that the composition resembles a pathway, but also has modified the composition through shifting to enhance the proportions, enclosure, illumination and depth of the space creating a more complicated and innovative solution (Fig. 5d). In addition, some students tested the stability of the proposed platform not only through the given weight, full plastic bottle of water, but also by applying additional weights to the original one (Fig. 6).

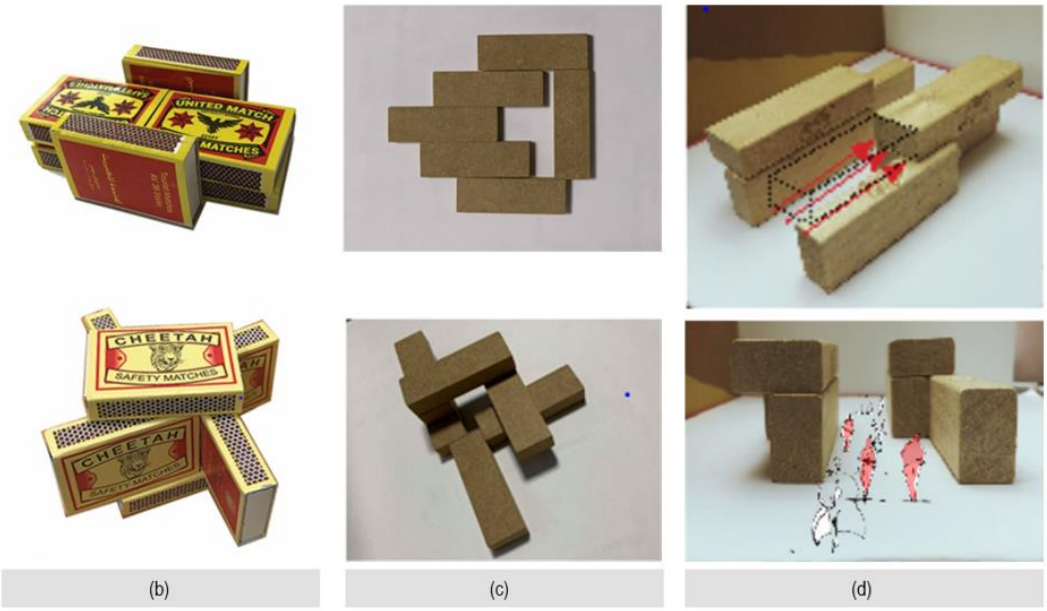


Fig. 5. Sample of students' work taking the challenge beyond the given constrains or rules (a) associating and linking their abstract compositions to the surrounding environment *(workshop one, problem 2 and 3, group 4),* (b) modifying the valid alternative trying to restructure it to generate a composition with a base, body and a cover in addition to meeting the original criteria *(workshop one, problem 3, group 9),* (c) developing a static solution to a dynamic self-supported structure, referring to it as cantilevered blocks, and imagining it as an architectural composition that is sitting lightly on the ground *(workshop one, problem 1, group 1)* and (d) imagining the abstract composition as a pathway, and accordingly shifting one of the blocks backwards to enhance proportions, enclosure, illumination and depth of the space *(workshop one, problem 1, group 17)*

)

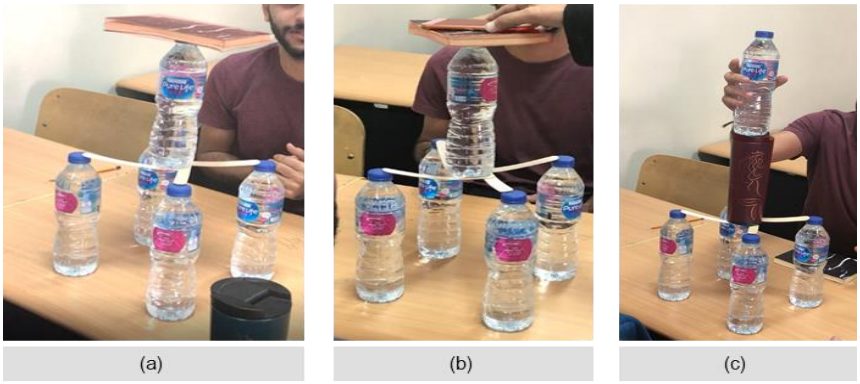


Fig.6. Experimenting and playing with extra weights to test the stability and strength of the platform (a) adding a notebook, (b) adding a notebook and a mobile phone and (c) full car mug in addition to the bottle *(workshop two, problem 2, group H*

4.2 Key Issues and Lessons Emerging from Students' Consciousness of the Cyclical Problem Solving Process

The analysis revealed four key issues and lessons emerging from students' awareness of the mental processes that happened during the cyclical problem solving process: experimentation and discovery, challenging the obvious, discussion and collaboration and deferring early judgment.

4.2.1. Experimentation and Discovery

Nearly, all the students have valued the importance of experimentation in the creative problem solving process and considered it as an integral part of the process. They understood the critical role that many operational verbs, such as trying, retrying, flipping, shifting, rotating, mixing, merging and integrating, play in proposing innovative solutions. Based on the analysis, many of those operations and experimental approaches done by students, yielded outstanding results. Through experimentation and discovery, they were able to generate, explore, modify, develop, diversify, fine tune and refine their outcomes. Some students were unable to find an alternative that fulfils the given criteria and through trial and error they were able to arrive at valid solutions (Fig. 7). Others were able, through play, to arrive at a solution adopting a sequential manner and based on this they were able to figure out a rule through which an infinite number of alternatives were generated (Fig. 8). Those who were not satisfied with their simple proposals kept trying and experimenting in order to discover new approaches and possibilities. As emphasized by Michalko (1991), a simple or mild change stimulated and provoked an endless number of ideas. This change might be in relational properties such as orientation, visual inertia, and position (Fig. 9a), or could even be in the nature, material or type of objects used. Change in visual inertia, degree of stability and balance of a composition with an identical arrangement yet a different object was observed, creating an interesting diversity in the proposed solutions (Fig. 9b).

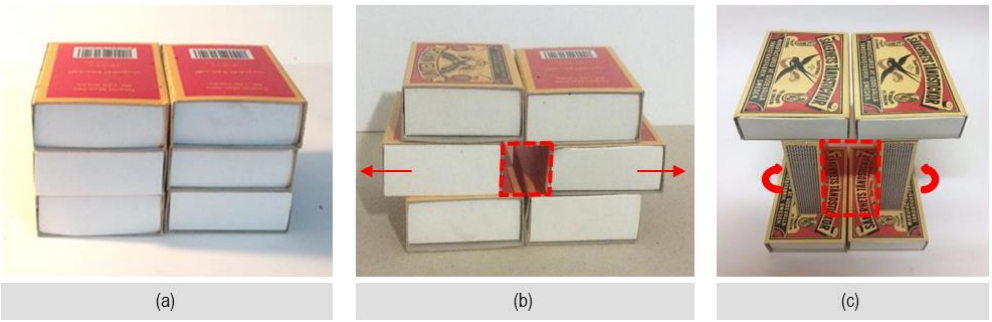


Fig.7. Fixing an (a) incorrect alternative through shifting to (b) create voids generating a valid solution in addition to (c) changing the proportions of the void, through rotating and flipping, creating a new alternative (*workshop one, problem 1, group 11*)

It was observed that play, fun and free investigations were always involved when a better solution was proposed by students. This is consistent with numerous studies arguing that students are deeply engaged in a learning environment that encourages play, discovery and having fun (Carroll & Thomas, 1988; Taborda et al., 2012). For example, one of the groups used different objects to test the strength of the platform. They started with a light object and when they succeeded they placed the given weight showing that they gradually became more confident (Fig. 10). Another group was experimenting and playing with extra weights to further test the stability and strength of the platform. Not only were they using different objects as extra weight to add more complexity to the challenge, but they also used precious objects such as mobile phones showing increased self-confidence and high inclination towards risk taking (Fig. 11).

Adopting a loose attitude, as referred to by Lin (1993), rather than a tight one and a willingness to try and re-try and see a mistake as something positive is greatly important to develop as creative thinkers and designers. Fostering an environment for creative work requires providing a balance between structure and free investigation, encourages play and fun methods and emphasizes the importance of reflection and iteration (Edward de Bono, 1991; Puja Khatri & Sumedha Dutta, 2018).

According to Cross (1999), design is opportunistic and exploratory in nature and cannot be predicted or anticipated in advance. Thus, a creative designer needs to think about what might lie ahead, discover something new instead of recycle

something that he/she already knows. This confirms the importance of experimentation and discovery to creative problem solving in design and how it helps in improving self-confidence, encourages risk taking and enhances tolerance for ambiguity and they are all critical to creativity in design (Cross, 1999).

Fig.8. Figuring out a general rule (a) proceeding in a sequential manner, (b) and accordingly an infinite number of alternatives was generated (*workshop one, problem 5, group 18*)

4.2.2. **Challenging the obvious**

It was found that raising questions was extremely helpful in the problem solving process. According to Michalko (1991), questions stretch one's eyes wide open. Questions helped in viewing the challenges from different perspectives, thus introducing new possibilities. This was either achieved through an insight which was quite rare, or mainly through critically revisiting and analyzing their proposed alternatives. Some students started to group and classify their proposals highlighting similarities and accordingly concluding the self-imposed constrains that they were imprisoned by and consciously started to challenge them. For example, they started wondering does it necessarily have to be a loop! Why not vertical? Should it be only orthogonal? Why symmetrical? In addition to many other similar questions (Fig. 12a, 12b and 12c). In other cases, students tried to analyze their incorrect trials, what didn't they try, challenging how else and what else is missing (Fig. 12d).

This curious and skeptical behavior allowed students to reverse the different conventional assumptions that they have subconsciously imposed on themselves, and this in turn, has helped them not only in generating and developing their proposals but also in proposing unique ideas and breakthroughs. Based on this, most of the students became more attracted to complexity, patient even if they do not know the answer yet and more willing to overcome any obstacles. They, as emphasized by Edward de bono (1970), started to use provocative manners instead of simple ones and believed that no matter how good something is, there is always a potential of doing it better.

As a reflection, being skeptical and challenging the obvious along with how this positively affects the tolerance to ambiguity and willingness to overcome problems is very important in addressing architectural design problems. The design process is an indeterminist one, as referred to by Goldschmidt (1997), which is characterized by uncertainty and ambiguity (Cross, 1999) and thus requires such skills. Preconceptions and prejudgments are issues that need to be widely addressed in design education (Kowaltowski et al., 2010). Liberation from conventional assumptions and preconceptions help designers to expand their possibilities. Although this is quite overwhelming for designers, yet this leaves many options open for as long as possible and that is a merit that usually leads to creative and successful designs.

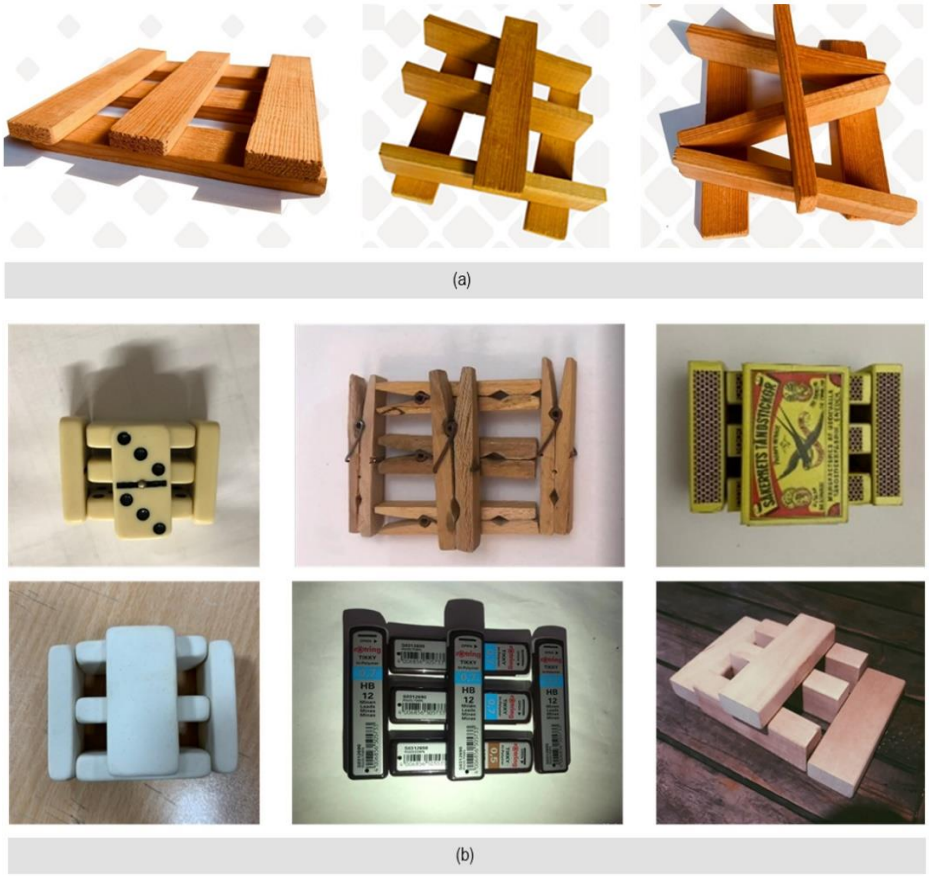


Fig.9. Simple change stimulating and provoking new ideas (a) change in position, visual inertia and orientation leading to more dynamic alternatives *(workshop one, problem 2, group 15)* (b) experimenting and playing with different objects and observing the change in visual inertia, degree of stability and balance of the compositions (*workshop one, problem 5, group 5*)

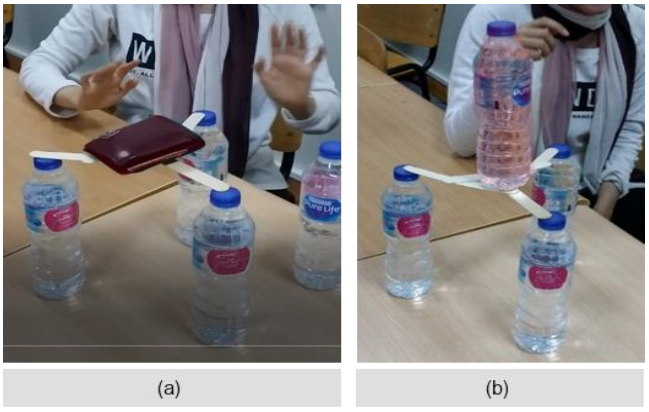


Fig.10. Testing the stability and strength of the platform (a) first using a light object (wallet) (b) using the given weight (bottle) showing gradual increase in self-confidence *(workshop two, problem 1, group C)*

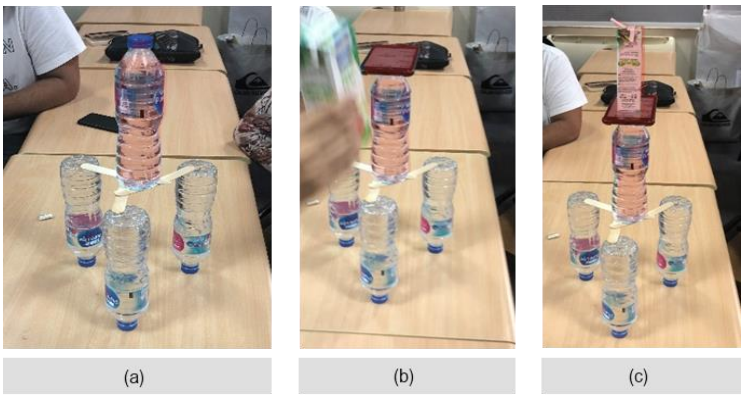


Fig.11. Testing the stability and strength of the platform (a) using only the given weight (bottle), (b) using the given weight (bottle) and a mobile phone, (c) using the given weight (bottle), a mobile phone and on top of it, a small juice box showing increased self-confidence and high inclination towards risk taking *(workshop two, problem 1, group F)*

4.2.3. Discussion and Collaboration

Based on the students' reflections and the author's observations, collaboration, discussion and free-wheeling were very useful and helped many students to arrive at their solutions or even develop them. While working in pairs, during the first workshop, several groups have highlighted the importance of discussion in the problem solving process. One of the students mentioned that the tinkering of his partner to the proposals that he offered was extremely inspiring. He declared that it was beneficial to both of them as mingling each other solutions usually helped them in proposing a valid solution and in developing it to more interesting and unique alternatives

Others mentioned how thinking out loud, talking about ideas to others was important. Listening to insights that emerge from the group rather than trying to push one's own idea, as emphasized by Sawyer (2006), was the spark that lead to numerous alternatives and solutions. For example, while trying to lift the weight on the platform and after nearly losing hope, one of the group members who was meticulously observing the trials yelled "Let us weave the sticks" and this was it (Fig. 13). Another group highlighted how discussion and collaboration has helped them in taking the challenge even further and applying heavier objects to the platform instead of only the bottle and when showed stability and strength; they became more confident and took higher risks and were driven to more complex challenges.

In that sense, collaboration and discussion are assets that significantly helps in generating genuine and new ideas which lies at the heart of any design discipline. Listening, talking, observing and accordingly developing solutions constantly leads to better results. Scientists, designers and professionals in all fields reported that their most innovative ideas and substantial results emerged from collaborations (John-Steiner, 2006).

4.2.4. Defer and Postpone Early Judgment

During those workshops and exercises, students developed a deeper understanding of how early and quick judgement specifically during the generation phase might deprive them of formulating a unique and creative solution. In fact, it is impossible to be curious and judgmental at the same time. They reflected that such an early assessment, especially within a group, that an idea does not work, negatively affected their level of engagement and suppressed their energy, spirit and contribution. In the second workshop, one of the groups arranged the sticks in a peripheral manner trying to lift the given weight and one of them quickly decided that this is impossible and the whole group got trapped, frustrated and lost the momentum for some time unable to propose other alternatives (Fig. 14a). On the contrary, another group reached the same point and they kept trying even if it looks impossible until they were able to lift the bottle using the sticks arranged only in a peripheral manner by placing the bottle horizontally and it worked (Fig. 14 b). Not only did they achieve this alternative, but they were also able to challenge themselves more and more presenting better and more complex alternatives, just because they deferred early judgement and had high tolerance to uncertainty.

Those findings are in accordance with other previous studies. For Puja Khatri & Sumedha Dutta (2018), a stress free or a non-judgmental environment helps students to express their ideas freely and this in turn opens the learning environment to new thoughts and opportunities. Furthermore, in a group work or brainstorming session, success is related to two main

principles one of which is deferring judgement (A.F. Osborn, 1963) and allowing the creative current to flow. Inevitably, this is very important in proposing solutions to ill-defined and tangled design problems.

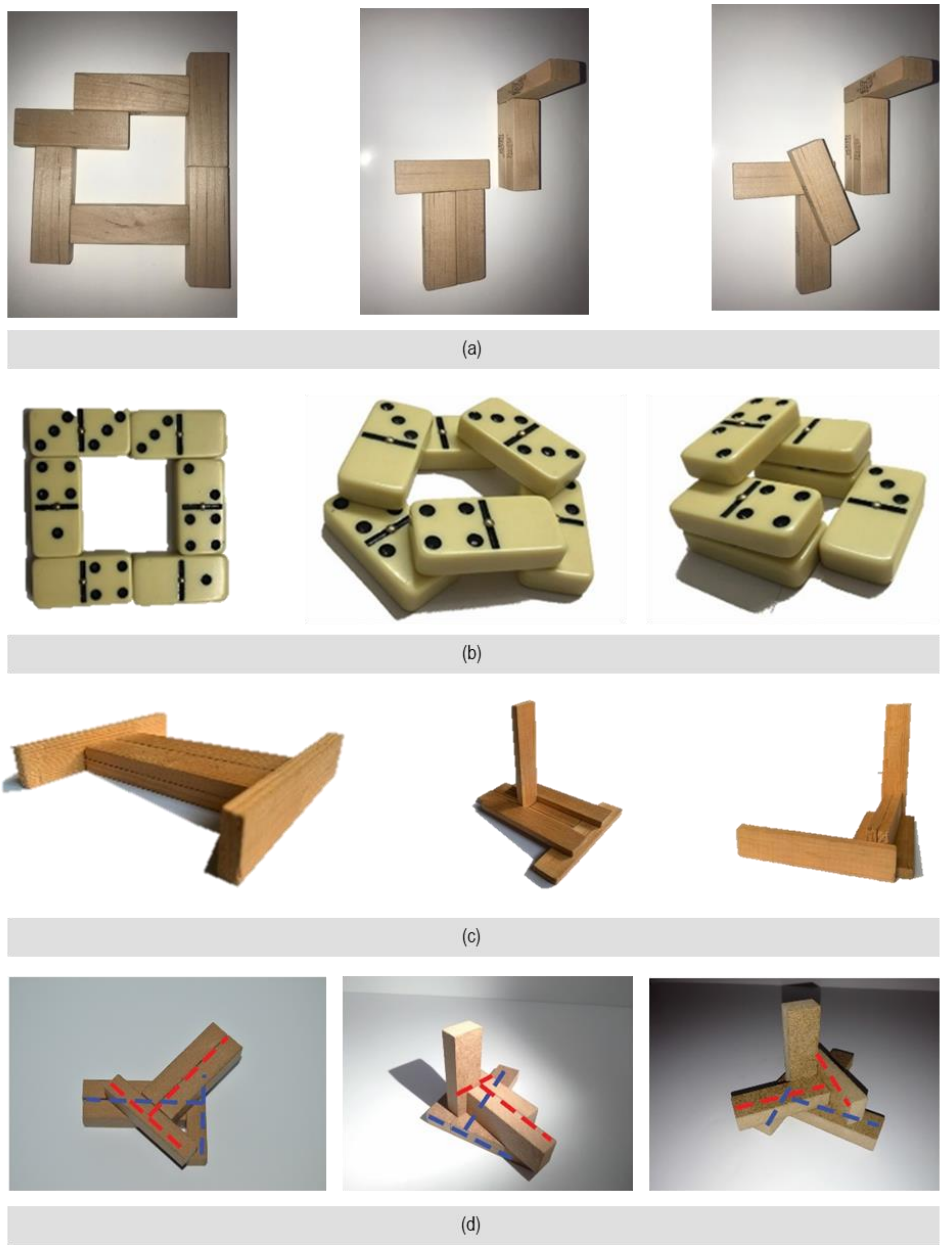


Fig.12. Challenging the conventional self-imposed constrains and assumptions by wondering (a) does it have to be a loop? *(workshop one, problem 1, group 7),* (b) does it have to be one level? *(workshop one, problem 1, group 9),* (c) does it have to be symmetrical? *(workshop one, problem 3, group 15)* and (d) can a diagonal relationship help in solving the problem? *(workshop one, problem 4, group 12)*

(a) (b)

Fig.13. Collaboration and discussion to lift the weight (a) as one of the members yelled "let us weave the sticks", (b) and they did *(workshop two, problem 1, group B)*

(a) (b)

Fig.14. Deferring and postponing judgement as a key issue in creative problem solving (a) quickly judging that it is impossible to lift the weight with sticks arranged only in a peripheral manner *(workshop two, problem 1, group D),* (b) another group lifting the weight using the sticks arranged only in a peripheral manner by placing the bottle horizontally *(workshop two, problem 1, group F)*

5 Conclusion

This paper discussed the importance of addressing creativity, specifically in architectural design education, from a cognitive perspective. It aimed at helping students in discovering their own way of solving problems and identifying its strengths and weaknesses. This stems from a belief that such awareness is considered as a significant step towards the improvement and development of their design thinking skills. The findings of the study, as summarized (Fig. 15)., have emphasized the positive impact of the cyclical behavior in the creative problem solving process and highlighted the different key issues and lessons emerging from students' consciousness of the mental processes that occurred during the problem solving process.

The iterative alternating nature between generation and exploration has resulted in a significant improvement in the product leading to more complex, creative and innovative solutions. Students empirically understood the importance of experimentation, play and discovery and its role in the creative problem solving process. They experienced how challenging the obvious offered more opportunities and provided new perspectives and insights to the situation. Furthermore, they witnessed and appreciated the important and vital role of collaboration, discussion and deferring judgement in developing, modifying and refining their solutions.

Based on the above analysis and discussion, the awareness and consciousness of those emergent lessons and of the cyclical nature of the creative problem solving process have assisted in developing and nurturing different personality traits. Students' self-confidence, inclination toward risk taking and tolerance for ambiguity have shown improvement and were observed blossoming throughout the process. Such traits are often critical to creativity and innovation in design.

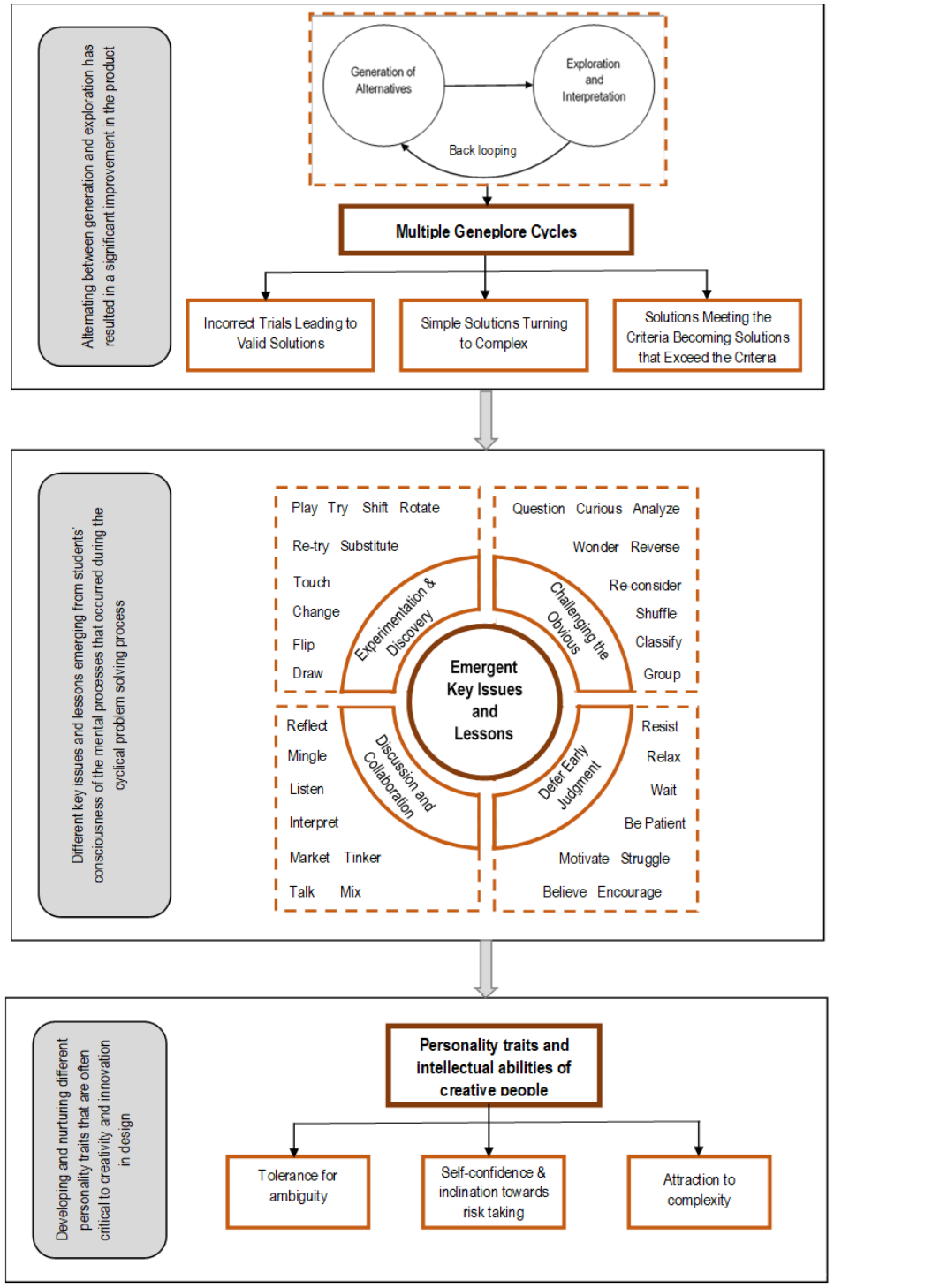


Fig.15. Summary of the core ideas and findings of the study

In conclusion, the findings of the study are expected to encourage conscious design and help students in creatively addressing ill-defined and tangled architectural design problems. Future research could extend the work presented here by exploring further strategies and approaches that couple creativity and architectural design education. More specifically, we need to encourage practical initiatives, in the educational agendas, which work on nurturing and developing the different individual skills and personality traits that often characterize creative and successful designers.

Acknowledgements: The author would like to thank her students for their spirit, contribution and level of engagement. She also extends her sincere acknowledgement to all AASTMT faculty members and teaching assistants for their continuous support and encouragement.

References

[1] A.F. Osborn. (1963). Applied Imagination: Principles and procedures of creative thinking. New York: Scribeners and Sons.
[2] Baillie, C. (2002). Enhancing creativity in engineering students. Engineering Science & Education Journal, 11(5), 185–192. https://doi.org/10.1049/esej:20020503
[3] Bourgeois-Bougrine, S., Buisine, S., Vandendriessche, C., Glaveanu, V., & Lubart, T. (2017). Engineering students' use of creativity and development tools in conceptual product design: What, when and how? Thinking Skills and Creativity, 24, 104–117. https://doi.org/10.1016/j.tsc.2017.02.016
[4] Buchanan, R. (1992). Wicked Problems in Design Thinking. Design Issues, 8(2), 5. https://doi.org/10.2307/1511637
[5] Cachia, R., Ferrari, A., Ala-Mutka, K., Punie, Y., & Institute for Prospective Technological Studies. (2010). Creative learning and innovative teaching: Final report on the study on creativity and innovation in education in the EU member states. Retrieved from http://dx.publications.europa.eu/10.2791/52913
[6] Carroll, J. M., & Thomas, J. M. (1988). FUN. ACM SIGCHI Bulletin, 19(3), 21–24. https://doi.org/10.1145/49108.1045604
[7] Cho, J. S., Hong, J., & Kwang-Soo, J. (2016). Creativity for Idea Generation in the Course of University Design Education. Journal of Integrated Design Research, 15(4), 73–81. https://doi.org/10.21195/jidr.2016.15.4.006
[8] Cross, N. (1999). Natural intelligence in design. Design Studies, 20(1), 25–39. https://doi.org/10.1016/S0142-694X(98)00026-X
[9] Edward de Bono. (1970). Lateral thinking: A textbook of creativity. London: Ward Lock Educational.
[10] Edward de Bono. (1991). The 5 day course in thinking (repr). London: Penguin.
[11] Edward de Bono. (2000). Six thinking hats (Rev. and updated ed). London: Penguin Books.
[12] Edward de Bono. (2007). How to Have Creative Ideas: 62 exercises to develop the mind (1 Aufl). London: Vermilion.
[13] Edward de Bono. (2009). I Am Right, You are Wrong: From this to the New Renaissance: From Rock Logic to Water Logic. Retrieved from https://books.google.de/books?id=HeJJPgAACAAJ
[14] Finke, R. A., Ward, T. B., & Smith, S. M. (1992). Creative cognition: Theory, research, and applications. Cambridge, Mass: MIT Press.
[15] Gardner, H. (1999). Intelligence reframed: Multiple intelligences for the 21st century. New York, NY: Basic Books.
[16] Goldschmidt, G. (1997). Capturing indeterminism: Representation in the design problem space. Design Studies, 18(4), 441–455. https://doi.org/10.1016/S0142-694X(97)00011-2
[17] Hocking, I., & Vernon, D. (2017). The right tool for the right task: Structured techniques prove less effective on an ill-defined problem finding task. Thinking Skills and Creativity, 26, 84–91. https://doi.org/10.1016/j.tsc.2017.08.001
[18] John-Steiner, V. (2006). Creative collaboration (Pbk. ed). Oxford ; New York: Oxford University Press.
[19] Kowaltowski, D. C. C. K., Bianchi, G., & de Paiva, V. T. (2010). Methods that may stimulate creativity and their use in architectural design education. International Journal of Technology and Design Education, 20(4), 453–476. https://doi.org/10.1007/s10798-009-9102-z
[20] Lau, K. W. (2017). Creativity Training in Design Education. In E. G. Carayannis (Ed.), Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship (pp. 1–8). https://doi.org/10.1007/978-1-4614-6616-1_359-2
[21] Lin, M. W. (1993). Drawing and Designing with Confidence: A Step-by-Step Guide. John Wiley & Sons.

[22] McAdam, R., & McClelland, J. (2002). Individual and team-based idea generation within innovation management: Organisational and research agendas. European Journal of Innovation Management, 5(2), 86–97. https://doi.org/10.1108/14601060210428186
[23] Michael Michalko. (1991). Thinkertoys: A Handbook of Business Creativity. California: Ten Speed Press.
[24] Paletz, S. B. F., & Peng, K. (2009). Problem Finding and Contradiction: Examining the Relationship Between Naive Dialectical Thinking, Ethnicity, and Creativity. Creativity Research Journal, 21(2–3), 139–151. https://doi.org/10.1080/10400410902858683
[25] Puja Khatri, & Sumedha Dutta. (2018). Divergent Thinking – It's Time to Change the Box! https://doi.org/10.5281/ZENODO.1884416
[26] Reiter-Palmon, R., & Illies, J. J. (2004). Leadership and creativity: Understanding leadership from a creative problem-solving perspective. The Leadership Quarterly, 15(1), 55–77. https://doi.org/10.1016/j.leaqua.2003.12.005
[27] Rittel, H. W. J. (1972). On the planning crisis: Systems analysis of the 'first and second generations'. Berkeley, Calif. : Univ. of California.
[28] Sawyer, R. K. (2006). Explaining creativity: The science of human innovation. Oxford ; New York: Oxford University Press.
[29] Schumacher, P. (2012). The autopoiesis of architecture. Vol. 2: A new agenda for architecture. Chichester: Wiley.
[30] Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In Handbook of creativity. (pp. 3–15). New York, NY, US: Cambridge University Press.
[31] Taborda, E., Chandrasegaran, S. K., Reid, T., Ramani, K., & Kisselburgh, L. (2012). Enhancing visual thinking in a toy design course using freehand sketching. Proc. ASME Des. Eng. Tech. Conf. Proceedings of the ASME Design Engineering Technical Conference, 7, 267–276. Retrieved from /z-wcorg/.
[32] Taylor, C. W. (1988). Various approaches to and definitions of creativity. The Nature of Creativity: Contemporary Psychological Perspectives., 99–121.
[33] Villalba, E. (2008). On creativity towards an understanding of creativity and its measurements. Luxembourg: Publications Office.
[34] Wallas, G. (1926). The art of thought,. New York: Harcourt, Brace and Company. /z-wcorg/.

Detection of Anomalies in the Computer Network Behaviour

Danijela Protić

MSc, Centre of Applied Mathematics and Electronics, Belgrade, Serbia

Miomir Stanković

PhD, Mathematical Institute, Serbian Academy of Science and Arts, Belgrade, Serbia

Abstract

The goal of anomaly-based intrusion detection is to build a system which monitors computer network behaviour and generates alerts if either a known attack or an anomaly is detected. Anomaly-based intrusion detection system detects intrusions based on a reference model which identifies normal behaviour of the computer network and flags an anomaly. Basic challenges in anomaly-based detection are difficulties to identify a 'normal' network behaviour and complexity of the dataset needed to train the intrusion detection system. Supervised machine learning can be used to train the binary classifiers in order to recognize the notion of normality. In this paper we present an algorithm for feature selection and instances normalization which reduces the Kyoto 2006+ dataset in order to increase accuracy and decrease time for training, testing and validating intrusion detection systems based on five models: k-Nearest Neighbour (k-NN), weighted k-NN (wk-NN), Support Vector Machine (SVM), Decision Tree, and Feedforward Neural Network (FNN).

Keywords: intrusion detection, Kyoto 2006+, k-NN, wk-NN, SVM, decision tree, FNN

Introduction

Intrusion detection systems (IDSs) protect computer networks from malicious activities which compromise network security and affect the confidentiality, integrity and availability of the data. IDSs can be grouped into the signature-based, anomaly-based, and hybrid (Ganpathy et al., 2015, 44-50).The basic idea of signature-based detection is to represent an attack in the form of pattern in such a way that any known attack and its variation can be detected. The main disadvantage of this approach is difficulty for detecting unknown attacks. Anomaly-based IDS detects changes in the network behaviour. The goal of anomaly-based detection is to build a statistical model that describes the normal behaviour of the computer network and then looks for activities which differ from the created model. It detects both intrusions and/or misuse, and classifies them as either 'normal' or 'anomaly'. The biggest challenge in anomaly-based detection is to identify what is considered normal network behaviour. Machine learning (ML) models can be trained as binary classifiers in order to recognize the notion of normality. In a supervised ML, the data have to be collected over a period of time to create a model of normal behaviour of users, hosts, and networks.

A number of records needed for training the complex computer networks can be large, which makes evaluation of the IDS computationally expensive since the processing time and memory usage rise with a size of the dataset. In anomaly-based detection some of the recorded data can be discarded to decrease time needed for training and increase accuracy of IDSs. This paper presents one pre-processing technique applied to reduce the size of the Kyoto 2006+ dataset. The proposed algorithm cuts off all categorical features, features which are intended for further analyses of the evaluated models, and features containing instances which cannot be normalized into the range [-1, 1], excluding the feature 'Label' that flags either normal network behaviour or an anomaly (Ramasamy & Rani, 2018, 1060-1067). After the pre-processing, of 24 features describing each instance nine features left. In this paper we present results on the accuracy and computation time testing for five supervised learning models: k-Nearest Neighbour (k-NN), weighted k-NN (wk-NN), Support Vector Machine (SVM), Decision Tree and Feedforward Neural Network (FNN).

1 The Kyoto 2006+ Dataset

The Kyoto 2006+ dataset is built on the three years (November 2006 to August 2009) of the real network traffic data, collected on five different computer networks inside and outside the Kyoto University. The data set is designed to provide

evaluation of the network-based intrusion detection systems (NIDS). It consists of 14 statistical features derived from the KDD Cup '99 dataset (1999) and 10 additional features which can be used for evaluation and further analyses of NIDS (Protic, 2018, 580-595). The Kyoto 2006+ dataset is captured using honeypots, darknet sensors, e-mail servers, web crawlers and other intrusion detection systems (Sing, 2014, 31-35). During the observation period, more than 50 million sessions of normal traffic, over 43 million sessions of known attacks and almost 426 thousand sessions of unknown attack were recorded (Song et al, 2011). Of the 41 features derived from the KDD Cup '99 dataset, authors discarded redundant data and content features which are not suitable for NIDS (See Table 1). In this way, the 14 features left consisted of one categorical feature 'Flag' and 13 continuous features. Moreover, authors extracted 10 additional features: 'Label' which indicated normal traffic or attacks, four features describing source and destination addresses and port numbers, two features describing start time and duration of the session, and three features for IDS, malware and Ashula detection (See Table 2).

2 Feature Selection

One of the major issues associated with the Kyoto 2006+ dataset is its size. Features selection reduces the data dimensionality by determining whether a feature is relevant or not for evaluation of the NIDS model. Using effective features in designing classifiers not only reduce the dataset but also improve performances of the classifier (Jayakumar, Revathi & Karpagam, 2015, 728-734).

In this paper we present two-step pre-processing algorithm for feature selection given as follows:

Step 1: Discard all categorical features and all features which are intended for further analyses, excluding feature 'Label'.

Step 2: Cut features containing instances which cannot be normalized into the range [-1, 1].

Of the 24 features of the Kyoto 2006+ dataset, 17 features are discarded after the first algorithm step. Nine features (5-13) left after the pre-processing is done. These features are normalized to fall into the range [-1, 1] by applying the hyperbolic tangent function given with Eq. (1):

$$\tanh(n) = \frac{2}{1+e^{-2n}} - 1 \qquad (1)$$

In this way the values of instances are scaled so that the effect of one feature cannot dominate the others. Moreover, normalized instances speed up FNN. Network training is more efficient when this pre-processing is performed on inputs. If the inputs are greater than 3 ($e^{-3} \approx 0.05$) sigmoid functions, which are used in the hidden network layer, become essentially saturated. If this happens at the beginning of the training process the gradients will be very small and the network training will be slow.

3 Machine Learning Models

Supervised ML algorithms use the known dataset to evaluate a model that generates prediction of unknown data. Assume that all data points belong either to the class 'normal' or class 'anomaly' (binary classification). Than each training data point x_i, from a vector of d-dimensional feature space, can be labelled by y_i as follows (see Eq. (2)):

$$y_i = \begin{cases} y_n, & x_i \in class_{normal} \\ y_a, & x_i \in class_{anomaly} \end{cases} \qquad (2)$$

The training dataset is denoted as follows $\{(x_i, y_i),\ i = 1, \cdots, k\}$.

This paper presents five models used for binary classification: k-NN, wk-NN (Hechenbichler & Schliep, 2004), SVM (Burgess, 1998, 283-298), Decision Tree (Sebastiani, 2002, 13) and FNN (Protic & Milosavljevic, 2006, 643-646).

3.1 k-Nearest Neighbour

k-NN is nonparametric method where new observation is placed into the class of observation from the learning set (Hechenbichler & Schliep, 2004). In this paper we present results on k-NN based prediction of unknown instances which finds the largest similarity of instances based on the Euclidean distance measure (Eq. (3)):

$$d(x_i, y_i) = \sqrt{\sum_{s=1}^{p} (x_{is} - y_{is})^2} \qquad (3)$$

3.2 Weighted k-Nearest Neighbour

The main idea of wk-NN is to extend the k-NN method in so that the observations within the learning set, which are particularly close to the new observation, should get a higher weight in the decision than such neighbours which are more distant one from that observation (Hechenbichler & Schliep, 2004). To reach this aim the distances have to be transformed into the similarity measures (Eq. (4)), which can be used as weights (Eq. (5)):

$$dist = \sqrt{\sum_{i=1}^{p} (x_i - y_i)} \qquad (4)$$

$$w = \frac{1}{dist^2} \qquad (5)$$

3.3 Support Vector Machine

The goal of SVM algorithm is to find a hyperplane that distinctly classifies the data points into various classes (Burgess, 1998, 283-298). To separate instances into the classes 'normal' or 'anomaly' the algorithm finds a linear hyperplane that has the maximum distance ρ=2/||w|| between data points of both classes. For the training set

$\mathbf{x}(x_i, y_i), x_i \in R^d, y_i \in \{-1,1\}$, the discriminant function takes the form (Eq. (6)):

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \sum_i w_i x_i + b \qquad (6)$$

where ***w*** is normal vector to the hyperplane that can be determined as follows (Eq. (7)):

$$\begin{aligned} & x_i \mathbf{w} + b \geq +1, && y_i = +1 \\ & x_i \mathbf{w} + b \leq -1 && y_i = -1 \\ & y_i (x_i \mathbf{w} + b) - 1 \geq 0, && \forall i,\ 1 \leq i \leq n \end{aligned} \qquad (7)$$

The idea is to find min||w|| which maximizes the distance *ρ*.

3.4 Decision Tree

Decision Tree prediction is based on the principle of recursive partitioning by monitoring decisions from the root to the last node (Sebastiani, 2002, 13). It is one of the graph-like algorithms which use branching methods to illustrate every possible outcome of the decisions, where nodes represent features, links represent decision rules and leafs represent the outcomes. In the experiments we applied the Iterative Dichotomy 3 algorithm (ID3) (Colin, 1996, 107-110) which calculates entropy and information gain to build a tree. Entropy is a measure which controls how the tree decides to split the data. If the target feature can take on k different values then entropy of S relative to this k-wise classification can be calculated as follows (Eq. (8)):

$$Entropy(S) = -\sum_{i=1}^{k} p_i \log_2(p_i) \qquad (8)$$

where p_i is the proportion S belonging to class i. Information gain represents expected reduction in entropy based on the decrease in entropy after the dataset is split on the feature. The feature with highest information gain will split first. Information gain can be calculated with the formula (Eq. (9)):

$$Gain(S, A) = Entropy(S) - \sum_{v \in avlues(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \qquad (9)$$

Gain(S,A) of a feature A relative to a collection of examples S provides information about the target function value, given the value of some other feature A (when A splits the set S into the subsets S_v).

3.5 Feedforward Neural Network

FNN is based on the back-propagation algorithm. The nonlinear transfer function of the FNN is given with Eq. (10):

$$y_i(\mathbf{w},\mathbf{W}) = F_i\left(\sum_{j=1}^{q} W_{ij} f_j\left(\sum_{l=1}^{m} w_{ji} x_l + w_{j0}\right) + W_{f0}\right) \quad (10)$$

where x_l are inputs, y_i are outputs, **w** and **W** are weight matrices, f_j and F_i denote transfer functions of hidden and output layers, m represents the number of inputs, q represents the number of outputs, and w_{j0} and W_{f0} denote biases (Protic & Milosavljevic, 2006, 643-646). The objective of FNN presented in this paper is to minimize output error in accordance with the Levenberg-Marquardt algorithm (Protic, 2015, 11-28). It should be pointed out that the FNN has to be large enough to improve network generalization and provide an adequate fit.

4 Results

In our experiments, performances of the models are measured based on accuracy (ACC), which represents the ratio of number of instances correctly classified to the total number of instances given with Eq. (11) (Ambedkar & Babu, 2015, 25-29):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

where true positive (TP), true negative (TN), false positive (FP) and false negative (FN) denote detected network behaviour as follows:

TP – 'anomaly' is detected as 'anomaly',

TN – 'normal' is detected as 'normal',

FP – 'normal' is detected as 'anomaly' and

FN – 'anomaly' is detected as 'normal'.

In our previous work we have presented results based on normalized and not-normalized dataset of instances and four ML models, namely k-NN, wk-NN, SVM and Decision Tree (Protic & Stankovic, 2018, 43-48). Here we present results of the experiments conducted to the normalized instances and five models: k-NN, wk-NN, SVM, Decision Tree and FNN. Accuracy of the models and the corresponding computation time (sum of training, testing and validation time) are given in Table 3. The experiments are conducted on 10-days records from the Kyoto 2006+ dataset (991.395 instances in total). All models are trained so that out of the total number of randomly selected instances 70% are used for training, 15% for testing and 15% for validation of the models. Experiments are performed using Intel(R) Core(TM) i7-2620M CPU 2.70GHz processor with 16GB RAM Installed Memory.

Results show the highest accuracy (99.5%) of wk-NN model. The results also point to high accuracy for k-NN model (99.36%). Computation time for evaluating Decision Tree model is significantly shorter comparing to the computation time for other model's evaluation, except for the FNN. Number of parameters in the network structure with nine inputs, one hidden-layer and one output is large enough to provide an adequate fit. The highest accuracy of FNN (99.21%) is achieved when the network is trained with the largest subset (136.625 instances in total). As expected, the time period needed for network learning is longer than for the networks trained with the smaller datasets (20s). Time period of training, testing and validating the FNN is significantly shorter than k-NN, wk-NN and SVM, and fall into the range [2s, 26s]. The results also show that the SVM model has a lower accuracy and longer computation time comparing to the other models.

5 Conclusion

Anomaly-based intrusion detection systems recognize deviations from normal computer network behaviour. A main challenge in anomaly-based detection is to determine the normal network behaviour and flag the anomaly. Machine learning models can be trained to classify the data into categories 'normal' or 'anomaly'. In supervised machine learning the data have to be recorded over a period of time to create a model of normal behaviour. This process can take significant time and may be computationally expensive for complex computer networks. To decrease the period of learning and increase the accuracy of the model the number of features can be significantly reduced. In this paper, we present one pre-processing technique based on the feature selection. A subset containing nine features and normalized instances is used for evaluation of IDSs based on k-NN, wk-NN, SVM, Decision Tree and FNN models. The results show that the highest accuracy gives the wk-NN model, while the shortest computation time has Decision Tree model. Overall, FNN shows higher accuracy and

shorter computation time, compared to the other models. In our further work we will present the results of experiments conducted on hybrid model based on wk-NN and FNN which detects variation in the decision on detected anomalies.

References

[1] Ambedkar, Ch. & Babu V. K. (2015). Detection of Probe Attacks Using Machine Learning Techniques. *International Journal of Research Studies in Computer Science and Engineering*. *2*(3), 25-29.

[2] Burgess, M. (1998). Computer Immunology. *12th USENIX Conference on System Administration*, Boston, MA, USA, 06-11 December 1998, 283-298.

[3] [Dataset] The UCI KDD Archive Information and Computer Science University of California, Irvine. Last modified: October 28, 1999. Retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[4] Ganpathy, S., Vijayakumar, P., Yogesh, P. & Kannan, A. (2015). An Intelligent CRF Based Feature Selection for Effective Intrusion Detection. *The International Arab Journal of Information Technology*, published online March 8, 44-50.

[5] Hechenbichler, K. & Schliep, K. (2004). Weighted *k*-Nearest-Neighbor Techniques and Ordinal Classification. *Sonderforschungsbereich* 386, Paper 399 (2004). Retrieved from https://epub.ub.uni-muenchen.de/1769/1/paper_399.pdf

[6] Jayakumar, K., Revathi, T. & Karpagam, S. (2015). Intrusion Detection using Artificial Neural Networks with Best Set of Features. *The International Arab Journal of Information Technology*. *12*(6A), 728-734.

[7] Protic, D. (2015). Feedforward neural networks: The Levenberg-Marquardt optimization and the optimal brain surgeon pruning. *Military Technical Courier*. *3*(63), 11-28.

[8] Protic, D. (2018). Review of KDD Cup '99, NSL-KDD and Kyoto 2006+ datasets. *Vojnotehnički glasnik/Military Technical Courier*, *66*(3), 580-596. DOI: 10.5937/vojteh-16670.

[9] Protic, D. & Milosavljevic, M. (2006). NNARX model of speech signal generating system: test error subject to modeling mode selection. *25th International conference on microelectronics*, 643-646.

[10] Protic, D. & Stankovic, M. (2018). Anomaly-Based Intrusion Detection: Feature Selection and Normalization Influence to the Machine Learning Models Accuracy. *European Journal of Formal Science and Engineering*. *1*(3), 43-48. ISSN 2601-8675 (Print), ISSN 2601-8683 (Online).

[11] Ramasamy, R. & Rani, S. (2018). Modified Binary Bat Algorithm for Feature Selection in Unsupervised Learning. *The International Arab Journal of Information Technology*. *15*(6), 1060-1067.

[12] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Comput.Surv. 34*(1),1-47.

[13] Singh S., Khan Z. & Saxena K. (2014). Intrusion Detection Based on Artificial Intelligence Technique. *International Journal of Computer Science Trends and Technology*. *2*(4), pp. 31-35.

[14] Song, J., Takakura, H., Okabe, Y., Eto, M., Inoue, D. & Nakao, K. (2011). Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation. In: *Proc.1st Work-shop on Building Anal. Datasets and Gathering Experience Returns for Security*. Salzburg. April 10-13, 29-36.

Tables

Table 1 First 14 features from the KDD Cup '99 dataset

No	Feature	Description
1	Duration	The length of the connection (seconds).
2	Service	The connection's server type (dns, ssh, other).
3	Source bytes	The number of data bytes sent by the source IP address.
4	Destination bytes	The number of data bytes sent by the destination IP address.
5	Count	The number of connections whose source IP address and destination IP address are the same to those of the current connection in the past two seconds.
6	Same_srv_rate	% of connections to the same service in the Count feature.
7	Serror_rate	% of connections that have 'SYN' errors in Count feature.
8	Srv_serror_rate	% of connections that have 'SYN' errors in Srv_count (% of connections whose service type is the same to that of the current connections in the past two seconds) feature.
9	Dst_host_count	Among the past 100 connections whose destination IP address is the same to that of the current connection, the number of connections whose source IP address is also the same to that of the current connection.
10	Dst_host_srv_count	Among the past 100 connections whose destination IP address is the same to that of the current connection, the number of connections whose service type is also the same to that of the current connection.
11	Dst_host_same_src_port_rate	% of connections whose source port is the same to that of the current connection in Dst_host_count feature.
12	Dst_host_serror_rate	% of connections that have 'SYN' errors in Dst_host_count feature.
13	Dst_host_srv_serror_rate	% of connections that have 'SYN' errors in Dst_host_srv_count feature.
14	Flag	The state of the connection at the time of connection was written (tcp, udp).

(Source: Song et al., 2011)

Table 2. Additional features

No	Feature	Description
1	IDS_detection	Reflects if IDS triggered an alert for the connection; '0' means any alerts were not triggered and an arabic numeral means the different kind of alerts. Parenthesis indicates the number of the same alert.
2	Malware_detection	Indicates if malware, also known as malicious software, was observed at the connection; '0' means no malware was observed, and string indicates the corresponding malware observed at the connection. Parenthesis indicates the number of the same malware.
3	Ashula_detection	Means if shellcodes and exploit codes were used in the connection; '0' means no shellcode nor exploit code were observed, and an arabic numeral means the different kinds of the shellcodes or exploit codes. Parenthesis indicates the number of the same shellcode or exploit code
4	Label	Indicates whether the session was attack or not; '1' means normal. '-1' means known attack was observed in the session, and '-2' means unknown attack was observed in the session.
5	Source_IP_Address	Means source IP address used in the session. The original IP address on IPv4 was sanitized to one of the Unique Local IPv6 Unicast Addresses. Also, the same private IP addresses are only valid in the same month; if two private IP addresses are the same within the same month, it means their IP addresses on IPv4 were also the same, otherwise are different.
6	Source_Port_Number	Indicates the source port number used in the session.
7	Destination_IP_Address	It was also sanitized.
8	Destination_Port_Number	Indicates the destination port number used in the session.
9	Start_Time	Indicates when the session was started.
10	Duration	Indicates how long the session was being established.

(Source: Song et al., 2011)

Table 3 Accuracy and computation time

No	Size of the dataset	Accuracy [%] Comp. time [s]	FNN	k-NN	wk-NN	SVM	Decision Tree
1	158572	Accuracy [%]	98.8	98.3	98.4	98.1	97.2
		Comp. time [s]	26	275.72	277.32	449.35	3.8452
2	129651	Accuracy [%]	97.67	91.8	91.8	98.3	97.3
		Comp. time [s]	20	175.84	173.32	254.3	3.3104

3	128740	Accuracy [%]	98.32	98.2	98.1	97.8	97.2
		Comp. time [s]	8	193.82	194.81	280.82	3.3033
4	136625	Accuracy [%]	99.21	99.3	99.4	99.1	98.3
		Comp. time [s]	20	194.83	194.23	217.32	8.3169
5	90129	Accuracy [%]	98.99	99.0	99.1	99.0	98.4
		Comp. time [s]	11	101.28	101.753	86.283	2.2308
6	93999	Accuracy [%]	98.12	96.5	96.5	98.0	97.5
		Comp. time [s]	7	109.25	108.77	111.83	2.2613
7	81807	Accuracy [%]	98.3	98.8	98.8	97.9	98.9
		Comp. time [s]	10	91.25	91.26	227.28	2..2615
8	57278	Accuracy [%]	99.14	99.36	99.3	99.2	99.3
		Comp. time [s]	2	42.704	43.235	33.754	1.743
9	58317	Accuracy [%]	98.97	99.1	99.2	99.1	98.9
		Comp. time [s]	3	31.714	31.738	34.234	1.7482
10	57278	Accuracy [%]	99.2	99.4	99.5	99.2	99.4
		Comp. time [s]	2	43.734	43.272	30.239	1.2901

Using a fast elitist non-dominated genetic algorithm on multiobjective programming for quarterly disaggregation of the Gross Domestic Product

Raïmi Aboudou Essessinou

Institute of Mathematics and Physical Sciences, University of Abomey-Calavi (UAC), and

Department of Statistics an Economics Studies at National Institute of Statistics and Economic Analysis

Guy Degla

Institute of Mathematics and Physical Sciences, University of Abomey-Calavi (UAC)

Abstract

This research paper we use a fast elitist multiobjective genetic algorithm to solve the new approach that we propose to quarterly disaggregating of the Gross Domestic Product (GDP) by multiobjective programming. Thus, the quarterly disaggregation of the GDP is described as a quadratic multiobjective programming problem that generalizes Denton's proportional method. The proposed approach has the advantage reduce to one the number of optimization programs to be solved. Our proposed method can be applied to the national accounts of any country that has adopted the National Accounting System. The simulation results are compared to those obtained using Denton's proportional method and these results revealed the overall performance of the multiobjective programming approach for the quarterly disaggregation of GDP. Our approach is more suitable for taking into account the links between branches of national accounts, in terms of volumes and prices of products demanded during the production process. Also, it reduces forecast error and volatility of quarterly GDP. Besides, it is worth noting that our method is a usfull step for data processing such as chain-linked measures, overlap growth techniques, seasonal adjustment and calendar effects adjustment, in time series and econometrics analysis.

Keywords: Quarterly disaggregation, quarterly national accounts, optimization, quadratic program, multiobjective programming, genetic algorithm.

1. Introduction

In economics analysis, the main aggregate indicator resulting from the quarterly national accounts (QNA) is the quarterly Gross Domestic Product (GDP). Thus, the problem is that of quarterly GDP estimation for global cyclical economics analysis. Generaly, for having quarterly GDP, the economy is divided into several branches or sectors and the quarterly accounts estimation is done per branch defined throught the System of National Accounting (SNA).

The procedures for compiling quarterly national accounts can be subdivided into direct and indirect procedures. The use of direct procedures requires the availability, at quarterly intervals, of the same data sources as those used to prepare the annual accounts, subject of course to the necessary simplifications. Direct procedures are often used in countries with sufficiently developed statistical systems. Indirect procedures is based on quaterly disagregation of passed annual national accounts and extrapolation for current quater by using mathematical or statistical techniques, with or without reference indicators. Indirect procedures are used in countries without sufficiently developped statistical systems as there is in sub-Saharan Africa. That are subdivided into two groups of methods: mathematical methods and econometric methods.

In the literature, mathematical methods and econometric methods are distinctly developed for the quarterly disagregation of national accounts. A full presentation is given in the quarterly national accounts manual [11], [12]. The main difficulty in applying econometric methodologies is the need to have data available over a long period. Thus, the mathematical methods presents the advantage for applying them even with short data series [15]. Overall, mathematical methods based on numerical approach are adapted to less developed statistical systems [15] as in most of sub-Saharan Africa countries. The quarterly disaggregation according to mathematical methods preserves the infra-annual evolution of the quarterly indicator by ensuring consistency with the annual value of the national accounts [15]. Furthermore, the most praticable mathematical

method using indicators is that developed by Denton (1971) which is presented in [11], [12] and the modified version done by Sjoberg (1982), Adriaan M. Bloem and al (2001), Cholette-Dagum (2006) can be seen in [12]. Denton method estimates quarterly accounts by proportional divisions based on the related quarterly indicator and also following a fairly smooth quarterly path based on certain criteria. Thus, the quarterly disagregation of national accounts with Denton's method becomes a minimization problem with constraints. However, it is not evident to find a global composite quaterly indicator related to the GDP in order to quarterly disagregate directly the GDP by this method.

As we can see, the traditional Denton's proportional method for the quarterly disaggregation of Gross Domestic Product (GDP) have many weaknesses related to the independent resolution of several optimization sub-problems resulting from the decomposition of GDP into branches of the economy according to the nomenclature defined for quarterly national accounts. The resolution of all these sub-problems one by one could be tedious due to the time required for individual processing. Therefore, this approach could lead to differences between the annual series of observed GDP and the GDP estimated from the quarterly accounts after aggregation due to simulation errors which occur at the level of the branches treated separately.

Based on the economy breakdown into several branches or sectors, this paper proposes a novel theoretical mathematical programming approach where the quarterly disaggregation of GDP is formalized as multiobjective optimization problem. Our paper solves the weaknesses of methods based on indirect procedures mainly Denton's proportional method for the GDP quarterly disaggregation. The overall quarterly disaggregation problem is presented using an indirect approach with indicators, leaning on Denton's proportional method. A Multi-Objective Programming for Temporal Disaggregation (MOPTD) model is therefore proposed (section 2) by generalizing Denton's proportional method. Moreover, we manage to build a non-trial hybrid algorithm based on an evolutionary fast elitist non-dominated genetic algorithm (NSGA-II) for solving the model (section 3).

After quarterly decomposition of national accounts by the MOPTD model, the derivated data can be used for some important data processing such as chain-linked measures, overlap growth techniques, seasonal adjustement and calendar effects adjustement, in order to get more consistency of quarterly time series for econometrics analysis. This paper proceeds neither chain-linked measures nor seasonal and calendar effects adjustment. The MOPTD model and the algorithm have been applied to Benin's real data of national accounts, breaking down the GDP into three sectors (primary, secondary and tertiary) (see section 4). The data were obtained from the National Institute of Statistics and Economic Analysis in short INSAE, of Benin.

2. Formulation of the Multi-Objective Programming for Temporal Disaggregation

In this research, the quarterly disaggregation of branches accounts is based on an indirect approach with indicators. The basic idea of quarterly accounts is to "adapt" the indicators to the annual accounts, estimating the statistical relationship that links in the past, the annualised indicator to the corresponding account, and assuming that this relationship observed on annual data is also valid on quarterly data. Among the indirect methods, one of them retains our attention. This is the proportional method of Denton (1971) in [11, 12]. The choice is guided on the one hand by the fact that it minimizes the prediction error of the current year, so that the forecast annual growth rate is very close to the realization and on the other hand, by its ease of implementation because it is a method adapted to short time series. Several works in the literature have used Denton's basic formulation to propose modifications, as have Cholette and Dagum (1994) in [11] and Dagum and Cholette (2006) in [12]. It should therefore be noted that if the time series available on the annul national accounts are not long enough, the econometric approach could lead to inconsistent estimators. Therefore, this research is based on Denton's proportional method, which is also implemented in several countries.

First of all, it is necessary to make a description all things we need for the formulation of the problem.

Description of the entities, cardinalities and indexing of the problem

We have the following entities :

M : the number of branches of national accounts, corresponding to the number of objective functions,

T : the number of years for national accounts observed,

The following indexes are chosen for the problem formalization:

$y \in \{1, 2, 3, \dots, T\}$, year generic index,

$i \in \{1, 2, 3, 4\}$, quarterly index,

$t \in \{1, 2, 3, \dots, 4T\}$, generic index of quarters on the *T* years' period: the indexes i and y allowing to reference an observation are restated using an operator proposed by Raimi et al. [20]

$k \in \{1, 2, 3, \dots, M\}$, generic branch index of the quarterly accounts nomenclature.

Description of the model variables

Data of the problem

Value of the annual account per branch: let $Z_{k,y}$ be the (known) value added of the branch account k , for the year *y*;

quarterly indicator: let $I_{k,t}$ be the value of the branch indicator k, for the quarter $t = 1, 2, 3, \dots, 4T$;

$I_k = (I_{k,t})_{t=1,2,..4T}$: the vector of quarterly indicator related the branch k over the entire period ;

Inter-branch interaction: we note $\bar{W}_{j,k}$ the interaction of the branch *j* on the branch k, considered as the average share of the branch k demand of product coming from the branch j : $\bar{W}_{j,k} = 1 \ \ if \ \ j = k$ and $0 \leq \bar{W}_{j,k} < 1 \ \ if \ \ j \neq k$.

Target variables

Value added per branch: let $X_{k,t}$ be the value of national account for branch k at quarter $t = 1, 2, 3, \dots, 4T$

The vector of national account value for branch k over the entire period is noted:

$X_k = (X_{k,t})_{t=1,2,..4T}$, $X_k \in \mathbb{R}^{4T}$; $for \ \ k = 1, \dots, M$;

The vector of quarterly national accounts value for all branches over the entire period is noted: $X = (X_1, \ X_2, \ X_{3,} \dots . X_M)$, $X \in \mathbb{R}^{4T \times M}$;

Quarterly Gross Domestic Product for the quarter t : GDP_t .

Relationships between variables and parameters

The target quarterly national accounts are related to the indicators (input data) through proportionality relationships.

The proportionality relationship between estimated quarterly values added and quarterly related indicators is as follow:

$$\frac{X_{k,t}}{I_{k,t}} = \frac{X_{k,t-1}}{I_{k,t-1}} + u_{k,t} \ \ ; \qquad \forall \ \ k = 1, 2, \dots, M; \ \ \forall \ t = 2, \dots, 4T \tag{2}$$

$u_{k,t}$ is the random term not explained for the quarter *t.*

The first difference in the BI ratio results in the quarterly residues

$$u_{k,t} = \frac{X_{k,t}}{I_{k,t}} - \frac{X_{k,t-1}}{I_{k,t-1}} \ \ ; \qquad \forall \ \ k = 1, 2, \dots, N; \ \ \forall \ t = 2, \dots, 4T \tag{3}$$

The sum of the squares of the quarterly residues that must be minimal is as followed

$$\sum_{t=2}^{4T} \left(\frac{X_{k,t}}{I_{k,t}} - \frac{X_{k,t-1}}{I_{k,t-1}} \right)^2 , \text{ for } \ k = 1, \ 2, \dots, M \tag{4}$$

The quarterly Gross Domestic Product is expressed by the relationship:

$$GDP_t = \sum_{k=1}^{M} X_{k,t} \ ; \qquad for \quad t = 1, 2, \dots, 4T \tag{5}$$

Annual Gross Domestic Product for the year is obtained by aggregation over the four quarters

$$GDP_y = \sum_{t=4y-3}^{4y} GDP_t \;\; ; \;\; for \;\; y = 1, 2, \dots, T \tag{6}$$

Objective functions

For the analysis with a multiobjective approach, the interaction between all branches is taken into account. In addition, it is assumed that there is an influenced additivity between sectoral residues in the determination of quarterly national accounts. Thus, the objective functions are given by :

$$f_k(X_1, X_2, \dots, X_M) = \sum_{j=1}^{M} \sum_{t=2}^{4T} \bar{W}_{j,k} \left(\frac{X_{j,t}}{I_{j,t}} - \frac{X_{j,t-1}}{I_{j,t-1}} \right)^2 \;\; ; \;\; for \;\; k = 1, 2, \dots, M \tag{7}$$

Constraints of the model

In each branch, the sum of the quarterly accounts estimated for the four quarters of a year is equal to the annual account (value added) of the branch for that year:

$$\sum_{t=4y-3}^{4y} X_{k,t} = Z_{k,y} \;\; ; \qquad \forall\, y = 1, 2, \dots, T; \; k = 1, 2, \dots, M \tag{8}$$

At each year, the total of the quarterly values added towards all branches is equal to the GDP:

$$\sum_{k=1}^{M} \sum_{t=4y-3}^{4y} X_{k,t} = \sum_{k=1}^{M} Z_{k,y} \;\; ; \qquad \forall \; y = 1, 2, \dots, T \tag{9}$$

For the variables sign, since the values of the branches annual accounts are positive (if not they can be made positive), it is also assumed that the corresponding quarterly values are positive:

$$X_{k,t} \geq 0 \;\; ; \qquad for \;\; t = 1, 2, \dots, 4T \;\; and \;\; k = 1, 2, \dots, M \tag{10}$$

With regard to maximum values, for a given year, the quarterly values of the accounts may not exceed the annual sectoral value:

$$X_{k,t} \leq Z_{k,y} \;\; ; \qquad \forall\, t = 1, 2, \dots, 4T \; ; \; k = 1, 2, \dots, M \; ; \forall\, y = 1, 2, \dots, T \tag{11}$$

Using remark 1 (a), an additional constraint, relating to the weighted averages of the BI ratios, is added to Denton's traditional equilibrium ratios. It result in :

$$\sum_{t=4y-3}^{4y} \frac{X_{k,t}}{I_{k,t}} \eta_{k,t} = \frac{Z_{k,y}}{\sum_{t=4y-3}^{4y} I_{k,t}} \;\; ; \qquad \forall\, y = 1, 2, \dots, T, \qquad k = 1, 2, \dots, M \,,$$

The weights are given by : (12)

$$\eta_{k,t} = \frac{I_{k,t}}{\sum_{r=4y-3}^{4y} I_{k,r}} \;\; ; \qquad k = 1, 2, \dots, M$$

Remark 2

If the constraints translated by the relationships (8) and (10) are satisfied then the constraint translated by the relationship (11) is satisfied. Moreover, the relationship (8) implies the relationship (9). As a result, constraints (9) and (11) are not taken into account in the theoretical model but will be taken into account in the simulation algorithms in order to reinforce the constraints for minimizing the value of the objective functions.

Definition 3 (MOPTD)

By grouping together all the elementary objective functions identified by the relationship (7) for all branches, and the constraints identified by the relationship (8), (10) and (12), the quarterly disaggregation of GDP appears as a problem formulated in the form of multiobjective programming. Thus, the multiobjective programming temporal disaggregation model proposed in this research is as follows:

$$\min_{X} \{(f_1(X), f_2(X), \dots, f_k(X), \dots, f_M(X))\} \qquad \text{(MOP)}$$

subject to.

$$X = (X_1, \quad X_2, \quad X_{3,} \dots . X_M)$$

$$X_k = (X_{k,t})_t \;\; ; \;\; t = 1, 2, \dots, 4T \; ; \text{ pour tout } \; k = 1, 2, \dots, M$$

$$-X_{k,t} \leq 0 \;\; ; \quad \forall \; t = 1, 2, \dots, 4T \; ; \; k = 1, 2, \dots, M \qquad (i)$$

$$\sum_{t=4y-3}^{4y} X_{k,t} - Z_{k,y} = 0 \;\; ; \quad \forall \, y = 1, 2, \dots, T \;\; ; \quad k = 1, 2, \dots, N \qquad (ii)$$

$$\sum_{t=4y-3}^{4y} \frac{X_{k,t}}{I_{k,t}} \eta_{k,t} = \frac{Z_{k,y}}{\sum_{t=4y-3}^{4y} I_{k,t}} \;\; ; \quad k = 1, 2, \dots, M \, , \qquad \forall \, y = 1, 2, \dots, T \qquad (iii)$$

where $GDP_t = \sum_{k=1}^{N} X_{k,t} \; ; t = 1, 2, \dots, 4T$

As it can be seen, this problem is multiobjective quadratic programming.

To reconcile the estimated quarterly national accounts with the true annual observed values, an adjustment is made to the estimated quarterly values. The adjusted value $X_{k,t}^{Adj}$ of the estimated national accounts for quarter t for branch (sector) k is given by:

$$X_{k,t}^{Adj} = \hat{X}_{k,t} - \eta_{k,t} \times Abs \left(\sum_{t=4y-3}^{4y} \hat{X}_{k,t} - Z_{k,y} \right) \times sign \left(\sum_{t=4y-3}^{4y} \hat{X}_{k,t} - Z_{k,y} \right) \qquad (13)$$

$\forall \, y = 1, 2, \dots . T$

The weights are given by: $\eta_{k,t} = \frac{I_{k,t}}{\sum_{t=4y-3}^{4y} I_{k,t}}$

$\hat{X}_{k,t}$ is the estimated value of the account for the quarter and $Z_{k,y}$ the value for the year y

$Abs\,(x)$ is the absolute value of x : $Abs\,(x) = \max\{-x, x\}$

$sign\,(x) = 1 \; if \; x > 0 \; and \; sign\,(x) = -1 \;\; if \; x < 0$.

The interest of the quarterly national accounts lies in the fact that in a current year when the annual accounts are not yet available, the quarterly accounts for that year can be estimated by extrapolation from the values of the observed quarterly indicators. Thus, starting from Denton's basic extrapolation method and that presented by Di Fonzo, T. and Marco, M. [9] in their formula (5), the relationship presented by Marco [18] through scenario 3, is adapted to the quarterly accounts by considering the values of the BI ratios of the last nine previous quarters. Assuming that m is the last year of observed annual national accounts, the extrapolated value added of the branch (sector) k for the quarter $(4m + r)$ is given by :

$$X_{k,4m+r}^{Adj} = I_{k,4m+r}\left[0.5 \times \frac{X_{k,4m+r-1}^{Adj}}{I_{k,4m+r-1}} + 0.5 \times \left(\frac{1}{9}\sum_{t=(4m+r-1)-9}^{(4m+r-1)} \frac{X_{k,t}^{Adj}}{I_{k,t}}\right)\right] \quad (14)$$

where

$X_{k,t}^{Adj}$ is the adjusted interpolation value added of the branch *k* for the quarter *t*

r is the rank of the quarter of year $m + 1$ for which the extrapolation is made,

$r = 1, 2, 3, 4.$

Proposition 1

Under the assumption of a total absence of interactions between branches, the program (MOP) is reduced to the Denton proportional method applied to each of the *M* branches of national accounts.

Proof

Let suppose that there is no interaction between the M branches of national accounts.

So we have $\bar{W}_{j,k} = 1 \;\; if \;\; j = k$ et $\bar{W}_{j,k} = 0 \;\; if \;\; j \neq k \;;\; k = 1, 2, 3, \dots, M$

Consequently, f_k becomes

$f_k(X) = \sum_{t=2}^{4T}\left(\frac{X_{k,t}}{I_{k,t}} - \frac{X_{k,t-1}}{I_{k,t-1}}\right)^2 \;; \quad \forall \; k = 1, 2, \dots, M \;\; with \; X = (X_1, \; X_2, \; X_3, \dots . X_M),$

$f_k(X)$ therefore depends only on X_k , $\forall \; k = 1, 2, \dots, M$. The problem (MOP) can therefore be written as follows:

$$\underset{X_1, X_2, X_3, \dots . X_M}{\text{Min}} \{(f_1(X_1), f_2(X_2), \dots, f_k(X_k), \dots, f_M(X_M))\} \quad \text{(MOP)'}$$

subject to

$$\begin{cases} X = (X_1, \quad X_2, \quad X_3, \dots . X_M) \\ X_k = (X_{k,t})_t \;\; ; \;\; t = 1, 2, \dots, 4T \;\; ; \text{ for all } \; k = 1, 2, \dots, M \\ -X_{k,t} \leq 0 \;\; ; \quad \forall \; t = 1, 2, \dots, 4T \; ; \; k = 1, 2, \dots, M \\ \sum_{t=4y-3}^{4y} X_{k,t} - Z_{k,y} = 0 \;\; ; \quad \forall \; y = 1, 2, \dots, T \;\; ; \quad k = 1, 2, \dots, N \\ \sum_{t=4y-3}^{4y} \frac{X_{k,t}}{I_{k,t}} w_{k,t} = \frac{Z_{k,y}}{\sum_{t=4y-3}^{4y} I_{k,t}} \;\; ; \quad k = 1, 2, \dots, M \; , \quad \forall \; y = 1, 2, \dots, T \end{cases}$$

By relaxing the constraint $\sum_{t=4y-3}^{4y} \frac{X_{k,t}}{I_{k,t}} w_{k,t} = \frac{Z_{k,y}}{\sum_{t=4y-3}^{4y} I_{k,t}} \;\; ; \quad k = 1, 2, \dots, M \; , \; \forall \; y = 1, 2, \dots, T,$

we get the following new program:

$$\underset{X_1, X_2, X_{3,}, \ldots X_M}{\text{Min}} \{(f_1(X_1), f_2(X_2), \ldots, f_k(X_{\text{k}}), \ldots, f_M(X_{\text{M}}))\} \qquad \text{(MOP)''}$$

subject to

$$\begin{cases} X = (X_1, \quad X_2, \quad X_{3,} \ldots . X_M) \\ X_k = (X_{k,t})_t \ ; \ t = 1, 2, \ldots, 4T \ ; \text{ for all } k = 1, 2, \ldots, M \\ -X_{k,t} \leq 0 \ ; \quad \forall \ t = 1, 2, \ldots, 4T \ ; \ k = 1, 2, \ldots, M \\ \sum_{t=4y-3}^{4y} X_{k,t} - Z_{k,y} = 0 \ ; \quad \forall \, y = 1, 2, \ldots, T \ ; \quad k = 1, 2, \ldots, M \end{cases}$$

Since the objective functions have independent arguments, it is possible to optimize the f_k separately to determine the $X_k = (X_{k,t})_t$ for $k = 1, 2, \ldots, M$, so that we can have the GDP_t by summing the $X_{k,t}$, using relationship (5).

This means solving all the elementary optimization program $(EOP)_k$:

$$\underset{X_k}{\text{Min}} \{f_k(X_{\text{k}})\} \qquad (EOP)_k$$

subject to

$$-X_{k,t} \leq 0 \ ; \quad \forall \ t = 1, 2, \ldots, 4T \ ;$$

$$\sum_{t=4y-3}^{4y} X_{k,t} - Z_{k,y} = 0 \ ; \quad \forall \, y = 1, 2, \ldots, T$$

It can be seen that the program $(EOP)_k$ is similar to Denton 's program presented in [11], illustrating the Denton proportional method applied to branch $k = 1, 2, \ldots, M$ of national accounts.

Thus, the program (MOP) is reduced to Denton's proportional method applied successively to the *M* branches to deduct quarterly GDP from the previous relationship (5). ■

Remark 3

From the results of proposition 1, it can be deduced that the proposed MOPTD model is a generalization of Denton's proportional method.

3. Resolution approach and algorithm design

The problem is analysed on the basis of a hypothesis of cooperation between the objective functions. On this basis, Pareto optimal solutions are prospected in solving the problem. The following definition provides an understanding of the Pareto optimality concept.

3.1. Preliminary definition ([19])

Let's consider the vector function $F(x) = \left(f_1(x), f_2(x), \ldots, f_Q(x)\right)$ and U the constrained space (space of feasible solutions) of the problem (MOPP). It should be noted $F(U)$ the value space of the objective functions. We define on $F(U)$ a partial relationship. Let K be any blunt cone such that $K \subset \mathbb{R}^Q$. Let's consider the binary relationship $\preccurlyeq_K$ indexed by K defined as follow:

$a \preccurlyeq_K b \Leftrightarrow (b - a) \in K.$

Since it is not possible to find a solution that simultaneously optimizes all objective functions in the case of a multiobjective program, the notion of dominance in the sense of Pareto is used.

(i)- Pareto-dominance: for two feasible decision vectors x et y, we say that x dominates y in the sense of Pareto and we note $(x,\ F(x)) \preccurlyeq_K (y,\ F(y))$, if and only if $for\ all\ q \in \{1, 2, 3, \dots, Q\}$, $f_q(x) \leq f_q(y)$ and $\exists\ q_0 \in \{1, 2, 3, \dots, Q\}$ such that $f_{q_0}(x) < f_{q_0}(y)$.

(ii)- Pareto optimal solution: a solution $x \in \mathbb{R}^N$ is called Pareto optimal in $\mathbb{R}^N$ if and only if, there is no vector $y \in \mathbb{R}^N$ which dominates x.

(iii)- The Pareto optimal set is defined as the set of all Pareto optimal solutions.

(iv)- The Pareto optimal front is defined as the set of all objective function values corresponding to all solutions in the Pareto optimal set.

The resolution algorithms of the traditional quarterly disaggregation methods are presented in [1], [2]. In this paper, the simulation method for the model resolution is based on the NSGA-II algorithm developed in the literature for multiobjective optimization [3], [4], [8], [21].

3.2. A brief overview of the NSGA-II algorithm

The NSGA algorithm called "Non-dominant Sorting Genetic Algorithm"[10, 20], after several years of implementation, has become ineffective in ranking individuals [14] especially for large problems treated. To correct this deficiency, it elitism was introduced into the basic algorithm [8] in order to preserve the best solutions from generation to generation [14]. Thus, the improved version of the NSGA algorithm, called NSGA-II, classify the individuals into many levels [6], this classification does not require necessary the choice of sharing parameter [4]. The NSGA-II algorithm is illustrated by the pseudo code given in algorithm 1. It should be noted that elitism in the NSGA-II algorithm contributes to accelerating the rate of convergence and the overall performance of the genetic algorithm incorporated in it.

Algorithm 1 (pseudo-code NSGA-II) [8, 14]

While (total number of iterations not completed)

Generation of the initial population

Repeat

While (Population is not classified) **do**

Evaluation of all borders

Normalization of areas of constraint violation

Search for undominated individuals

Replacement of individuals

End while

Selection

Crossover

Polynomial mutation

Until (Criteria for shutdown achieved)

Recombination of optimal Pareto solutions

End while

3.3. The proposed MOPTD algorithm designing

The NSGA-II adapted to the problem of multi-target optimization for the quarterly disaggregation of *GDP* resulted in a fast elitist multiobjective programming algorithm for temporary disaggregation (MOPTD-NSGA-II). The choice of the NSGA-II

algorithm as the central nucleus is justified by the fact that elitist algorithms allow better results to be obtained on multiobjective problems [5]. The pseudo code of the MOPTD -NSGA-II algorithm is given in Algorithm 5. The interpolation of the accounts by the model takes into account the reconciliation of the quarterly accounts with the true annual values observed. This reconciliation is done by an adjustment procedure of which the pseudo-code is given in algorithm 3. In addition, the MOPTD model allows quarterly accounts to be extrapolated if necessary using the extrapolation procedure of which the pseudo code is given in algorithm 4.

In the NSGA-II algorithm, the initial population is generated from the bounded values (minimum and maximum) of the target variables. Since the problem (MOP) is relatively of large size, to ensure rapid convergence of the algorithm towards Pareto optimal solutions, the solution search domain has been narrowed using the annual data provided on account variables. Thus, the variable bounded values were generated from the annual values of the accounts by releasing the constraints (i) of the problem (MOP). The pseudo code of the limit value calculation procedure is given in algorithm 2.

As it can be seen, the NSGA-II procedure introduces a hazard when generating the reference population. Since the search for Pareto optimal solutions is based on this reference population and the Pareto optimal solution does not strictly achieve the minimum of each objective function, an adjustment procedure has been completed in order to reconcile the estimated quarterly values with the true observed annual values. The relationship (13) presented in the theoretical description is used. The pseudo-code of this adjustment procedure is given in algorithm 3.

Algorithm 2 (pseudo code of the Procedure Boundary_values_Calculation (abX, ny))

BEGIN (Boundary_values_Calculation)

**/Initialization of the values of the annual accounts (all branches/sectors))*

For k = 1 : M

For y = 1 : lyear

abX(k,y) ← annual values

end for

End for

**/Calculation of boundary values for quarterly accounts*

For k =1 : M

qbX_max (k) ← [ ] */ empty box

qbX_min (k) ← [ ] */ empty box

For y = 1 : ny

qbXmax ← [ ]

qbXmin ← [ ]

For t = 4y-3 : 4y

qbX_max (k,t) ← abX(k,y)/4+[SD (abX(k))]/(lyear-fyear+1)

qbX_min (k,t) ← abX(k,y)/4-[SD (abX(k))]/(lyear-fyear+1)

qbXmax ← [qbXmax qbX_max (k,t)]

qbXmin ← [qbXmin qbX_min (k,t)]

t ← t+1

End for

qbX_max (k) ← [qbX_max (k) qbXmax]

qbX_min (k) ← [qbX_max (k) qbXmin]

y ← y+1

End for

qbX_M(k) ← qbX_max (k, 1:nq)

qbX_m(k) ← qbX_min (k, 1:nq)

k ← k+1

End for

tXmax ← [qbX_M(1) qbX_M(2) qbX_M(3) ……. qbX_M(M)]

tXmin ← [qbX_m(1) qbX_m(2) qbX_m(3) ……. qbX_m(M)]

END (Boundary_values_Calculation)

Algorithme 3 (pseudo code of the procedure Adjustment (qX,ny,nq))

BEGIN (adjustment procedure)

Initialization

vector of the abX annual accounts

vector of quarterly accounts qX

vector of quarterly indicators qbI

For k = 1 : M

qX(k)_Adj ← [ ]

For cy = 1 : ny

fcount ←4*cy-3;

lcount ←4*cy;

qX_a(k, cy) ← [ ]

For j = fcount : lcount

w(k,j) ← qbI(k,j) /sum(qbI(k,j), j, fcount :1: lcount)

c(1, cy) ← sum(qX(k,j), j, fcount :1: lcount) - abX(k, cy)

If c(1, cy)>=0

qx_a(k, j)= qX(k, j)- w(k,j) * abs(c(1, cy))

Else

qx_a(k, j)= qX(k, j)+ w(k,j) * abs(c(1, cy))

End if

qX_a(k, cy) ← [qX_a(k, cy) qx_a(k, j)]

j ← j+1

End for

qX(k)_Adj ← [qX(k)_Adj qX_a(k, cy)]

cy ← cy+1

End for

k← k+1

End for

END (adjustment procedure)

The extrapolation of quarterly accounts for future quarters is based on the procedure described by the relationship (14) presented in the theoretical description. The pseudo code of the extrapolation procedure is given in algorithm 4.

Algorithm 4 (pseudo code of the procedure Extrapolation (qX_entry, qbl_new, nq ,s))

BEGIN (extrapolation procedure)

s ← 4 x (lyear-wyear)

While (s>0) do

Call for data on quarterly indicators I(1:nq+s)

For k = 1 : M

Colling of the vector I(k, 1:nq+s)

X(k, 1:nq) ←qX_adjus(k, 1:nq)

X(k) ← X(k, 1:nq)

X(k, nq+1:nq+s) ← [ ]

For r = 1 : s

X(k, nq+r) ← I(k, nq+r)* (0.5* X(k, nq+r-1) / I(k, nq+r-1)

+ 0.5*sum(X(k, t) / I(k, t), (nq+r-1):1: (nq+r-1)-9)/9)

X(k, nq+1:nq+s) ← [X(k, nq+1:nq+s) X(k, nq+r)]

r ← r+1

End for

X(k) ← [X(k) X(k, nq+1:nq+s)]

k ← k+1

End for

Recovery and backup of the matrix [X(1) X(2) X(3) …… X(M)]

End while

END (extrapolation procedure)

Finally, the pseudo code of the complete algorithm (MOPTD-NSGA-II) is given in algorithm 5 which is as follows :

Algorithme 5 (main pseudo code of MOPTD-NSGA-II)

BEGIN (main Algorithm)

**/Initialization time parameters*

fyear */ first year of the observation period of the annual accounts

lyear */ last year of the observation period of the annual accounts

wyear */ last year considered for interpolation

/ Calculation of all other associated parameters

wy ← (wyear - fyear +1) */ number of work years for interpolation

wq ← 4*wy */ number of quarters for the calculation period

s ← 4 x (lyear-wyear) */ number of quarters for extrapolation

If lyear > wyear **then**

ny← wy; */ number of years for interpolation

nq← wq; */ number of quarters for calculations

else

ny ← (lyear - fyear +1) */ number of years for interpolation

nq← 4*ny */ number of quarters for calculations

End if

Boundary_values_Calculation (abX, ny) */ Execution of the Procedure

/ Interpolation of accounts on quarterly indicators

Begin (interpolation)

- Initialization of the other NSGA-II input parameters

M ← number of objective functions

p1← number of constraints

V← nq×M */calculates the number of variables

Pop_size ← give the size of the population

run ← the number iteration

gen_max ← give the maximum number of generations

- Calling up the problem (objective functions and constraints))

Begin (NSGA)

Execution of the NSGA-II algorithm

End (NSGA)

/ Recovery of results provided by NSGA

If run==1

qX ← [new_pop(:,1:V)] */ Pareto optimal solution

Else

qX ← [pareto_rank1(1:V)] */ Pareto optimal solution

End if

/ Adjustment of quarterly accounts

For k = 1 : M

Adjustment (qX(k), ny, nq) */ Execution of the procedure

qX(k)_Adj ← Adjustment (qX(k), ny, nq)

k ← k+1

End for

End (interpolation)

**/ Extrapolation of accounts to the quarters of the following year*

If (s > 0)

Calling up indicator values qbl(1:nq+s)

Begin (extrapolation)

For k = 1 : M

Extrapolation (qX(k)_Adj, qbl(k), nq ,s) */ Execution of the procedure

qX1_fwd(k) ← Extrapolation (qX(k)_Adj, qbl(k), nq ,s)

k ← k+1

End for

End (extrapolation)

En if

END (main algorithm)

4. Application to Benin's national accounts

Simulations are carried out using the databases of the National Institute of Statistics and Economic Analysis (INSAE). The data were subject to prior statistical processing before proceeding to the actual simulations. The GDP is decomposed into three sectors. The results obtained with the proposed model were compared with those obtained using Denton's proportional method applied separately to each of the three sectors.

4.1. Statistical data and processing

This section presents the data sources we used and the processing carried out on data related to national accounts and related indicators.

The databases available at the National Institute of Statistics and Economic Analysis (INSAE) reveals that Benin has the series of annual national accounts (ANA) from 1999 to 2015, compiled according to the System of National Accounting (SNA) 93 with 2007 as the base year, as well as various (quarterly) indicators on economic activity. These data were therefore used for the simulation exercise. It should be noted that work undertaken by the INSAE national accounts department to move from SNA 93 to SNA 2008 [10], with a base year change to 2015, has not yet been completed to make available the new series on the definite accounts over the simulation period.

In the context of this research, the quarterly GDP obtained through the quarterly disaggregation of the values added. Thus, the proposed approach using three indicators is different from that of Abdelwahed Trabelsi and Leila Hedhili [1] for which GDP is quarterly in aggregate form with only one indicator, the industrial production index.

Thus, the resolution approach adopted is that relating to the research for efficient points (Pareto optimum) after problem reduction (MOP) with three objectives.

4.2. Our method results and comparison with Denton's proportional method

Then, several tests were performed on the key parameters of the algorithm in order to retain values that reduce the convergence time of the algorithm. Simulation results and statistics on the quality of the model are presented below. In the literature, some softwares are adopted for the quarterly disaggregation of national accounts using traditional methods. As part of this research, all simulations are carried out using the OCTAVE software.

Pareto optimality front

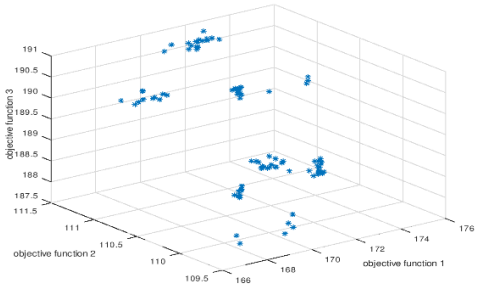
The NSGA-II algorithm uses the following key parameters as input parameters: population size (*pop_size*), iteration number (*no_rum*) and maximum number of generations (*gen_max*) beyond which the algorithm stops. The choice of the values of these parameters depends on the size of the problem. In the literature, problems are tested with the NSGA-II algorithm for minimum required values and set at 20 for *pop_size* and 5 for *gen_max* [4]. Sometimes, large values can be set: up to 200 for *pop_size*, 10 for *no_rum* and 1000 for *gen_max* in the case of problems with two objectives [8].

Based on these findings, several values were tested for the parameters *pop_size as* in [14]*, no_rum, and gen_max*, in this research. These tests have been illustrated by different representations of the Pareto front. The Pareto front obtained for the main values of (*pop_size, no_rum, gen_max*) are presented in Figure 1. The Pareto optimal points are represented in blue.

The analysis of the results shows that if *no_run*=1, the simulation gives several Pareto optimal points but the number of points increases with the population size (*pop_size*) and this regardless of the *gen_max* value, this situation is illustrated by the panels (a1) (b1), (c1) and (c2). It was used in the following *pop_size*=100 simulations, as adopted in [6] and [21].

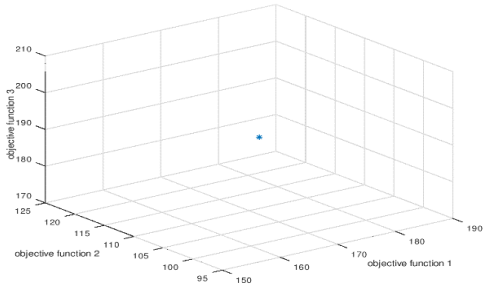
a1- (Pop_size, No_runs, gen_max)=(100,1,25)

Elapsed time=214.539 seconds



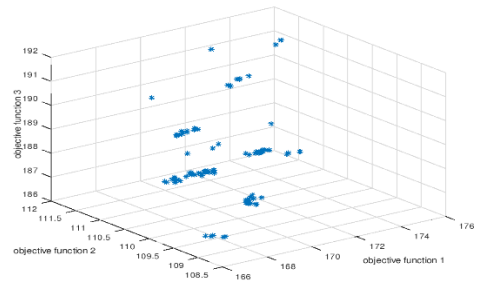
a2-(Pop_size, No_runs, gen_max)=(100,2,25)

Elapsed time=447.526 seconds



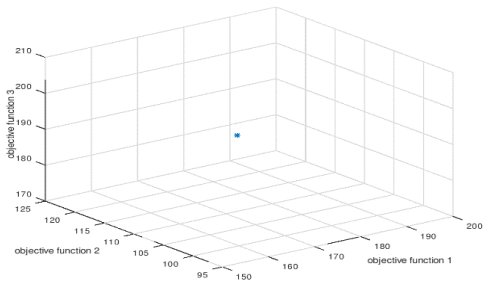
b1-(Pop_size, No_runs, gen_max)=(100,1,100)

Elapsed time=858.307 seconds



b2-(Pop_size, No_runs, gen_max)=(100,2,100)

Elapsed time=1645.12 seconds



c1-(Pop_size, No_runs, gen_max)=(10,1,100)

Elapsed time=124.319 seconds

c2-(Pop_size, No_runs, gen_max)=(10,1,25)

Elapsed time=61.8776 seconds

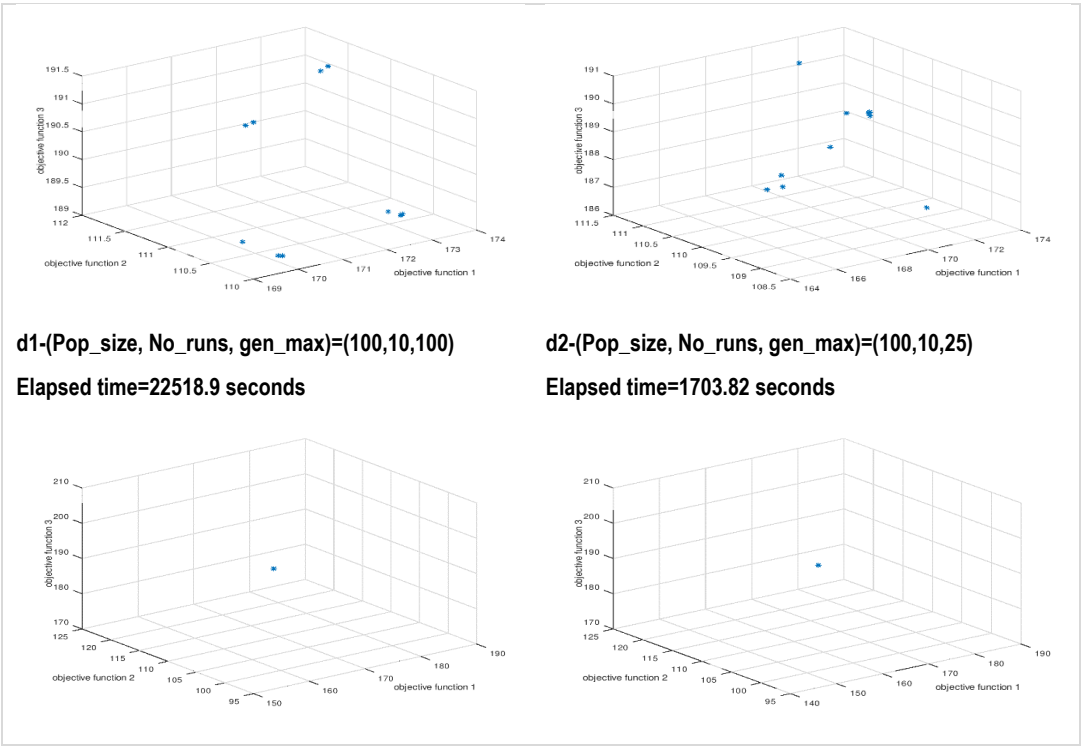


Figure 1 : Optimal pareto boundary for different simulation parameters

Source: *INSAE-data base,* Author's simulation

The Pareto boundaries presented in the panels (a2), (b2), (d1) and (d2) indicate that the algorithm converges to a single Pareto optimal point when *no_run*=100 for any *gen_max* value with *pop_size*=100, but the boundaries obtained with *gen_max*=25 and *gen_max*=100 have the same configuration; the points obtained for the different cases are all located in practically the same restricted space.

It should be noted that overall, the gen_max values set at 25 and 100 give Pareto front with the same characteristics; however, the runtime of the algorithm for displaying results is relatively longer for gen_max=100 than for gen_max=25. Thus, based on the situations presented above, the simulations which of results are presented below are performed with (pop_size, no_rum, gen_max) = (100, 10, 25).

Comparison of the evolution of interpolated series

For the comparison of the model results with those obtained using Denton's proportional method, MOPTD simulations are performed with the parameter (pop_size, no_runs, gen_max) = (100,10,25). Figure 2 below presents the results of both methods. The series of quarterly national accounts represented in panels (a), (b), (c) and (d) are those obtained after simulations are made over the period 1999-2014 and extrapolated for the quarters of 2015. The quarterly indicators are represented by red dotted lines. The series of quarterly national accounts obtained with the model are represented in black and those resulting from the application of the Denton's proportional method are the blue curves.

As shown in the results presented in Figure 5, the quarterly national accounts series obtained by both methods and the respective quarterly indicators maintain the trends and relationships obtained with the annual data. However, the analysis of the curves presented in the four panels shows that the MOPTD model produces relatively much smoother (less fluctuating) series compared to the series obtained using Denton's proportional method.

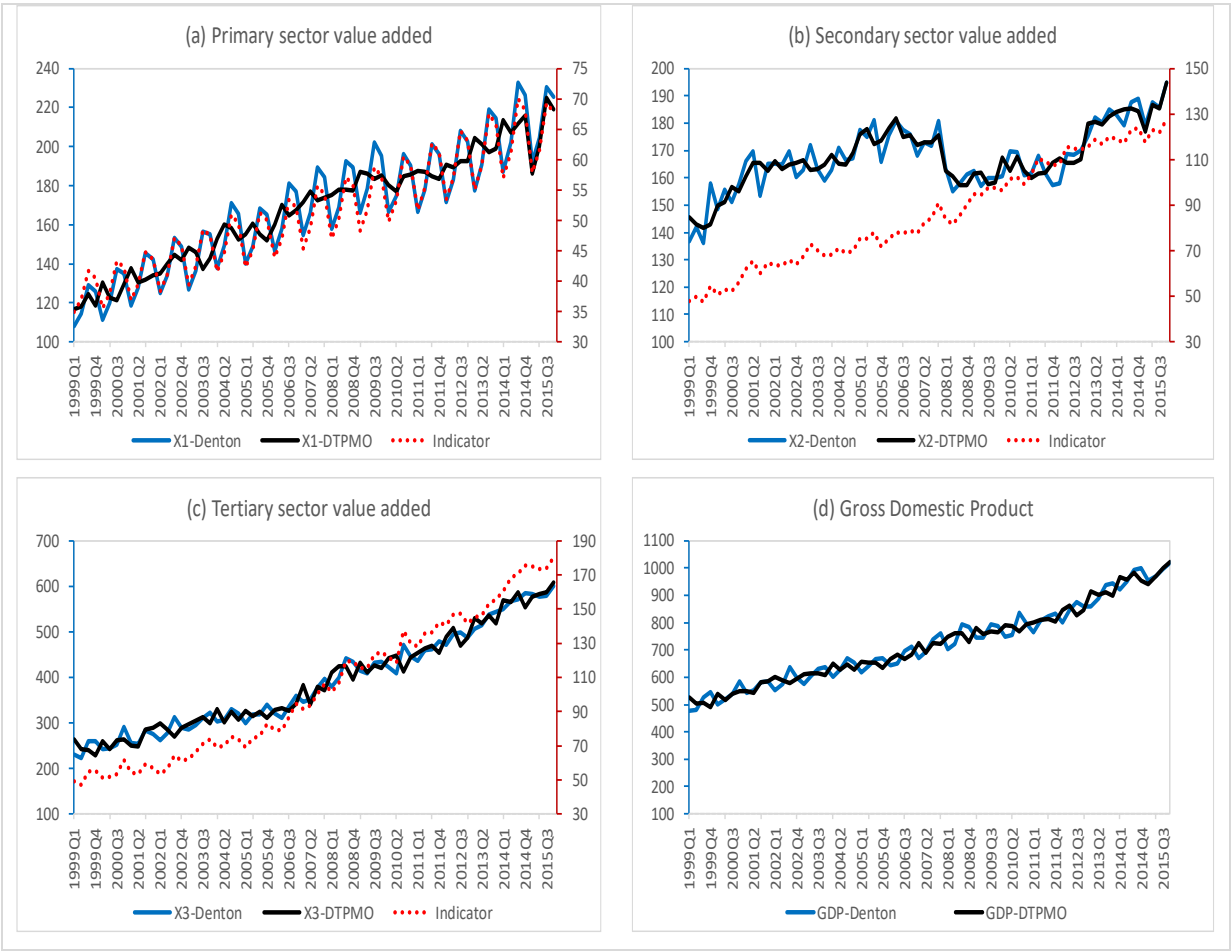


Figure 2 : Comparative evolution of quarterly accounts and indicators

Source: *INSAE-data base,* Author's calculations

The performance test of the MOPTD model compared to Denton's proportional method is performed in two steps.

The first step in performance analysis is based on a comparison of the level of GDP and growth rate obtained by extrapolation for the four quarters of 2015, after interpolation over the period 1999-2014. Table 1 summarizes the interpolation results compared to the true observed values and Table 2 summarizes the extrapolation results for 2015 in terms of values and growth rates. As the results in Table 1 show, both methods of quarterly disaggregation of GDP replicate exactly the actual observed account values for the retrospective periods (1999-2014).

With regard to the forecast for 2015, as shown in Table 2, for each sector, the annual values obtained by aggregating the quarterly national accounts are different from the real values observed for 2015. The order of magnitude of the difference in values is random, according to an analysis covering the sectors, on the one hand, and the methods, on the other hand. However, by the rule of additivity of accounts and global equilibrium, both methods give the GDP annual values relatively close to the value observed for 2015.

	Observed annual accounts				Annualized quarterly accounts-DENTON				Annualized quarterly accounts-DTPMO			
Year	X1	X2	X3	GDP	X1	X2	X3	GDP	X1	X2	X3	GDP

1999	477.64	572.96	975.02	2025.63	477.64	572.96	975.02	2025.63	477.64	572.96	975.02	2025.63
2000	503.60	612.75	1027.94	2144.28	503.60	612.75	1027.94	2144.28	503.60	612.75	1027.94	2144.28
2001	534.31	654.22	1070.11	2258.64	534.31	654.22	1070.11	2258.64	534.31	654.22	1070.11	2258.64
2002	561.24	659.70	1142.57	2363.51	561.24	659.70	1142.57	2363.51	561.24	659.70	1142.57	2363.51
2003	574.42	657.29	1213.19	2444.90	574.42	657.29	1213.19	2444.90	574.42	657.29	1213.19	2444.90
2004	623.90	667.25	1262.06	2553.20	623.90	667.25	1262.06	2553.20	623.90	667.25	1262.06	2553.20
2005	622.49	699.62	1274.83	2596.94	622.49	699.62	1274.83	2596.94	622.49	699.62	1274.83	2596.94
2006	662.47	710.03	1326.86	2699.36	662.47	710.03	1326.86	2699.36	662.47	710.03	1326.86	2699.36
2007	694.50	693.48	1472.97	2860.95	694.50	693.48	1472.97	2860.95	694.50	693.48	1472.97	2860.95
2008	708.50	637.89	1654.65	3001.04	708.50	637.89	1654.65	3001.04	708.50	637.89	1654.65	3001.04
2009	741.95	639.36	1689.33	3070.64	741.95	639.36	1689.33	3070.64	741.95	639.36	1689.33	3070.64
2010	727.59	660.37	1747.60	3135.56	727.59	660.37	1747.60	3135.56	727.59	660.37	1747.60	3135.56
2011	742.19	648.71	1837.60	3228.49	742.19	648.71	1837.60	3228.49	742.19	648.71	1837.60	3228.49
2012	765.07	664.75	1954.00	3383.82	765.07	664.75	1954.00	3383.82	765.07	664.75	1954.00	3383.82
2013	801.76	722.61	2102.79	3627.16	801.76	722.61	2102.79	3627.16	801.76	722.61	2102.79	3627.17
2014	848.41	738.97	2274.56	3861.93	848.41	738.97	2274.56	3861.93	848.41	738.97	2274.56	3861.93
2015*	**783.89**	**851.19**	**2303.51**	**3938.59**	**852.63**	**747.52**	**2340.89**	**3941.05**	**830.99**	**743.72**	**2355.94**	**3930.66**

(*) the account values for this year in the case of Denton and MOPTD are those extrapolated

Table 1 : Comparison of the values of the annual accounts actually observed and aggregated after quarterly reporting

The growth rates resulting from the forecast are also close to the GDP growth rate observed in 2015. The actual observed rate is 1.98% while for the Denton proportional method, the growth rate obtained is 2.05% showing 0.06 point of overrun in percentage.

	Annual national accounts				Annual national accounts growth rate			
	X1	X2	X3	GDP	X1	X2	X3	GDP
Real account value	**783.89**	**851.19**	**2303.51**	**3938.59**	**-7.60%**	**15.19%**	**1.27%**	**1.98%**
	One year-ahead prediction							
Denton	**852.63**	**747.52**	**2340.89**	**3941.05**	**0.50%**	**1.16%**	**2.92%**	**2.05%**
DTMPO								
Simulation 1	830.99	743.72	2355.94	3930.66	-2.05%	0.64%	3.58%	1.78%
Simulation 2	827.30	745.18	2370.19	3942.66	-2.49%	0.84%	4.20%	2.09%
Simulation 3	828.01	744.34	2368.94	3941.29	-2.40%	0.73%	4.15%	2.05%

MEAN	**828.77**	**744.41**	**2365.02**	**3938.20**	**-2.32%**	**0.74%**	**3.98%**	**1.97%**

Note : the MOPTD results are those of simulation 1 with (Pop_size, No_runs, gen_max)=(100,10,25)

Table 2 : Summary of extrapolated GDP for 2015

With the MOPTD method, the direction of variation of the GDP growth rate differential obtained by extrapolation, compared to that really observed, is mixed. Indeed, in order to test the robustness of the MOPTD model, the simulation was carried out successively three times with the same parameters. But the random nature conferred on the NSGA-II algorithm induced different values of the *GDP* extrapolated for 2015. The *GDP* growth rate extrapolated for 2015 from these three simulations is set at 1.78%, 2.09% and 2.05% respectively, giving an average of 1.97% which is close to the real level observed. This reflects that the MOPTD model is relatively robust.

The second step of performance analysis of the MOPTD model is based on the analysis of statistics on the quality of quarterly series adjustments by interpolation and extrapolation. To do this, the mean coefficient of variation which referred to the mean relative standard deviation (*MSD*), the mean absolute error (*MAE*) and the mean absolute revision (*MAR*) were calculated for the comparison of the performance of the MOPTD model with that of Denton's proportional method.

The standard deviations (*SD*) of the interpolated quarterly accounts over the period 1999-2014 are calculated in order to compare the volatility of the quarterly value added series *X1, X2, X3* and *GDP*. The *MSD* resulting from the quarterly *GDP* adjustment operation is given by the expression (15) :

$$MSD = 0.5 \times \left(\sum_{k=1}^{3} SD_{X_k} \bar{w}_k \right) + 0.5 \times SD_{GDP} \tag{15}$$

with $\bar{w}_k$ the average relative share (over the period 1999-2015) of sector k value added out of *GDP*

$$SD_Y = \frac{\sum_{t=1}^{4(T-1)} (Y_t - \bar{Y})^2}{\bar{Y}} \; ; \; \bar{Y} \; being \; the \; average \; of \; the \; series \; Y$$

With regard to forecast quality, the analysis of predictive power for each of the two quarterly methods is based on the formula proposed by Marco [18] for calculating the average absolute error. For each series extrapolated to a subsequent year, the estimated prediction error (*E*) is estimated by calculating the relative difference (in percentage) between the extrapolated value and the true value [18]. In practice, the estimated prediction error of the *Y* series is given by:

$$\hat{E}_{Y,T} = \frac{\hat{Y}_T - Y_T}{Y_T} \times 100$$

$\hat{Y}_T$ being the estimated value of the series Y (representing either *X1, X2, X3* or *GDP*) by extrapolation for year T (last accounting year). Thus, since the relative difference between the extrapolated and observed of *GDP* values is not obtained by additivity of errors from the sectors in the *GDP* calculation, the formula (15) proposed by Marco [18] has been modified and the *MAE* in predicting quarterly *GDP* is given by the relationship :

$$MAE = 0.5 \times \left(\frac{1}{3} \sum_{k=1}^{3} |\widehat{E}_{X_k,T}| \right) + 0.5 \times |\widehat{E}_{GDP,T}| \tag{16}$$

In addition, as presented above, the extrapolated values of the account series are different from the true observed values. The annual and quarterly national accounts are therefore subject to revision, due to extrapolation errors, seasonal adjustment or revision of quarterly indicators [16], once the temporary or final annual national accounts are available. As mentioned by Marco [18], it is desirable that the impact of short-term movements be limited so that the quarterly disaggregation method minimizes the size of revisions. Thus, the revision effect is captured by the mean absolute revision (*MAR*) of the last three years of quarterly changes [18]. The proposed formula for the calculation of *MAR* is a modification of the one used by Marco [18] and is presented in the following relationship (17):

$$MAR = 0.5 \times \sum_{k=1}^{3} \overline{w}_k \left[\left(\frac{1}{12} \sum_{t=4(T-2)+1}^{4(T+1)} \left| \frac{\hat{X}_{k,t}^{(T+1)}}{\hat{X}_{k,t-1}^{(T+1)}} - \frac{\hat{X}_{k,t}^{(T)}}{\hat{X}_{k,t-1}^{(T)}} \right| \right) \right]$$
$$+0.5 \times \frac{1}{12} \sum_{t=4(T-2)+1}^{4(T+1)} \left| \frac{\widehat{PIB}_t^{(T+1)}}{\widehat{PIB}_{t-1}^{(T+1)}} - \frac{\widehat{PIB}_t^{(T)}}{\widehat{PIB}_{t-1}^{(T)}} \right| \quad (17)$$

where the indexes $(T+1)$ and (T) by exponent represent the last years of observation used; $\hat{X}_{k,t}$ refers to the value added of the sector k for the quarter t.

The Table 3 presents statistics on the *MSD*, *MAE* and *MAR* for sector value added and *GDP*. The resulting average values are in bold. The analysis of the results exposed in Table 3 shows overall that the MOPTD model performs relatively better than Denton's proportional method in the quarterly disaggregation of GDP process. Indeed, when reading Table 3, the MOPTD model minimizes the overall *MAE* and *MSD* statistics (absolute mean). However, Denton's proportional method minimizes the size of revisions *MAR*.

		Prediction error (%)	*Standard Deviation (%)*	*Absolute Revision (%)*
Primary sector value added (X1)				
	Denton	8.77%	18.17%	1.10%
	DTPMO	5.72%	16.33%	3.77%
Secondary sector value added (X2)				
	Denton	-12.18%	6.69%	1.79%
	DTPMO	-12.54%	6.20%	4.07%
Tertiary sector value added (X3)				
	Denton	1.62%	26.58%	0.22%
	DTPMO	2.67%	26.42%	4.27%
Gross Domestic Product (GDP)				
	Denton	0.06%	18.97%	0.05%
	DTPMO	-0.01%	18.58%	3.17%
Mean Absolute				
	Denton	**2.94%**	**19.41%**	**0.43%**
	DTPMO	**2.88%**	**18.90%**	**3.64%**

Note : the DTPMO results are those of simulation 1 with (Pop_size, No_runs, gen_max)=(100,10,25)

Table 3 : Summary of statistics in relation to the accounting forecasting exercise

6. Conclusion

The estimation of quarterly GDP using an indirect approach based on quarterly indicators is often done by quarterly disaggregation of the accounts of the various branches that make up GDP. This approach leads to the resolution of several optimization programs. On retrospective data, the quarterly disaggregation of national accounts, based on branch-by-branch disaggregation for the determination of quarterly GDP, does not take into account the link between branches of

national accounts in the production process. This research presented a brand new approach to quarterly disaggregate GDP broken down into several branches or sectors that takes into account the links between branches and generalizes the Denton proportional method.

The proposed approach describes the method of quarterly disaggregation of GDP as a quadratic multiobjective programming. In applying the model to Benin's national accounts for which GDP is broken down into three sectors (primary, secondary and tertiary), the resolution was made using a multiobjective evolutionary algorithm using dominance in the Pareto sense. Since the theoretical formulation is a generalization of Denton's proportional method, the simulation results of the MOPTD model are compared with those obtained by applying Denton's proportional method on the value added of each sector.

It should be noted that the model proposed in this research paper is better suited for the quarterly disaggregation of GDP. Indeed, the quarterly GDP obtained by the MOPTD model is less volatile than that obtained by Denton's proportional method. In addition, on an absolute average, the forecast errors are small for the model compared to that of Denton proportional. However, it should be noted that the Denton method gives small revision errors for series compared to the MOPTD model. This situation could be explained by the random nature of the model's resolution algorithm.

Finally, this research shows that the extrapolations of GDP using the Denton proportional method and those made using the multiobjective programming approach produce similar results when the quarterly indicators used are strongly correlated to the annual accounts. It should be noted that the multiobjective programming approach is better suited to take into account the links between branches of national accounts in the quarterly disaggregation process and to reduce the volatility of the quarterly GDP obtained. The proposed MOPTD model can be applied to the national accounts of any country using the System of National Accounts (SNA) [10] which describes a uniform methodology for the compilation of annual national accounts.

Acknowledgments

All simulations are performed with Octave software installed on HP Intel Core i7 (vPro) notebook PCs, 16 Gb RAM, under Windows system.

We would like to mention that any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of nor the Institute of Mathematics and Physics Sciences (IMSP) nor the National Institute of Statistics and Economic Analysis.

References

[1] Abdelwahed, T., Leila, H. : A Benchmarking Approach for Temporal Disaggregation of Economicc Time Series by Related Series. Working paper and studies, ISSN 1725-4825, ISBN 92-79-01306-8, Luxembourg: Office for Official Publications of the European Communities, 5-10, 29-41 (2005). http://europa.eu.int.

[2] Ana, A., Enrique M. Q. : Software to perform temporal disaggregation of economic time series. Working paper and studies, ISSN 1725-4825, ISBN 92-894-8626-0, Luxembourg : Office for Official Publications of the European Communities , 3-40 (2005). http://europa.eu.int.

[3] Ankur, S., Pekka, M., Kalyanmoy, D. : A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications. arXiv: 1705.06270v1 [math.OC] 17 May 2017, 4-7 (2017)

[4] Aravind, S. : A Fast Elitist Multiobjective Genetic Algorithm: NSGA-II. Mathlab source code (2012). http://www. mathworks. com/matlabcentral/fileexchange/10429-nsga-iia-multi-objective-optimization-algorithm/content

[5] Barichard, V. : Approches hybrides pour les problèmes multiobjectif. Thèse de Doctorat Num. d'ordre : 593, école doctorale d'Angers, Laboratoire d'Etude et de Recherche en Informatique d'Angers, Univesite d'Angers, 1, 7, 22, 32-38 (2003)

[6] Belaidia, I., Mohammedib K., Brachemi, B. : Mise en œuvre de l'Algorithme NSGA-II pour l'Optimisation Multi-Objectif des paramètres de coupe en Fraisage en bout. 19ème Congrès Français de Mécanique, Marseille, 24-28 août 2009, 3-5 (2009)

[7] https://www.researchgate.net/publication/260553237_Mise_en_oeuvre_de_l'Algorithme_NSGAII_pour_l'Optimisation_Multi-Objectif_des_parametres_de_coupe_en_Fraisage_en_bout

[8] Christoph, Sax., Peter, S. : Temporal Disaggregation of Time Series. Contributed research articles, The R Journal Vol. 5/2, December 2013. ISSN 2073-4859, 80-81 (2013), https://journal.r-project.org/archive/2013-2/sax-steiner.pdf
[9] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. : A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Trans on Evol. Comp, Vol. 6, No. 2, april 2002, 183 - 188 (2002), https://www.iitk.ac.in/kangal/Deb_NSGA-II.pdf
[10] Di Fonzo,T., Marco, M. : On the Extrapolation with the Denton Proportional Benchmarking Method. IMF Working Paper Series, WP/12/169, 5-6 (2012), https://www.imf.org/external/pubs/ft/wp/2012/wp12169.pdf
[11] European Communities, United Nations and al. : System of National Accounts 2008. All rights reserved, New York, 1-7, 39-66 (2009), https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf
[12] Fonds Monétaire International : Manuel des comptes nationaux trimestriels : concepts, sources statistiques et compilation. Compilation de Adriaan M. B., Robert J. D., Nils O. Fonds Monétaire International, Washington, D.C., 93-139 (2001), https://www.imf.org/~/media/Websites/IMF/imported-publications-loe-pdfs/external/pubs/ft/qna/2000/textbook/fra/text.ashx
[13] International Monetary Fund: Quarterly National Accounts Manual. 2017 Edition, Pre-Publication, november 2018, 86-126, (2018), https://www.imf.org/external/pubs/ft/qna/
[14] Fréderic, B. : Optimisation continue, Cours et problème corrigé. Sciences Sup, Maths. Appl. pour le Master/SMAI, 327 pages Dunod, Paris, 20-21 (2006)
[15] Hafid, Z. : Représentation de solution en optimisation continue, multi-objectif et applications. Thèse de Doctorat cotutelle, Ecole Mohammadia d'Ingenieurs, LERMA- Rabat-Maroc, Institut National des sciences appliquées de Rouen, HAL Id: tel-00939980,30-31, 50-51 (2013), https://tel.archives-ouvertes.fr/tel-00939980/document, 31 Jan (2014)
[16] Institut National de la Statistique-Cameroun. : Les Comptes Nationaux Trimestriels. Document de méthodologie, République du Cameroun, Décembre 2012, 25, 32-39 (2012), http://www.stat.cm/downloads/CSS/Methodologie_comptes_trimestriels_Cameroun_version_13dec2012.pdf
[17] Insee : Comptes nationaux trimestriels. La méthode en bref. Dernière mise à jour : 29 janvier 2016, 1-2 (2016), https://www.insee.fr/fr/metadonnees/source/fichier/comptes_trim_29012016.pdf
[18] Julien, C. : Analyse post-Pareto en Optimisation Vectorielle Stochastique et Déterministe : Étude Théorique et Algorithmes. Thèse de doctorat, Mathématiques, Université de la Nouvelle-Calédonie, 19 (2014), https://tel.archives-ouvertes.fr/tel-01337505/document
[19] Marco M. : Nowcasting Annual National Accounts with Quarterly Indicators: An Assessment of Widely Used Benchmarking Methods. IMF Working Paper, WP/16/71, march 2016, 6-7, 12, 14, 16-17 (2016), https://www.imf.org/external/pubs/ft/wp/2016/wp1671.pdf
[20] Robin C. P., Kalyanmoy D., Maszatul M. M., Sanaz M., Rui W. : A Review of Hybrid Evolutionary Multiple Criteria Decision Making Methods. COIN Report Number 2014005, 1-3, 6 (2014)., https://www.egr.msu.edu/~kdeb/papers/c2014005.pdf
[21] Raïmi Aboudou Essessinou, Guy Degla2 and Babacar Mbaye Ndiaye: From Two to One Index Isomorphism in Optimization Program for Quarterly Disaggregation of Annual Times Series. Journal of Advances in Mathematics and Computer Science, 34(1&2): 1-15, 2019; Article no.JAMCS.51632, http://www.journaljamcs.com/index.php/JAMCS/article/view/30199
[22] Srinivas N., Kalyanmoy D.: Multiobjective optimization using Non-dominated Sorting Genetic Algorithm. Journal of Evo. Comput. Vol. 2, No. 3, Pages 221-248, 7-12 (1994), https://web.njit.edu/~horacio/Math451H/download/SrinivasDeb_GA.pdf

Design and Application of Optimally-Tuned Variable Parameters PID Controller for Nonlinear Engineering Systems

Magdy A. S. Aboelela

Faculty of Engineering, Cairo University, Electrical Power Dept., Giza 12613, Egypt

Abstract

The goal of this article is to investigate the implementation of the Cuckoo Search Algorithm (CSA) as an optimization technique to determine the parameters of variable parameters PID (VP-PID) controller. The VP-PID has three parameters that have to be optimally evaluated. A case of three physical imbedded nonlinearities in a single area electric power system has been selected to test the suitability of the proposed technique. The integral-square error (ISE) criterion has been considered as a part of the objective function together with the percentage overshoot and settling time. Matlab/Simulink software has been used in the simulation process. The simulation results show that the proposed VP-PID controller furnishes a better performance than the conventional PID controller.

Keywords: Proportional-Integral-Derivative controllers (PID); Variable Parameters PID controller (VP-PID); Cuckoo Search Algorithm; Nonlinear physical Systems; Matlab/Simulink

Introduction

In recent years, the problem of maintaining the power and frequency of a power system free from oscillations has become rapidly crucial need because of irregular load variations and imbedded system nonlinearities [1]. The unexpected load variations result in many undesired behaviours such as the mismatch of generated power and load demand for consumption. This can be achieved by the load frequency control (LFC) methods. Nowadays, plenty research work is going on to make the systems intelligent so the systems can successfully serve the benefits of all customers [2].

Several optimization approaches have been recorded in literature that can be applied to tune the conventional PID controller. This includes but not limited to:

Particle Swarm Optimization (PSO),

Genetic Algorithm (GA),

Bacterial Foraging Optimization (BFO) [2].

Furthermore, the LFC and AGC of a single area power system been investigated by a variety of techniques such as gravitational search algorithm [3], the modified particle swarm optimization [4], the artificial neural network [5], optimal control design [6], fuzzy controllers [23], proportional-integral-observer techniques [7], as well as LQR and Legendre wavelet function [8]. Also, the AGC control of single area power system with distributed generation has been investigated in [9]. All these researches deal only with no nonlinearities in the control loops.

On the other hand, the application of VP-PID controller has been studied in different fields by many research workers. The contribution of the variable parameters class of controllers is the cope the desired characteristic of dynamic systems. Based on the desired behavior of the system output, we can assume variable parameter controller. In general, most of the systems will start their response with peak overshoot and sustained oscillations. This can be reduced by implementing a conventional PID controller [10]. The tuned parameters of this class of controllers may be reduced after sometime when the response tends to have steady state. This is the philosophy behind the use of VP-PID controller [11-12].

In this paper, a new evolutionary optimization technique has been used for tuning the variable parameters PID controller. This algorithm is inspired by lifestyle of a bird family called cuckoo. Specific breeding and egg laying of cuckoos bird is the main basis of this optimization algorithm [13-14].

THE CLASSICAL PID CONTROLLER

The PID controller is considered to be an important component in industrial control systems because of its capability of reducing the steady state error and enhancing the dynamic response and other static characteristics of systems. The PID controller is defined, mathematically, by the next equation [15-17]:

$$u(t) = K_{\mathrm{p}}\mathrm{e}(\mathrm{t}) + K_{\mathrm{i}}\int_0^{\mathrm{t}}\mathrm{e}(\tau)\mathrm{d}\tau + K_{\mathrm{d}}\frac{\mathrm{d}\,\mathrm{e}(\mathrm{t})}{\mathrm{dt}} \qquad (1)$$

Where $\mathrm{e}(\mathrm{t})$ is the system error, K_{i} is the integral constant, K_{p} is the proportional constant, K_{d} is the derivative constant and $u(t)$ is the controller output.

On the other hand, the VP-PID controller can be described by changing controller gains as follows

Case I : The proportional gain parameter (*Kp*) must be high firstly and then decrease to the normal value in order to accelerate system response. Under these conditions, the basic formula of *Kp* is:

$$K_p = C_1 + C_2 * e_{f(i)} \qquad (2)$$

Where C_1 and C_2 are considered constants achieving the following bounderies

$$C_{1_{min}} \leq C_1 \leq C_{1_{max}} \qquad (3)$$

$$C_{2_{min}} \leq C_2 \leq C_{2_{max}} \qquad (4)$$

and $e_{f(i)}$ is the final value of the system error at optimal tuning step *i*.

Case II : The derivative gain parameter (K_{d}) is increased to prevent the oscillation and overlapping. But, this increase causes the slowdown of the system response. Thus, change of K_{d} must be from the highest value to the lowest. K_{d} can be formulized as

$$K_d = C_3 + C_4 * e_{f(i)} \qquad (5)$$

Where C_3 and C_4 are considered constants achieving the following bounderies

$$C_{3_{min}} \leq C_3 \leq C_{3_{max}} \qquad (6)$$

$$C_{4_{min}} \leq C_4 \leq C_{4_{max}} \qquad (7)$$

Case III : Integral gain parameter (K_{i}) is provided to initialize the steady state error but big values of this parameter causes oscillation and higher overshoot. Thus, change of K_{i} must be from the highest value to the lowest. K_{i} can be formulized as

$$K_i = C_6 - C_5 * e_{f(i)} \qquad (8)$$

Where C_5 and C_6 are considered constants achieving the following boundaries

$$C_{5_{min}} \leq C_5 \leq C_{5_{max}} \qquad (9)$$

$$C_{6_{min}} \leq C_6 \leq C_{6_{max}} \qquad (10)$$

CUCKOO SEARCH ALGORITHM (CSA)

The cuckoo search algorithm (CSA) is inspired by the cuckoos breeding behavior [13]. Cuckoos select a random nest for laying their eggs; and a cuckoo can place only one egg at the time. Only the eggs with the highest quality (solutions) are carried over to the next generation. The available number of host nests must be fixed. The host can detect an alien egg with a probability $P_a \in [0, 1]$. In this case, the host bird can either throw the egg out or give up the nest so as to construct a fully new nest in a new location.

In the case of generating new solutions x^{t+1} for a cuckoo i, an L*é*vy flight method is applied as follows [14].

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus \text{Lévy}\ (\lambda) \qquad (11)$$

Where α is the step size ($\alpha \geq 0$) and it must be related to the scales of the studied problem. In most cases, it may be selected to be as $\alpha = O(\frac{L}{10})$ where L refers to the studied problem characteristic scale. λ is a variable such that $1 < \lambda \leq 3$.

In fact, the Lévy flight in essence presents a random walk whose random step length is drawn from an Lévy distribution. It has an infinite variance with an infinite mean.

Lévy $\backsim u=t^{-\lambda}, (1 < \lambda \leq 3)$ (12)

A number of the new solutions should be generated by Lévy around the best solution obtained so far. This will speed up the local search.

The objective function J_i at iteration *i* is given as

$$J_i = \frac{1}{ISE_i} \quad (13)$$

so that

$$ISE_i = \int_0^\infty {e_i}^2(t)\, dt + \text{Settling Time}_i + \text{P.O.}_i \quad (14)$$

Where P.O._i and Settling Time_i represent the percentage overshoot and settling time at iteration *i* of optimization procedure [15].

APPLICATION

The controllers described in this research work can be applied on a single area power system. The system consists of the governor; the turbine and the generator; and the load. The block diagrams implementing the conventional PID and the VP-PID controllers is delineated in Fig. 1. The parameters for this system as well as the characteristics of the turbine saturation (GRC), the GDB and the time delay as non-linear elements are listed in [1]. The system is subjected to a sudden load change of 0.05 p.u. The parameters of the Cuckoo Search Algorithm (CSA) are delineated in Table 1. The flow chart of the algorithm is illustrated in Fig. 2.

Table 1. The parameters of the Cuckoo Search Algorithm (CSA)

Parameter	Value
Number of individuals	10
Number of generations	100
The probability (P_a)	0.25

Results and Discussion

The frequency control single area power system with GRC, GDB, and time delay nonlinearities incorporating the classical and VP-PID controllers has been studied in this research. The Cuckoo Search Algorithm has been utilized to tune the proposed controllers. The tuning process of the two controllers has been achieved using the Matlab/Simulink software. The results of the tuning procedure using the ISE criterion are summarized in Table 2.

As stated in the literature, a proportional controller (K_p) will reduce the rise time and will decrease, but never eliminate, the steady-state error. An integral control (K_i) will eliminate the steady-state error, but it may make the transient response worse. A derivative control (K_d) will increase the stability of the system, reducing the overshoot, and improving the transient response [15-17].

As mentioned earlier, the first alternative is to implement the conventional PID controller as noted in Fig. 3. The parameters of the first controller $(K_p,\ K_d,$ and $K_i)$ are tuned using the Cuckoo Search Algorithm but those of the second controller, PID(s), are obtained a special optimization technique assigned to this block in Matlab.

The frequency deviation curve of the PID controller as shown in Fig. 3 is somewhat acceptable. There are two main notes. The first is the irregular effect at steady state. Second is the high value of ISE. The estimated optimal parameters of this controller are shown in Table 2. The tuning procedure using the CSA gives zero value for the integral gain K_i.

On the other hand, the frequency deviation given by the VP-PID controller is shown in Fig. 4. The estimated parameters variation of the VP-PID using CSA are delineated in Fig. 5. This controller will yield a smaller value for the ISE compared with the conventional PID controller as shown in Table 2. This, of course, can be achieved with little high computation time.

The frequency deviations obtained by the classical and VP-PID controllers are illustrated in Fig. 3 and 4. These curves illustrate that we cannot arrive, exactly, to a zero-state error. This is due to the imbedded system nonlinearities and the absence of integral controller. One more thing, the increase of the time delay will result in unstable system.

Conclusion

In this paper, two classes of controllers have been applied to a single area LFC. The system under study has three sources of nonlinearities, the GRC, GBD, and time delay. The first controller is the classical PID controller while the second is the VP-PID controller. The parameters of these two controllers have been optimally evaluated using the Cuckoo Search Algorithm. Results show that the second controller behaves better than the first one. Furthermore, the effect of dealing with the three embedded nonlinearities, the GRC, GBD, and time delay, have been investigated but more detailed research is recommended.

References

[1] Aboelela, M. A. S., "Implementation of Artificial Intelligence Based Optimally Tuned Controllers to a Class of Embedded Nonlinear System", *7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Thessaloniki, Greece, 7-9 May, 2018.

[2] Ahmed Husseina, S. Sadeem Saliha , and Y. Ghazi Ghasma, "Implementation of Proportional-Integral-Observer Techniques for Load Frequency Control of Power System", Procedia Computer Science 109C, pp. 754–762, 2017.

[3] B. C. Kuo and F. Golnaraghi, "Automatic Control Systems, 9th Edition, John Wiley and Sons, NY, 2010.

[4] B. Ogbonnaa, S.N. Ndubisib " Neural network based load frequency control for restructuring power industry" Nigerian Journal of Technology (NIJOTECH), Vol. 31, No. 1, pp. 40-47, 2012.

[5] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, "GSA: a gravitational search algorithm," Information Sciences, vol. 179, 2009, pp. 2232-2248.

[6] Hong Zhang and Bo Hu, "The Application of Nonlinear PID Controller in Generator Excitation System", Energy Procedia, Vol. 17, pp. 202 –207, 2012.

[7] K. Ogata, "Modern Control Engineering", 5th Edition, Pearson, 2010.

[8] K. Vijayakumar and T. Manigandan, "Nonlinear PID Controller Parameter Optimization using Enhanced Genetic Algorithm for Nonlinear Control System", CEAI, Vol.18, No.2 pp. 3-10, 2016.

[9] L. H. Hassan, Haider A.F. Mohamed, M. Moghavvem, and S.S. Yang "Load Frequency Control of Power Systems with Sugeno Fuzzy Gain Scheduling PID Controller" ICROS-SICE International Joint Conference, 2009, Fukuoka, Japan.

[10] M. M. Ismail, M. A. Mustafa Hassan "Load frequency Control Adaptation using Artificial Intelligent Technique for one and two different areas power system" International Journal of Control, Automation and Systems, Vol. 1, No. 1, Jan. 2012.

[11] O. Aydogdu and M.Korkmaz, "A Simple Approach to Design of Variable Parameter Nonlinear PID Controller", *IPCSIT vol.20, IACSIT Press, Singapore, 2011.*

[12] R. Boruah and B. Talukdar, " Modelling of Load Frequency Control Using PI Controller and Stability Determination with Linear Quadratic Regulator", *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), Volume 9, Issue 4 Ver. VI (Jul – Aug. 2014), PP 30-36*

[13] R. C. Dorf, and R. H. Bishop, "Modern Control Systems, 13th Edition, Pearson, 2017.

[14] Santosh Kumar and S. Mishra, "AGC for distributed generation", In Proceedings of the Sustainable Energy Technologies, ICSET 2008. IEEE International Conference, 24-27 Nov. 2008. Singapore.

[15] T. S. Gorripotu, S. Gopi, H. Samalla, A. V. L. Prasanna, and B. Samira, "Applications of Computational Intelligence Techniques for Automatic Generation Control Problem-A Short Review from 2010 to 2018", In book: Computational Intelligence in Pattern Recognition, A.K. Das, J. Nayak, B. Naik, S.K. Pati, D. Pelusi (Eds.), Springer, Singapore, January 2020.

[16] X. S. Yang and Deb, S., "Cuckoo search via Lévy flights". In *Nature & Biologically Inspired Computing, World Congress,* IEEE., 210-214, 2009.

[17] Y. Abdelaziz and Ali, E. S., "Cuckoo Search algorithm based load frequency controller design for nonlinear interconnected power system". " International Journal of Electrical Power & Energy Systems, 73: 632-643, 2015.

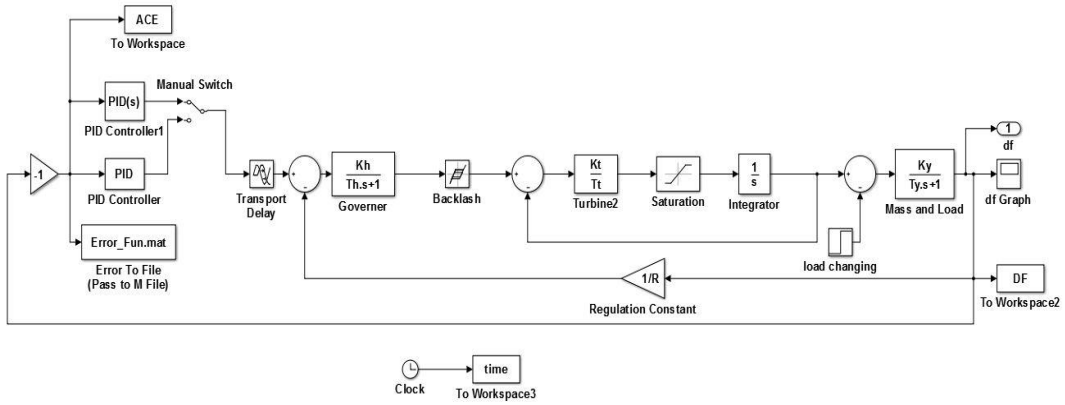


Fig. 1: Diagram of the LFC under investigation using Classical and VP-PID controllers

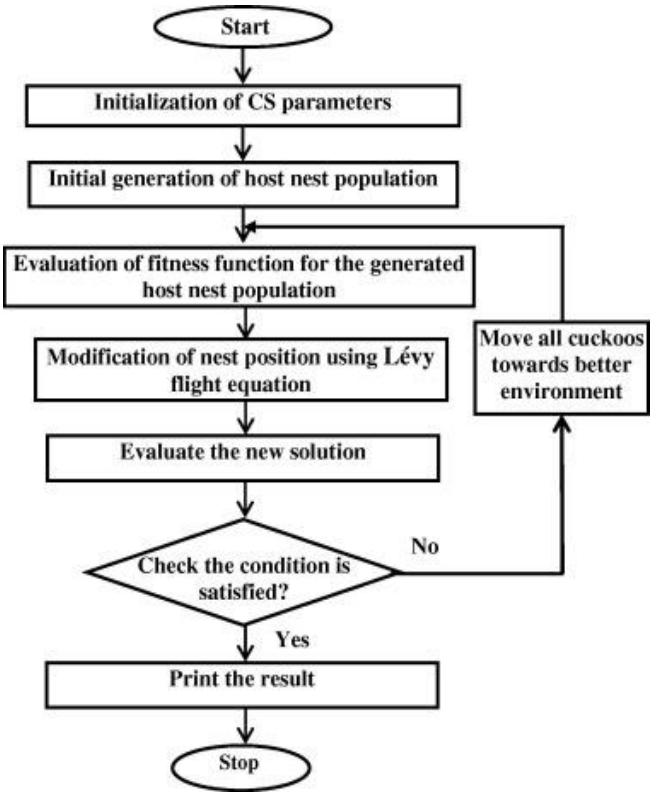


Fig. 2: The flow chart of the Cuckoo Search (CS) Algorithm [13]

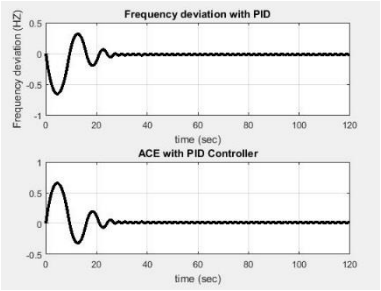


Fig. 3: Results of the PID controller (1.5 seconds time delay)

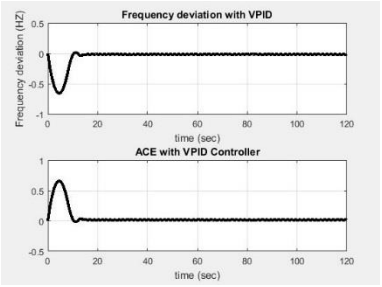


Fig. 4: Results of the VP-PID controller (1.5 seconds time delay)

(VPID=VP-PID)

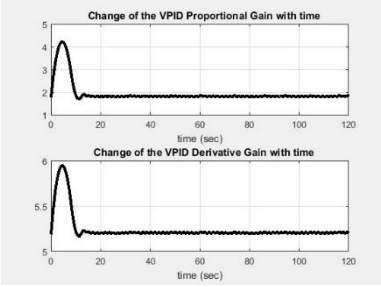


Fig. 5: Variation of VP-PID controller gains (1.5 seconds time delay)

(VPID=VP-PID)

Table 2. Results of the two proposed controllers

Controller	Parameters					
PID Controller	K_p	K_d	K_i	-	-	-
	2.4491	1.8840	0	-	-	-
ISE	12.2269					
	C_1	C_2	C_3	C_4	C_5	C_6
VP-PID controller	1.7509	3.7527	5.1860	1.1559	0	0
ISE	9.4966					

Biodynamic Synchronized Coupled Model for Crowd-Footbridge Interaction

Marcelo André Toso

Federal Institute of Santa Catarina, Xanxerê, SC, Brazil

Herbert Martins Gomes

Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil.

Abstract

Nowadays there are growing interests in vibration serviceability assessments of composite footbridges. The new design trends of composite footbridges make them slender civil structures that may be affected by the load action of walking pedestrians resulting in large deflections or even uncomfortable vibrations. Furthermore, the presence of people on the footbridges causes the addition of mass to the structural system and due to the human body's ability to absorb vibrational energy, an increase in structural damping. In this paper, the interaction between pedestrian and structure is modelled using data from pedestrian characteristics and vibration data from a measured footbridge as a comparison basis. A previously developed numerical model was used, this model called Biodynamic Synchronized Coupled Model (BSCM) consists of a fully synchronized force model in the longitudinal and lateral direction of pedestrian's movement and a biodynamic model with mass, damping and stiffness parameters. The model is coupled with the structure using the Finite Element Method at the feet's contact points. Pedestrians are treated as individuals with intrinsic kinetic and kinematic parameters following a measured correlation matrix obtained by the use of an especially designed force platform. Finally, the adequacy of the proposed model to represent the pedestrians as BSCM for the walking effects on the structure is investigated by experimentally measured accelerations on a footbridge (freely walking). The numerical results show good agreement with the experimental results.

Keywords: footbridges, human-induced vibration, crowd-footbridge interaction, walking, natural frequencies.

1. Introduction

Several cases of excessive vibrations have been studied in the past that are related to pedestrians' footbridges. Cases of unstable footbridges such as the Millennium Bridge in London and the Solferino Footbridge in Paris have attracted professional attention. These footbridges presented large vibrations on their opening days (Dallard et al. [1]). Many footbridges have natural frequencies that are coincident with the dominant frequencies of the pedestrian induced load and therefore they have the potential to undergo excessive vibrations. It is noteworthy that humans are quite sensitive to vibration in a low-frequency range of whole-body vibrations where natural frequencies of the human body limbs and systems can be observed. Therefore, there are recommendations related to human body vibration, for example, ISO 2631 [2] which define limitation curves for exposure times in some frequency ranges. This paper presents a study on the interaction between walking pedestrians and a flexible footbridge. The paper uses some previous investigation results (Toso et al. [3] and Toso and Gomes [4]). A Biodynamic Synchronized Coupled Model (BSCM) is used to analyze the pedestrian structure interaction. This model brings together interaction in 3D with several 1 DOF models to model human crowds walking and interacting with structures. In the longitudinal and transversal direction of movement a Fully Synchronized Force Model (FSFM) is used and in the vertical direction, it is combined with a biodynamic model (mass-spring-damper parameters) that is coupled to the structure. Besides, experimental data for a real footbridge considering a pedestrian crowd are presented which confirms the trends suggested by the numerical modeling.

2. Literature Review

Zivanovic et al. [5] state that the pedestrian-structure-interaction needs to consider in footbridges design, mainly in the design of slender structures that are dynamically excited by humans. Excessive vibrations may cause discomfort to pedestrians and potential deterioration of the footbridge's structural integrity. Stoyanoff and Hunter [6], report that the natural frequencies of short-span footbridge are usually not susceptible to pedestrian vibrations. However, when the

distance between the spans increases, the natural frequencies of the structure decrease, and the human occupation and interaction on them becomes a concern due to the resonance phenomena. The authors state that, during walking, vertical forces produce frequencies between 1.5 and 4.0 Hz, while lateral forces produce frequencies between 0.75 and 2.0 Hz. Considering lateral oscillations, Bodgi et al. [7], state that a crowd on a footbridge imposes a lateral dynamic excitation on the structure at a frequency close to 1.0 Hz. When the first vibration mode decreases, being in the same range as the human step rate, the resonance may occur. Consequently, there is an increase in the oscillation amplitude of the structure and the pedestrians are forced to change their natural gait. If the amplitude of oscillation is large enough, the phenomenon called pedestrian-structure synchronization occurs. Researchers such as, Wheeler [8] state that the crowd effect is not significant unless the pedestrian's step rate is close to 2.0 Hz. The author affirms that the effect of the pedestrian crowd on a footbridge with fundamental frequency away from the typical step rate (2.0 Hz) may be negligible because the vibratory response of the structure may be lower when compared to a single pedestrian walking at a frequency identical to the fundamental frequency of the analysed structure. Qin et al., [9], evaluated the dynamic response of a footbridge (analytical study) considering a single pedestrian and the effects of interaction between human-structure. The analysed footbridge was modelled as a simply supported Euler-Bernoulli beam with uniform cross-section. The pedestrian damping varies over time, considering the individual's walking speed. A controlled force was used to compensate the energy dissipated during walking. The effects of stiffness and damping of the human body were investigated. According to the authors, for a flexible structure, the pedestrian-structure dynamic interaction is greater when compared to a rigid structure. There is also potential elastic energy of the individual's legs, "tending to separate the pedestrian from the structure," being greater when the individual is near to footbridge mid-span. As result of the interaction between pedestrians and flexible structures, it is noted that the individual must impose more external energy and modify the walking pattern to maintain his steady gait and a relatively uniform dynamic behavior of the body's center of mass. Recently, Tubino [10] proposed a numerical model that accounts for pedestrian structure coupling in the vertical direction. The footbridge was modelled as a continuous unidimensional beam dynamic system, while pedestrians were schematized as moving single-degree-of-freedom systems with random dynamic properties. The paper results show possible variations of damping ratio and natural frequency in the coupled system based on the random pedestrians' parameters. Regarding pedestrian loads, these forces have been determined from investigations using force platforms, treadmill machines, and even prototype footbridges, in which the applied force is the amount produced by a single walking pedestrian. The combined force applied by individuals is considered for groups of pedestrians or crowds. Thus, the design load is a force model. To analyse the analytic human-induced loads, most of guidelines, for instance, SETRA [11] and ISO 10137 [12] often consider three to five harmonics of the frequency spectrum of ground reaction forces (GRF). During walking, a pedestrian produces dynamic forces with components in three directions: vertical, lateral and longitudinal to the footbridge. The vertical component is generated by the impact of supporting the body weight on each leg alternately. In the lateral direction, the forces are generated by the periodic balance of the body when changing legs. Finally, in the longitudinal direction, the force is the result of friction between the foot and the floor, as well as the acceleration and deceleration of the body in this direction. Using force platforms, some researchers (Harper et al., [13]; Galbraith and Barton, [14]; Blanchard et al., [15]; Kerr, [16]) conclude that the vertical component of the resultant force of an individual has two peaks and a valley as shown in Figure 1. The other force components are also present in the figure.

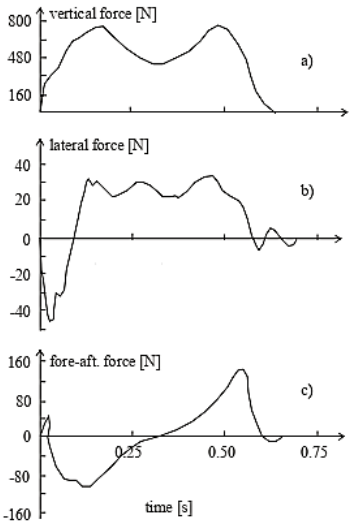


Figure 1: Typical forces during walking: (a) vertical, (b) lateral and (c) longitudinal direction. (Zivanovic et al. [5]).

3. Methodology

This paper proposes to assess the pedestrian-structure interaction using data from an experimentally measured footbridge as a basis for comparisons. In the longitudinal and transversal direction of movement a Fully Synchronized Force Model (FSFM) is used and in the vertical direction, it is combined with a biodynamic model (mass-spring-damper parameters) that is coupled to the structure (herein called, BSCM, Biodynamic Synchronized Coupled Model). The biodynamic model considers the synchronization of the three force components applied in space (positions where they should be applied) and in time (peak and valleys of the three force components occurring synchronously at the proper time) with a spring-mass-damper model, coupled continuously with the structure's FEM. Following the use of the BSCM, the mid-span RMS (root-mean-square) acceleration of a footbridge structure is evaluated to check its serviceability. For the applied BSCM, kinetic (forces) and kinematic (speeds, pacing rate, step length, and step width) parameters are used. Then the BSCM results are compared with the FSFM and these results are compared with experimental data.

3.1. Force Models

Design Codes use simplified force models to represent the force magnitude from successive footfalls. These models assumes forces acting in a straight line along the direction of walking at a constant speed (SETRA Guideline [11] and ISO 10137 [12]). This is a very common assumption for the analysis and design of footbridges. It is obvious the major disadvantages in using this simplified model as it lacks for dynamic interaction (only time-varying forces are used) and the spatiality and synchronization of application of the three forces components. SETRA Guideline [11] and ISO 10137 [12]) use force models to vertical, lateral and longitudinal direction of the walking, based in Fourier series. Overall, these models produces acceleration unsafe values, since they are based on excessively simplified load models. Then, it is necessary to incorporate other parameters to simulate the pedestrian-structure interaction (for instance, considering biodynamic models). Furthermore, in a recent publication Toso et al. [3] proposed a force model called Fully Synchronized Force Model (FSFM). In that model, the changes in velocities during the walking (single and double stance phase) were considered and they were synchronized in time and space. The authors pointed that the pedestrian speed in the double stance phase is greater than at single stance phase. This speed is also greater than the average speed of the pedestrian. Another important characteristic is that human walking does not occur in a straight line in the direction of walking (as proposed by Guidelines). There are parameters like step length and step width that influence the application of the resulting force. Using a specifically designed force platform, these kinematic parameters were measured, and average values for a test campaign with 54 subjects were presented by Toso et al. [3]. In that proposed force model, peak and valley values from each force component should be placed accordingly in the right position of the contact surface and the model's reference time adjusted to the correct phase. Thus, there is a spatial and temporal synchronization of the three ground reaction force components. More details about the Fully Synchronized Force Model (FSFM) can be found in Toso et al. [3].

3.2. Biodynamic Synchronized Coupled Model

In the literature, there is a consensus that shifts in the structural natural frequencies are not observed when modelling the pedestrians using force-only models. This will only be present, as indicated by experimental measurements of a crowd of pedestrians crossing footbridges, if mass, dissipative effects, and synchronism of applied loads are taken into account. Another observed feature is the structural damping increase when considering pedestrians like biodynamic models. This is attributed to the human body's ability to absorb energy, a feature represented only when introducing the biodynamic models. In this paper, the FSFM (Fully Synchronized Force model) is merged to a biodynamic model resulting in the so-called BSCM (Biodynamic Synchronized Coupled Model) composed of mass, damping, and stiffness, with a single degree of freedom (SDOF) that represents the action of a walking pedestrian in the vertical direction. Afterward, the model's degree of freedom is continuously coupled to the structure in the places where the feet have contact with the floor and replicated to represent a pedestrian crowd. This replication takes into account intra and inter-subject variability (mainly due to subject's weight, height, step length, step width and pacing rate). This makes possible to analyze the structural behaviour according to distinct crowd densities in a systematic way. Toso and Gomes [4] proposed this model, where a biodynamic model with synchronization of the three force components applied in space (positions where they should be applied) and in time (peak and valleys of the three force components occurring at the proper time) was presented. This biodynamic model is coupled continuously with the structure's FE Model. The biodynamic model has been conceived in a way that the actuator force is the main source that drives the human will for walking, generating oscillations in the vertical direction. The actuator reasoning is based on experimentally obtained vertical force (using a force platform) and acceleration data (using an accelerometer attached to the pedestrian's body waist). This allows the fit of the biodynamic model (with actuator) in order to match measured vertical ground reaction force and acceleration along time, in a rigid platform. It is important to note that, for flexible structures, the interaction force will change as the flexible structure add a relative displacement to the biodynamic model. A biodynamic model for flexible structure, including mass, damping, stiffness and actuator that is coupled to the footbridge in the vertical direction is considered, according to Figure 2.

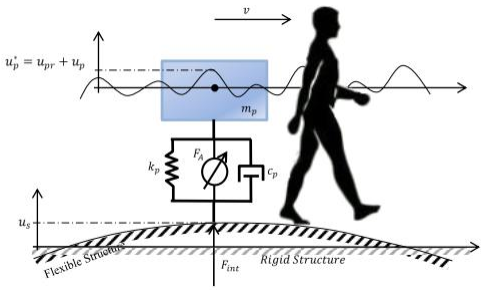


Figure 2: The biodynamic single degree of freedom system composed of mass, damping, stiffness and actuator and flexible structure situation.

In this figure, u_p^* is the pedestrian vertical displacement around the center of mass in the initial equilibrium rest configuration, considering a rigid and a flexible structure; c_p and k_p are the damping and stiffness of pedestrian; F_A is the vertical actuator force; v is the pedestrian speed; m_p is the pedestrian modal mass; u_s is the structural displacement and F_{int} is the interaction force.

The model proposed allows assessing interactions with both rigid and flexible structures. Toso and Gomes [4] presented a complete description of the model; here it is showed some equations to evaluate the pedestrian structure interaction. Equation (1) represents the pedestrian's interaction force $\{\boldsymbol{F_{int}}\}$ when walking on a flexible structure, according to presented in Figure 2.

$$\{\boldsymbol{F_{int}}\} = \{\boldsymbol{H}(e,\xi_0)\}^T[-c_p(\dot{u}_p - \{\boldsymbol{H}(e,\xi_0)\}\{\dot{\boldsymbol{u}}_s\}) - k_p(u_p - \{\boldsymbol{H}(e,\xi_0)\}\{\boldsymbol{u}_s\}) + F_R] \quad (1)$$

Here: $\{\boldsymbol{H}(e,\xi_0)\} = [0\ 1\ 0]_{1\times3}[\boldsymbol{N}^e(\xi_0)]_{3\times6}[\boldsymbol{Q}]_{6\times n}$. Where $[\boldsymbol{N}^e(\xi_0)]_{3\times6}$ is the shape function, assuming a plane structural beam Euler-Bernoulli finite element; $[\boldsymbol{Q}]_{6\times n}$ is used in order to evaluate only the displacements in the vertical direction at the contact point; u_p and $\dot{u}_p$ are the displacements and velocities related to the interaction between pedestrian

and flexible structure; $\{\boldsymbol{u}_s\}$ and $\{\dot{\boldsymbol{u}}_s\}$ are the displacements and velocities related to structure; F_R is the pedestrian vertical reaction force, measured in a rigid structure.

3.2.1. Linking the biodynamic model to the structural equations of motion

Equations (2), (3) and (4) represent the matrices of the coupled system (pedestrian and structure), that takes into account mass, damping and stiffness respectively.

$$[\mathbb{M}]_{\boldsymbol{n+1}\times\boldsymbol{n+1}} = \begin{bmatrix} [\mathbf{M}]_{\boldsymbol{n}\times\boldsymbol{n}} & \{\mathbf{0}\}_{\boldsymbol{n}\times\mathbf{1}} \\ \{\mathbf{0}\}^{\boldsymbol{T}}_{\mathbf{1}\times\boldsymbol{n}} & m_p \end{bmatrix} \tag{2}$$

$$[\mathbb{C}(t)]_{\boldsymbol{n+1}\times\boldsymbol{n+1}} = \begin{bmatrix} [[\boldsymbol{C}]_{\boldsymbol{n}\times\boldsymbol{n}} + \boldsymbol{C}^*(i)_{\boldsymbol{n}\times\boldsymbol{n}}] & -\{\boldsymbol{H}(e,\xi_0)\}^T c_p \\ -\{\boldsymbol{H}(e,\xi_0)\}c_p & c_p \end{bmatrix} \tag{3}$$

$$[\mathbb{K}(t)]_{\boldsymbol{n+1}\times\boldsymbol{n+1}} = \begin{bmatrix} [[\boldsymbol{K}]_{\boldsymbol{n}\times\boldsymbol{n}} + \boldsymbol{K}^*(i)_{\boldsymbol{n}\times\boldsymbol{n}}] & -\{\boldsymbol{H}(e,\xi_0)\}^T k_p \\ -\{\boldsymbol{H}(e,\xi_0)\}k_p & k_p \end{bmatrix} \tag{4}$$

One way to solve the time domain Equation of motion is by using direct numerical integration. In this paper, the Newmark integration scheme (Bathe [17]) is used in order to evaluate the structural response considering all the structural vibration modes.

Toso and Gomes [4] showed that considering a three-dimensional truss element, the human-structure interaction model can be evaluated using the Equation (5):

$$\begin{bmatrix} [\mathbf{M}] & [\mathbf{0}] \\ [\mathbf{0}] & diag\{\boldsymbol{m}_p\} \end{bmatrix} \begin{Bmatrix} \{\ddot{\boldsymbol{u}}_s\} \\ \{\ddot{\boldsymbol{u}}_p\} \end{Bmatrix} + \begin{bmatrix} \left[[\boldsymbol{C}] + [\boldsymbol{H}]^T diag\{\boldsymbol{c}_p\}[\boldsymbol{H}]\right] & -[\boldsymbol{H}]^T diag\{\boldsymbol{c}_p\} \\ -diag\{\boldsymbol{c}_p\}[\boldsymbol{H}] & diag\{\boldsymbol{c}_p\} \end{bmatrix} \begin{Bmatrix} \{\dot{\boldsymbol{u}}_s\} \\ \{\dot{\boldsymbol{u}}_p\} \end{Bmatrix} +$$

$$\begin{bmatrix} \left[[\boldsymbol{K}] + [\boldsymbol{H}]^T diag\{\boldsymbol{k}_p\}[\boldsymbol{H}]\right] & -[\boldsymbol{H}]^T diag\{\boldsymbol{k}_p\} \\ -diag\{\boldsymbol{k}_p\}[\boldsymbol{H}] & diag\{\boldsymbol{k}_p\} \end{bmatrix} \begin{Bmatrix} \{\boldsymbol{u}_s\} \\ \{\boldsymbol{u}_p\} \end{Bmatrix} = \begin{Bmatrix} \{\boldsymbol{F}(t)\} \\ \{0\} \end{Bmatrix} - \begin{Bmatrix} \{\boldsymbol{F_R}(t)\} \\ \{0\} \end{Bmatrix} \tag{5}$$

where $diag\{\boldsymbol{m}_p\}$, $diag\{\boldsymbol{c}_p\}$ and $diag\{\boldsymbol{k}_p\}$ will contain the mass, damping and stiffness parameters of each of the pedestrians.

4. Numerical and experimental results

The analyzed structure is located in the city of Brasília, Brazil. It is a composite footbridge (34.08 m in length, 2.4 m in width and 2.25 m in height). The footbridge roof is built with a curved reinforced concrete shell. The deck floor consists of reinforced concrete planks that are simply supported on the truss members that link the left and right side of the footbridge. The handrails consist of hollow tubular steel bars. Additional information, design details, materials properties, etc. can be found in Brasiliano et al. [18]. Figure 3 shows the analyzed footbridge.

Figure 3: Analyzed footbridge.

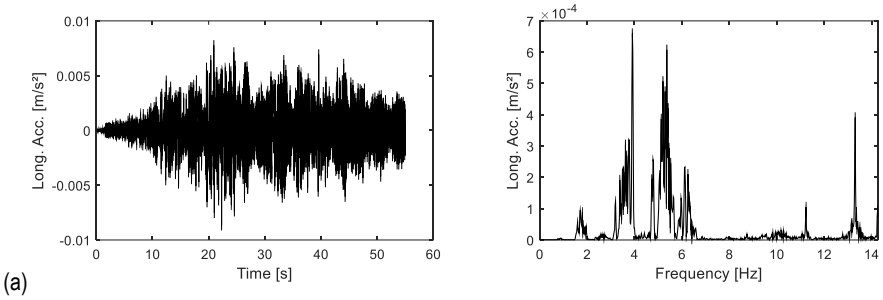
4.1. Numerical model results

In this topic, a dynamic analysis is performed to investigate the footbridge dynamic behaviour under a pedestrian crowd (0.25 pedestrians/m^2). This analysis consider: (a) a Biodynamic Synchronized Coupled Model (BSCM) that considers an actuator in the vertical direction of each pedestrian, and a fully synchronized force model as mechanisms for the willingness of walk, in a number that represents some crowd; (b) a Fully Synchronized Force model (FSFM) to vertical, longitudinal

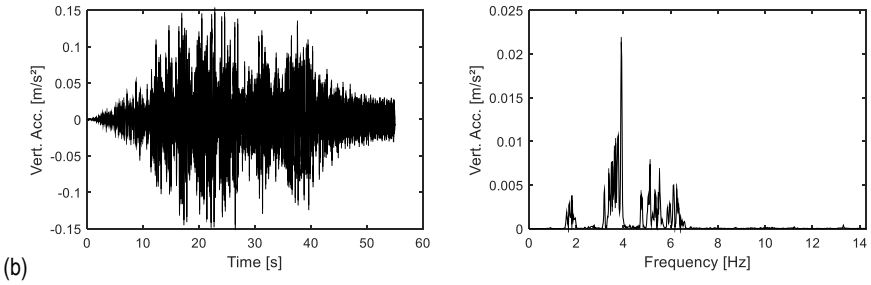
and lateral directions. The two models use the dynamic load factors proposed to SETRA Guideline [11] for force estimation. Both models assume the entry of the pedestrians on the footbridge is randomly generated, considering the effective width of the structure and keeping the specified crowd densities. Pedestrians and their characteristics are represented using random variables, following a Gaussian distribution based on average parameters, coefficients of variation and a correlation matrix of experimental data. Monte Carlo method is used to obtain the pedestrian's kinematic parameters, which results in different pacing frequency, step length, step width etc. for each individual (Toso et al. [3]). Regarding to biodynamic parameters, Toso et al. [19] proposed a single degree of freedom (SDOF) biodynamic model to obtain the following parameters: mass (m), damping (c) and stiffness (k). In this paper, those equations were used to corresponding to the data of the analysed crowd. These biodynamic parameters are used in Equation 5.

Individuals have different characteristics, for instance height, mass, pacing rate etc. According to Toso et al. [3], the crowd characteristics are modelled based on mean value and standard deviation of the group of pedestrians being modelled. Each random variable follows a correlated Gaussian distribution based on experimentally obtained data. So, for the number of pedestrians that are to be modelled, samples are generated based on a correlated Gaussian distribution for these parameters. Besides, it is also reasonable to assume that for a specific individual, there are some trends in the kinetic and kinematic parameters such as, body mass, step length, step width, pacing rate, etc., that justifies the use of a correlation matrix (Toso et al. [3]).

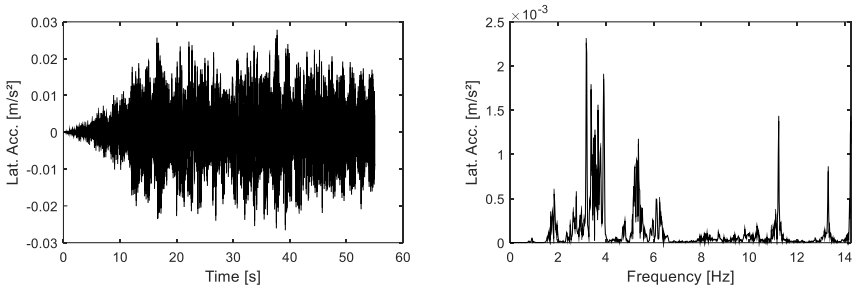
Figure 4 presents the results of the mid-span accelerations in the longitudinal, vertical and lateral directions. This numerical simulation uses the Biodynamic Synchronized Coupled Model (BSCM). This model is couple to the structure in the vertical degree of freedom. In other directions remaining the fully synchronized force model. This analysis assumes that the three force components are completely synchronized (time and space), and uses the kinetic and kinematic parameters for the pedestrians. A pedestrian crowd of 0.25 pedestrians/m^2 is use in this simulation.



(a)



(b)



(c)

Figure 4. Acceleration response for BSCM and crowd density 0.25 ped/m^2:

a) longitudinal, b) vertical and c) lateral directions.

Figure 5 shows the results of mid-span acceleration for longitudinal, vertical and lateral direction, considering just the Fully Synchronized Force Model (FSFM) in three directions. As in the previous example, there is no synchronism between pedestrians, that have their own pacing rates, and crosses the footbridge only once.

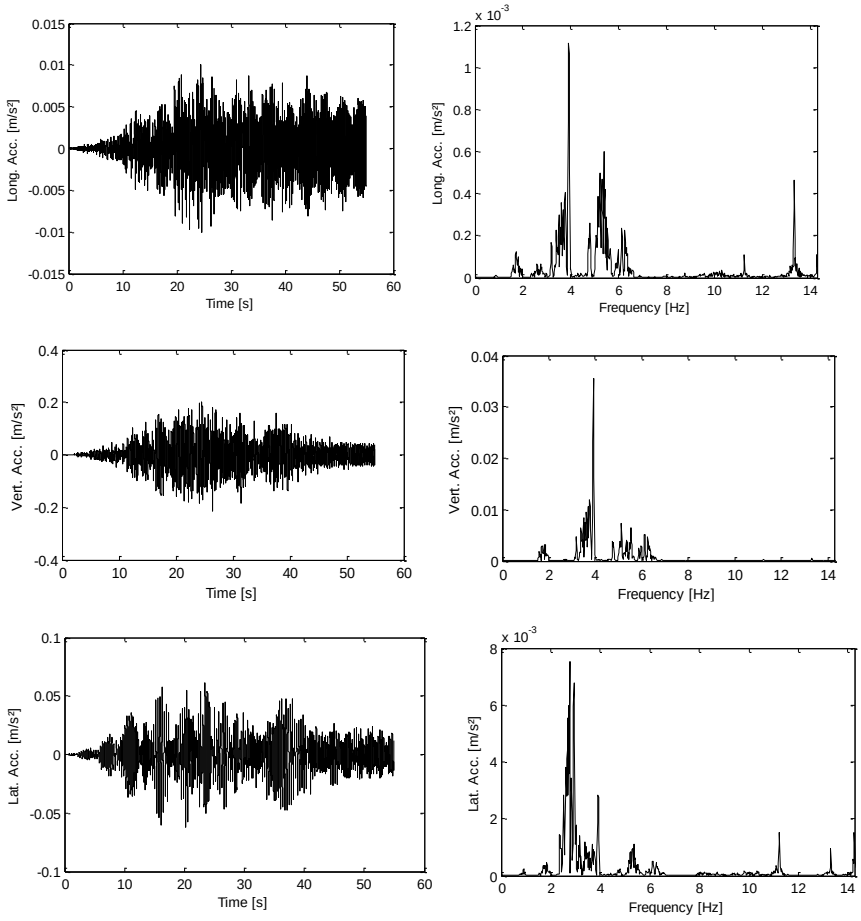


Figure 5. Acceleration response for FSFM and crowd density 0.25 ped/m^2:

a) longitudinal; b)vertical; c) lateral direction.

The numerical results presented in Figures 4 and 5 shows natural frequencies to around 2.0 Hz. These frequencies are associated with pedestrians' pacing rates. A frequency about 3.91 Hz corresponds to the natural frequency of the first vertical bending mode of the structure that was excited by the pedestrian crowd. Furthermore, the fundamental frequency of the structure (frequency about 3.0 Hz which corresponds to the lateral bending mode) was also excited by the pedestrian crowd, using both models. Considering the lateral and longitudinal directions the previous results shows frequencies around 11.0 and 13.0 Hz. These frequencies are excited due to the presence of the pedestrian thrust and deceleration phases in the human step and the zig-zag pattern in the force application. These frequencies are not excited if Design Codes (simplified force models) are used, because Design Codes assumes that the forces acting in a straight line along the direction of walking at a constant speed. Furthermore, the torsional mode (6.20 Hz) was also excited using both models, due to the spatiality of application of the pedestrian load that produces torsion in the footbridge. These frequencies, again, will not appear if simplified force models be used, since this model assumes the application of the load along a straight line.

Table 1 shows the RMS acceleration (footbridge's mid-span response) obtained with the two models (FSFM and BSCM). The acceleration was obtained in longitudinal, vertical and lateral directions. These results show BSCM presented lower acceleration values than FSFM in all directions.

Table 1. RMS acceleration response for a density of 0.25 pedestrians/m^2.

	RMS mid-span acceleration (m/s²)		
Model	Longitudinal	Vertical	Lateral
FSFM	0.0026	0.0526	0.0158
BSCM	0.0019	0.0394	0.0070

* FSFM: Fully Synchronized Force model; BSCM: Biodynamic Synchronized Coupled Model.

Regarding the influence of the BSCM, considering a crowd density of 0.25 ped./m^2, it is noted that both models (Biodynamic Synchronized Coupled Model and Fully Synchronized Force model) presented a fundamental frequency about 3.91 Hz, i.e., there was no modification in the structural fundamental frequency. Additional simulations (Table 2) show that for the crowd density of 0.50 and 0.75 ped./m^2 the use of FSFM still results in structural fundamental frequencies at 3.91 Hz, as expected. However, the BSCM presents a decrease in the structural fundamental frequency, according to Table 2, proving the previous trend. It is concluded that this is due to increase in overall mass as the crowd density increases. Maybe such decrease in natural frequency was not observed in a crowd density of 0.25 ped./m^2 due to a low number of pedestrians.

Table 2. Structural fundamental frequency (Hz).

	Crowd Density (ped./m^2)		
Model	0.25	0.50	0.75
FSFM	3.91	3.91	3.91
BSCM	3.91	3.61	3.47

* FSFM: Fully Synchronized Force model; BSCM: Biodynamic Synchronized Coupled Model.

4.1. Experimental results

The results presented in Figure 6 correspond to the investigated pedestrian density (0.25 ped./m^2). In a previous publication, Brito et al. [20] measured the vertical acceleration of this footbridge at the mid-span. In this case and crowd density, the subjects crossed the structure freely, maintaining their own pacing rates.

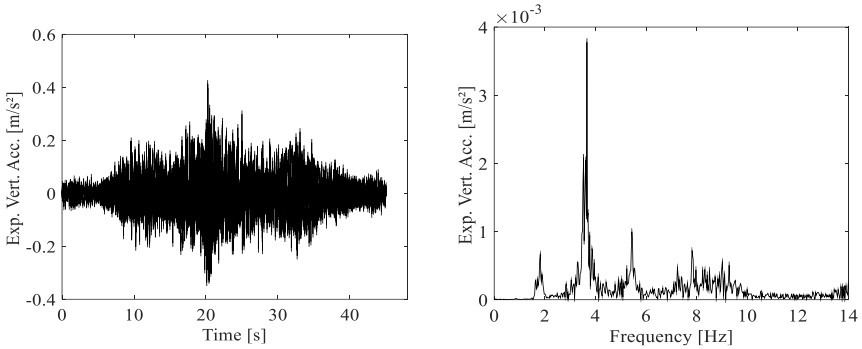


Figure 6. Mid-span vertical acceleration: experimental measurement (0.25 ped./m^2).

Both models BSCM and FSFM presents a similar shape for the vertical acceleration. The results for BSCM from Figure 4(b) shows a peak acceleration of 0.15 m/s². Using the FSFM (Figure 5b) the acceleration peaks are 0.20 m/s² and the experimental values are about 0.40 m/s^2. One can note that these numerical results are underestimating actual peak vibration. In terms of vertical RMS acceleration, the values became closer to the experimental ones (0.068 m/s^2), with FSFM resulting in 0.0523 m/s^2 and BSCM 0.0394 m/s^2. It should be emphasized that this comparison was harmed because most of the BSCM parameters were not available and the experimental campaign was performed previous to the development of the BSCM.

Figure 7 shows the spectrogram for the corresponding mid-span vertical acceleration: experimental measurement. This shows how the natural frequency varies along time according to the mass variation due to crowd flow. One can notice that the main lower frequency only happens during the crowd entrance in the footbridge (approximately from 6 s to 35 s). The second main frequency happens along all the experiment.

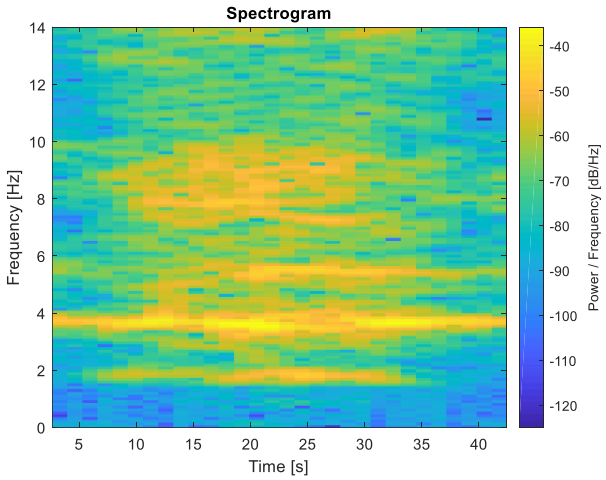


Figure 7. Spectrogram for the mid-span vertical acceleration: experimental measurement (0.25 ped./m^2).

5. Conclusions

In this paper, it is used a biodynamic model with synchronization of the three force components in space and in time with a stiffness-mass-damper model coupled with the structure's FEM to assess the pedestrian structure interaction. Both models used here (Biodynamic Synchronized Coupled Model and Fully Synchronized Force Model underestimated the peak vertical acceleration values when compared to the experimental ones and this suggest other sources of uncertainty may affect the final behavior. The effect of the BSCM in other simulation for higher crowd densities (0.50 and 0.75 ped./m^2) could be noticed since the simulated structure presented modifications in the dynamic behaviour as a decrease in first

natural frequency and increase in the overall damping as the crowd density increases. This suggests the human body effectively dissipates part of this vibration energy. The results of the dynamic analysis using FSFM rendered larger RMS accelerations, mainly in the lateral and longitudinal direction if compared with the BSCM. Additional frequencies present in both spectral acceleration in the lateral and longitudinal directions can be attributed to the use of the synchronized force model, used in both models. The observed differences in spectral accelerations in the vertical direction are attributed to the biodynamic effect that is present in BSCM and not in the FSFM.

Regarding to natural frequencies of footbridge, the vertical fundamental frequency of 3.91 Hz was identified experimentally and numerically. This is a relatively high frequency when considering structures with similar spans according to CEB [21]. Possibly, this characteristic is a consequence of the balance of stiffness and mass obtained when using a composite footbridge (steel truss and concrete floor). Such characteristics are desirable, avoiding that the structural fundamental frequency be a value close to the frequencies of excitation of the first harmonic of the pedestrian walking. Considering the lateral fundamental frequency, experimentally, the frequency of 3.12 Hz was obtained. This value is far from the frequency range considered to be critical, which are values close to 1.0 Hz. As for the vertical structural acceleration, the footbridge has vibration amplitudes below the values set by standards for comfort limits, which are of the order of 0.5 to 0.7 m / s² (OHBDC [22]; BS-5400 [23]; Eurocode 5 [24] and SETRA Guideline [11]). Numerically, it was observed that the excitation range of the second harmonic, excited by some pedestrians, are frequency values very close to the vertical structural fundamental frequency. However, the results of numerical simulations with different pedestrian's densities, walking on the footbridge indicate that it is not subject to excessive vibrations.

References

[1] Dallard P, Fitzpatrick AJ, Flint A, Le Bourva S, Low A, Ridsdill-Smith RM, et al. The London millennium footbridge. J Struct Eng 2001; 79 (22), 17-33.

[2] ISO 2631-2. International Organization for Standardization. Evaluation of human exposure to whole-body vibration, Part 2: human exposure to continuous and shockinduced vibrations in buildings (1 to 80 Hz), 1989.

[3] Toso MA, Gomes HM, de Brito JLV. Crowd-structure interaction: investigating the spatiality and synchronization of a pedestrian force model. Int J Constr Steel Res 2017;133: 510-21.

[4] Toso MA, Gomes HM. A coupled biodynamic model for crowd-footbridge interaction. Engineering Structures 2018; 177: 47-60.

[5] Zivanovic S, Pavic A, Reynolds P. Vibration serviceability of footbridges under human-induced excitation: a literature review. J Sound Vib 2005; 279: 1-74.

[6] Stoyanoff S, Hunter M. Footbridges: Pedestrian induced vibrations. Rowan Williams Davies and Irwin Inc, Ontario, Canada, 2003.

[7] Bodgi J, Erlicher S, Argoul P. Lateral vibration of footbridges under crowd-loading: Continuous crowd modeling approach. Key Engineering Materials 2009; 347, 685-690.

[8] Wheeler JE. Prediction and Control of Pedestrian Induced Vibration in Footbridges. J Structural Division 1982; 108, n° ST9, 2045-2065.

[9] Qin JW, Law SS, Yang QS, Yang N. Pedestrian-bridge dynamic interaction, including human participation. J. Sound Vib 2013; 332, 1107-1124.

[10] Tubino F. Probabilistic assessment of the dynamic interaction between multiple pedestrians and vertical vibrations of footbridges. J Sound Vib 2018; 417:80–96.

[11] SETRA-Service d'Études techniques des routes et autoroutes. Footbridges. Assessment of vibrational behaviour of footbridges under pedestrian loading, 2006.

[12] ISO 10137, International Organization for Standardization. Bases for Design of Structures-Serviceability of Buildings and Walkways against Vibrations, 2012.

[13] Harper FC. The mechanics of walking. Research Applied in Industry 1962; 15 (1), 23-28.

[14] Galbraith FW, Barton MV. Ground loading from footsteps. Journal of the Acoustic Society of America 1970; 48,1288-1292.

[15] Blanchard J, Davies BL, Smith JW. Design criteria and analysis for dynamic loading of footbridges. In: Proceedings of the DOE and DOT TRRL Symposium on Dynamic Behaviour of Bridges, Crowthorne, England, 1977, 90-106.

[16] Kerr SC. Human Induced Loading on Staircases. PhD Thesis (Mechanical Engineering Department) University College London, London, 1998.

[17] Bathe KJ. Finite element procedures. 2nd ed. New Jersey: Prentice-Hall; 1996.

[18] Brasiliano A, Doz G, Brito JLV, Pimentel RL. Role of non-metallic components on the dynamic behaviour of composite footbridges. In: Proceedings of the 3rd international conference footbridge; 2008.

[19] Toso MA, Gomes HM, da Silva FT, Pimentel RL. Experimentally fitted biodynamic models for pedestrian-structure interaction in walking situations. Int J Mech Syst Signal Process 2016; 72-73: 590-606.

[20] Brito JLV, Doz GN, Ávila SM, Pimentel RL, Brasiliano A, Roitman N, et al. Evaluation of the dynamic behaviour of a composite footbridge located in the city of Brasília-Brazil. In: XXXIV Jornadas Sul Americanas de Engenharia Estrutural; 2010 [in Portuguese].

[21] CEB, Comite Euro-International Du Beton. Bulletin D'Information N° 209. Vibrations Problems in Structures. Practical Guidelines, Lausanne, Switzerland, August, 1991.

[22] OHBDC, Ontario Highway Bridge Design Code, Highway Engineering Division, Ministry of Transportation and Communication, Ontario, Canada, 1991.

[23] BS 5400, British Standards Association. Steel, Concrete and Composite Bridges - Part 2: Specification for Loads; Appendix C: Vibration Serviceability Requirements for Foot and Cycle Track Bridges, London, 1978.

[24] EUROCODE 5-2. Design of Timber Structures – Part 2: Bridges, ENV 1995-2, European Committee for Standardization, Brussels, Belgium, 1997.

Continuous Quality Improvement System (CQIS) and the Hospital Information System (HIS)

Tomi Thomo, Ph.D. Cand.

John Sarivougioukas, Ph. D.

Abstract

The adoption of Hospital Information Systems (HIS) aims at the achievement of the desired level of quality. Quality is a major concern for HIS attempting to continuously control the large number of defining parameters. However, quality is characterized by its tendency to be redefined according to the holding nosocomial conditions, the way hospitalization services offered to the patients, the adopted medical technology, the availability of resources, and the applying workload. Thus, there is a continuously holding need for reconsidering the way of offering the hospitalization services. The Continuous Quality Improvements System (CQIS) is dedicated to monitor the operation of HIS and provide the exact required information to make the proper decision in order to achieve the desired, each time, level of Quality offered by the HIS to the end users and the patients. Hence, an information technology system must be provided to oversee the operational results of HIS and support the decision making process of the hospital's administration reconsidering the HIS functional terms.

Keywords: Continuous Quality Improvement System (CQIS) and the Hospital Information System (HIS)

1. Introduction

The Continuous Quality Improvement System (CQIS) includes the set of functionalities allowing to ameliorate the quality level of the provided hospitalization services. The improvement is achieved performing a cyclic processing operation following the well-known and clearly define Quality process (Plan-Do-Act-Check) [Ref-1]. In other words, CQIS provides the sufficient and necessary functionalities to apply the four stages cyclic quality process. First, define the desired level of quality, second, evaluate the obtained operational level based on actual measured performance, third, make decisions supported by evidences to achieve the desired level, and eventually, forth, set the plan to obtain the desired level of quality. The operation of CQIS must be facilitated by a Hospital Information System (HIS) since it documents the activities of all important and vital processes carried within the nosocomial organization to succeed in the provision of the necessary and required medical services. CQIS sensors must be fused within the HIS organizational structure to achieve the desired level of Quality of Services (QoS) indices.

The operational area of hospitals is characterized by dynamism due the continuously changing functional terms of operation such as the availability of hospital's resources, workload, and applying technology. The continuous improvement of the offered hospital services has always been a requirement by the hospitals administration. Hospitals resist to adopt methods, standards, protocols, and best practices without a sound and complete prepared plan. The desired, each time, results can be obtained with the applying organizational infrastructure, the availability of the necessary resources, the adequacy of the plan, and the obtained outcomes. CQIS needs to be coupled with HIS in order to support the hospitals administration with dependable means for decisions making on well-defined level of Quality of Service (QoS) and taking advantage of the currently available artificial intelligence.

The rest of the paper continues with the examination of the contemporary literature, it examines the hospital environment from the Quality point of view, it analyzes the functional and operational requirements of CQIS, it considers the impact on HIS, it proceeds with the implementation plans, and it concludes with considerations about future research.

2. Characteristics of the proposed CQIS

The characteristics of the performed work in this paper are provided below. There are presented the five more outstanding characteristics.

The nature of HIS is about a hospital's system consisted of skillful personnel, software and hardware devices, IT networks, data, processes, procedures, and specific work tasks [1]. The purpose of HIS is the storage, retrieval, transformation, and

dissemination of information within the hospital [1]. HIS is based on IT communication over virtually many and complex channels storing information, using various electronic devices, following predefined processes, procedures, and carrying specific tasks. Thus, HIS forms a seeming static framework to perform the tasks assigned by the administration of the hospital. However, the dynamic nature of the operation of the hospital requires frequent adjustments to meet the dynamic business - medical requirements. Hence, the first point of our work is about the need of a software system or a mechanism, performed by CQIS, to rectify the HIS operation in such a way that the desired level of quality is proved to be achieved by the evidential stored data.

The HIS rectified operation behavior due to CQIS presence is based on the proper processing of the stored data and the associated made decision. Nevertheless, the processing operation on the HIS data must be performed on a variety of different and incompatible types of data. For instance, the patients admission data must be compared and correlated to the analytical laboratory results to draw specific conclusions and make decisions accordingly. Processing requires the employment of the means to process dissimilar data from various sources within HIS to perform inferences, deductions, and inductions. In such cases, the extraction-transformation-loading (ELT) approach is employed. The ETL approach provides the necessary means for the adaptability of workflows exchanging data and integrating processes among different HIS sub-modules [2]. Hence, the second point of our work is about the integration of various HIS modules based on the data provided by the carrying workflows and the achievement of inter-process communication belonging to different modules of HIS.

The decisions required to be made for the rectification of the HIS operation must be well documented and based on objective evidences. Such evidences must be based on the stored data. However, the stored data schema, received from internal or external sources, must be realized by the hospital's administration supporting the decision making system. In other words, HIS must be equipped with an additional database that provides a single point of managerial information for reporting and analysis purposes supporting the decision-making process [3]. Hence, the third point of our work is about the availability of management information to support the decision-making process.

The necessary management tools for the hospital administration to perform the decision-making process must be readily available by CQIS. The architectural aspects of CQIS must provide the means to make available key-performance indicators (KPIs) to improve performance [4]. However, KPIs must be capable of representing quantitatively and qualitatively involved business magnitudes such as quality, cost, and delivery services [5]. Moreover, HIS operational adjustment needs the prioritization of the performed processes and it needs the application of evident tools such as Analytic Hierarchy Process (AHP) consisted of a structured technique for organizing the analyses of complex decisions [6]. Also, CQIS must be present the proper architecture facilitating the production route of offered services by the hospital providing the associated and necessary details for each workflow in a manner analogous to the one performed in the industry [7]. In addition, the CQIS architecture must provide the facilities to handle the failures of HIS to achieve its goals in a systematic manner [8]. Hence, the forth point of our work is about the availability of an architectural structure capable of providing readily available the necessary information to support the decision-making process.

The HIS functionality achieving to obtain the set aims and goals must be directed by a decision-making process based on evidential data. The applying methodologies require the proper software infrastructure sharing computer resources and data. The application of the theory of constraints [9] requires the availability of both the proper software architectural structure and the adequate data schemata. Hence, the fifth point of our work is about the availability of an architectural software structure along with the matching data structures allowing the accommodation of well-known methodologies such as ISO quality framework (ISO 9001), FMEA, Six Sigma and Lean Six Sigma.

3. Relative Literature

The approach of quality improvement is capable of transforming healthcare [10] provided the successful application of specific methodologies which are providing results rely on successive measurements of reliable data. The applying methodologies are based on the fundamental cycle of Quality (Plan-Do-Act-Check) and each method is specifically attempting to obtain the previously defined goals. The ISO standards (e.g. ISO 9001) are attempting to develop a managerial framework without providing the necessary warranties of the achieved level of quality since it is defined by the end-user based on the set requirements of operation [11, 12]. The reliability of ISO has been improved due to the carrying certification process by independent - notified bodies. The proposed CQIS approach is extending its capabilities beyond the ISO standards border lines to transform healthcare services in order to warranty the quality improvement.

Quality improvement is performed by repeatedly examining the obtained level of QoS applying at specific workflow point dedicated methods. However, the applied industrial methods of failure classifier [13] to assess the performing processes

reliability, the production route card [7] describing in details the conditions of the performing workflows, the FMEA method for preparing data to be fed for assessing and managing the associated risks of the carried processes [14], the Six Sigma method for smoothing the variation of procedures execution [15] meeting the patients' expectations, and the Lean Six Sigma method for minimization the waste of the resources [16] include time as a resource, they all have been considered independently. CQIS is called to facilitate the application of processing methods to provide intermediate or final results to support the decision-making process performed by the hospital's administration.

The achievement of a specified level of quality in the hospital, the organizational structure and the IT systems must be reconsidered reforming the operational terms [17]. There is the perception that quality improvement must be achieved by employing tools such the Disease Related Groups (DRGs) and internationally recognized codification systems such as the International Code of Diseases (ICD) [18] without providing the means to measure directly the level of quality. Instead, DGRs along with ICD provide the means for indirect measurement of the level of quality. However, the structural infrastructure of CQIS can accommodate both DRGs and ICD codifications. In the same direction, the employment of key-performance-indicators (KPIs) as evidences must be escorted by a clearly defined procedures including the activities of collecting data as evidences, analyzing and interpreting the data, auditing, an finally, feedback to managers [19]. On the other hand, CQIS provides the structural readiness to be capable of reforming the operational terms of HIS, accommodating tools such as DRGs, ICDs and KPIs.

CQIS requires support from a local multidisciplinary team of within the hospital to supervise total quality management and re-engineering activities [20]. In addition, hospitals must have available training services to support the operation of CQIS to overcome the inertia of change and the hesitation caused by the locally holding cultural conditions [20]. The local multidisciplinary team has a managerial instrument to examine the obtained achievements [21] which has to be implemented with software means and coupled with HIS. CQIS requires training services, as any other HIS software module, to explore the capabilities and apply the locally holding conditions. The experience and the accumulated knowledge earned from other sectors of business activities concerning quality, methodologies, and practices on IT systems has to be adjusted and applied to achieve improvement in the provided quality [22].

4. The HIS environment's requirements

The currently holding perception about the operation of a HIS is the adoption of as many international standards and protocols as possible. However, standards and protocols are addressing just a part of a HIS scope missing the entire spectrum of hospital operations and functionalities. In addition, the standards and protocols are incapable of meeting peculiarities revealed from the locally holding economic, social, and cultural conditions. Adopting standards and protocols in the everyday practice focuses on the application of well known and pre-approved good-practices. However, the hospitals are improvising applicable solutions with the locally holding conditions and thus jeopardizing the success of the standards and protocols application. Hence, the fidelity of the resulting application of the adopted standards and protocols depend on the embedment of the locally holding conditions at the hospital.

The hospitals choose the certification of the applying standards in order to prove the adaptation of specific standards and become more attractive to the relative market. The insurance sector and the market of healthcare urges hospitals to get certified with internationally recognized standards to ensure that the offered services follow predefined and acceptable methods.

The introduction of international standards assists the hospitals to adopt well recognized standards achieving the convergence to widely acceptable models such as Health Level Seven (HL7). The adoption of international standards in the HIS software provides the capability of minimal interventions to achieve, at least, semantic interoperability. The interoperability for hospitals presents two folds. Firstly, serving the patients when there are successive hospitalizations transferring vital information concerning the medical records from a hospital to another. Secondly, the comparison among hospitals requires the evaluation of compatible values base on analogously applying methods and procesures.

5. Quality and Quality measurements

The HIS user interfaces provide the means to acquire raw data through well and adequately designed forms. The collected data can be processed to reveal the holding level of quality. Also, the gathered data can feed properly designed indices to inform the hospitals' administration about more complex administrative mechanism, processes, and procedures. The HIS reporting system can present the obtained data in predetermined formats to express the level of the achieved productivity, pre-determined indices, and QoS of various operations and workflows.

The operation of HIS provides the capability to properly determine the parameters which are reflecting the level of the holding quality. Those parameters are fed with data from the HIS user interfaces and with proper processing can reveal the obtained performance. Therefore, the hospitals' administration can receive continuous flow of information regarding critical and important parameters from the hospitals' operations.

6. Functional performance of CQIS

The CQIS is a fictitious or virtual mechanism which is fused within HIS allowing to set the levels of acceptable performance. At the center of this virtual mechanism is evaluation providing the decision support means for continuous reconsideration of the set level of performance after reexamination of the set level of the achievements of execution of the applying parameters. Also, CQIS provides the organizational facilities for the examination and analyses of the embedded risk, hazards and opportunities. Hence, CQIS is a virtual mechanism with elements distributed in the entire body of HIS structure.

CQIS is giving the current level of quality by calculating and evaluating the available data obtained by HIS. The performed calculations provide indices -which are obtained from the gathered HIS data. The formed indices are assessed, appraised, and evaluated to provide additional information to support the decision support system of CQIS. The evaluation methodology includes well-known methods such as FMEA, PHA, FTA etc.

The selection of the quality indices must be chosen in a way that several confirmations and verifications are in place. The quality indices form an acyclic network structure characterized by interdependences among the participating nodes-indices. The inter-relationships among the indices provides the necessary cross-checking about the indices values and statistical trends.

7. Placement of a CQIS

CQIS must be considered as an independent module of HIS. However, CQIS must be coupled with HIS in order to provide the benefits of its use. The operation of CQIS is considered to be completely transparent to the end users of HIS. The CQIS must be facilitated by an adequate internal functional structure permitting the interconnection with HIS and match with the corresponding organizational structure.

The major building blocks of CQIS form an acyclic network. The fist building block refers to the declaration of the followed policies. The second building block refers to the aims and set targets. The third block concerns the management of the associated risk. The forth block is about the encountered exceptions. The fifth block refers to the internally and externally carried procedures. The sixth block concerns the duties of the users. The seventh block refers to the processing unit of the CQIS. The formed acyclic graph is depicted in Figure-1.

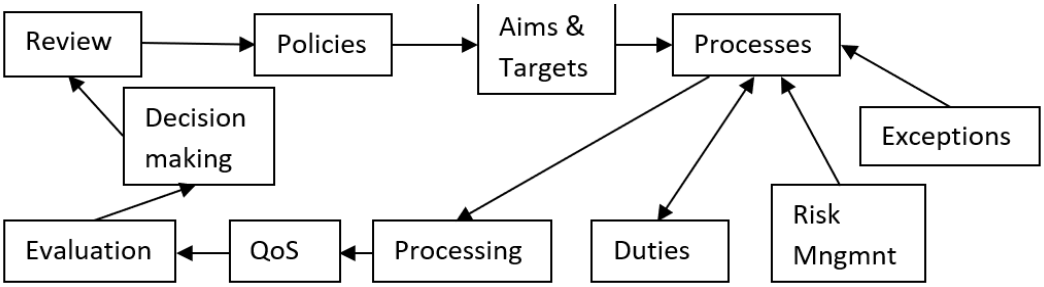


Figure-1. ***The CQIS internal structure***

CQIS is closely coupled with HIS. Such a coupling presents two layers, the first layer concerns the rules applying on the operation of HIS while the second layer is about the evidential data at the database level. Thus, for each HIS software module, there are two levels of correspondence between CQIS and HIS. A typical interaction of CQIS and HIS is depicted in Figure-2. CQIS supervises the operation of the applying rules for each HIS software module and obtains objective evidences about the precise application of the applying rules. The rules are consisted of hospital business instructions for performing the intended purposes. For instance, there is a set of rules applying on the patients' admission which directs the HIS software module to perform the intended work task in the pre-determined fashion. The applying rules contain and

manipulate the exceptions and the associated managerial restriction for mitigating the associated risk. The obtained evidences are collected and classified in order to be processed accordingly before the set evaluation procedures.

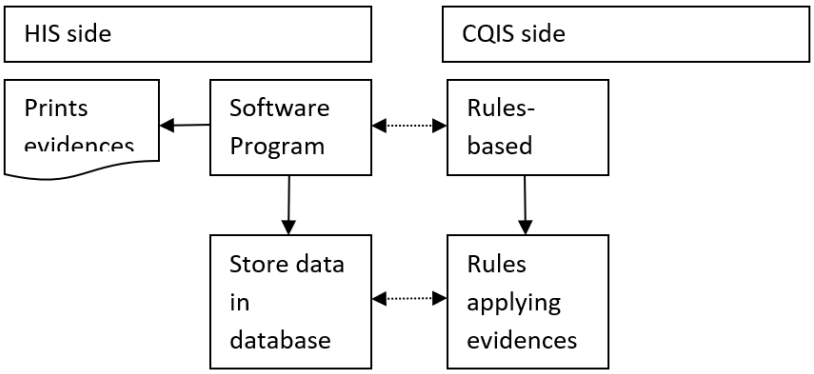


Figure-2. ***Coupling of CQIS and HIS***

8. HIS with a CQIS - organizational changes

The adoption of CQIS introduces additional roles in the organizational chart. Additional duties must be considered corresponding to the collection, evaluation, and review of the data to satisfy the decided and applying policies. Such duties are usually contributed to existing posts in the hospital's organizational chart forming committees to get synchronized under the supervision of an appointed manager. The newly introduced roles are about the supervision of the CQIS activities, decisions to be made for improving the offered QoS and advising of the hospital's administration. The representative of such a committee, usually a quality manager, presents the resulting performance data, the proposed decisions by the committee, and the set aims for improving the hospital's performance based on the evidential data. The rest of the roles which follow the HIS procedures are transparently supervised without realizing any CQIS intervention.

9. Implementation of a CQIS

The implementation of a typical CQIS involves at least two phases. The first one regards the development of the core modules of CQIS. The second phase is about the implementation of the interfaces with HIS. Both phases can be developed independently provided the adoption of proper standardization. In addition, the coupling of CQIS with HIS requires sufficient information about the data management and storage activity performed in HIS.

The implementation of CQIS must follow a bottom-up fashion. The processing unit must proceed to all the rest of the CQIS units since it receives data for processing from all the rest of the CQIS units. Then, it follows the development of the unit containing the users duties. The implementation carries on with the unit administering and managing the carried procedures since it must be connected with both the processing and the users' duties units. The implementation goes on with placing the exceptions handling unit which enriches further the carried procedures' unit. The implementation continues with the risk management unit. The development works carry on with the implementation of the quality targets' unit. The implementation ends with the development of the policies module. The described bottom-up development fashion allows the inclusion of ordinary quantified types of data along with the qualitative data that need some additional processing. Such a traversing requires the development of the proper qualitative scales to express properly the quantitatively obtained data.

The implementation of CQIS is usually following the installation of HIS expressing the need for continuous improvement. Thus, CQIS can either be an independent software module that eventually must be connected with HIS or be among the rest of the HIS software modules. In either case, CQIS must be coupled with HIS performing the necessary interconnections. The coupling of CQIS with HIS is obtained using software links either at the level of software applications or at the database level. In most of the cases. it is considered more efficient to perform the connection between CQIS and HIS at the storage level since it is completely transparent to the end-users.

10. Impact on HIS

The introduction of CQIS module impacts the HIS performance. The objective evidences obtained by the data gathered in HIS feed the data structures acting as criteria upon which decisions are made by CQIS. The decisions made upon the collected data guide the Hospital's Administration to change the applying policies or alter the operation of HIS in order to achieve the set goals. Such alterations in HIS operation impact the flow of the carried processes, the involved procedures, and the way certain works are performed with HIS. Consequently, accustomed processes, procedures, and specific works change or adapted to the expressed needs by CQIS that dictate the requirements to achieve the desired improvement.

The operation of CQIS provides valuable information to Hospital's Administration. The operation of HIS may change based upon the decisions made by the Administration in order to achieve the set quality characteristics. However, changes in HIS may cause changes to CQIS depending upon the extension of the performing changes. The performed changes in HIS must be documented in order to follow up the obtained versions of the resulting software system. Similarly, the decisions made by the Hospital's Administration must be properly documented to avoid repetitions and conflicts on the changing software code.

The changes dictated by the decisions made based on CQIS criteria cause changes in the HIS workflows and the sequences of the HIS interfaces. The performed changes in HIS operations must be adequately documented providing evidential support for the alterations and HIS versioning. The supporting documentation must provide the objective evidences along with the drafts of the existing and the changed process or procedure. The documentation must be classified according to CQIS demands. Hence, a reliable management scheme must continuously support the HIS maintenance.

The operation of CQIS provides valuable information regarding the performance of HIS and consequently, the efficiency of the Hospital's organizational goals referring to departments and clinics, as well as, the patients' treatments. The main effect from the operation of CQIS is about the demanded changes in the operation of HIS. In other words, HIS is no longer a static implementation but upon the obtained data, there are made decisions altering the flows of the carried processes. Therefore, the changes in HIS affect the way performed the personnel's daily works providing the required rationalized reasoning.

11. Implementation plan

The objective to develop CQIS is directly related with the desire to control the offered QoS of the hospital services applying HIS. Actually, CQIS is going to declare and dictate the necessarily required adjustments to the HIS' operation. CQIS provides the means to make available evidences about the required changes in the HIS' functionality. The evidences are received from the HIS operation and received transparently from the end-users. Then, the evidential data is presented properly to support the associated decision making process carried by the hospital administration. Hence, the aim of the implementation plan is the development of an IT system that receives transparently data from HIS, performs the adequate processing of the received data, presents the data to the decision makers, records the made decisions, and eventually, sets the level of QoS until the next observation of HIS performance. Therefore, the objective of the implementation plan is the development of an IT system that is going to follow the performance of the applying HIS and continuously regulating the its efficiency and effectiveness to obtain the set level of quality.

The operational characteristics of CQIS must be capable of allowing the achievement of coupling with the HIS that supports. Also, CQIS must present such a structure that it can cooperate with any HIS. The provision of CQIS with HIS data as well as feeding HIS with CQIS data can be accomplished with automatic set-ups, semi-automatic procedures, or manually. CQIS must be equipped with software modules that support the decision making process carried by the hospital's administration. In that sense, CQIS must be aware of the carried HIS processes, procedures, and work tasks. In addition, CQIS must accommodate and apply the decided polices as well as the level of quality desired to be achieved by HIS. Also, CQIS must facilitate the risk management and the periodic review of HIS operation in order to define the HIS operational exceptions. Moreover, CQIS must be equipped with software modules that achieve the development, after the proper processing, of the desired, in each case, report. Therefore, CQIS must operate in such a way that achieves cooperation with any installed HIS.

The implementation of CQIS is based on open source tools which are freely available over the internet. The operation of CQIS requires a browser since it uses HTML (hyper-text mark-up language) at the presentation level. Then, at the processing level PHP (hyper-text pre-processor) applies which is executed at the server supporting CQIS. Afterwards, at the storage level, CQIS is supported by a relational database management system such as MariaDB. The software structure of CQIS is consisted of dedicated software modules which are supported by web hooks or web services to exchange data

with the associated HIS. The coupling of CQIS and HIS depends on the HIS readiness to export operational data. In any case, CQIS must have available web services to receive plain text data for processing. On the other hand, CQIS must be capable of exporting data in the requested data formats which may require some additional programming effort. Hence, CQIS is going to be developed with well-known open source tools in order to provide the technical standardization which is necessary for splicing with HIS permitting the involvement of IT professionals without restrictions. Thus, CQIS must make publicly available the requirements for its connection with any HIS.

The implementation of CQIS provides the means to other information systems to transfer data. For each CQIS module there must be developed the corresponding software programs that allow the transfer of data from and to HIS. CQIS is planned to provide the set of RESTfull APIs along with the calling requirements, the characteristics of each API. and explanations about the usage of each API. Similarly, the authentication mechanism is going to be provided publicly for protection and security reasons. The authorized hospital IT administrators are going to be given access to dedicated web services to provide, automatically or manually, the necessary information required for the CQIS operation. The repertoire of the available web services are going to be provided at the CQIS administrator's maintenance web page.

The implementation plan for CQIS presents three stages. The first stage is consisted of eight (8) phases and concerns the development of the CQIS modules. The first phase involves the development of the database schema including registry of the carried activities and a dictionary for the adopted terms and concepts. The second phase refers to the development of the web services which are necessary for enriching the CQIS database tables with HIS data. The third phase concerns the development of the CQIS end-users interface. The fourth phase is about the development of the processing module. The fifth phase is dedicated to the development of the machine learning applications. The sixth phase is related to the development of the decision support system. The seventh phase concerns the reporting module. The last phase, is dedicated to the development of the administrator's software modules for the CQIS house-keeping.

The second stage is consisted of four (4) phases. The first phase of the second stage is about testing and debugging. The second phase concerns the trials of the developed CQIS including stress testing. The third phase refers to the development of the necessary on-line documentation of the developed system. The fourth phase of the second stage refers to the development of the training materials.

The last stage of the implementation plan is dedicated to the trial operation of the developed system. CQIS web services are going to be used to receive data from and send data to HIS. Then, the decision support module is going to be tested applying the predetermined use-cases. It follows the testing of the reporting module which indirectly provides evidences about the appropriateness in the operation of the processing module. At every step of the carried trials the CQIS documentation and the training material is going to be examined. The delivery of the developed system is going to include a training system environment, a programming development environment, and a completely functional CQIS system.

12. A practical Example

The implementation of the CQIS can applied on an existing HIS. The continuously carried process of the examination of quality for the inpatients and the financial sustainability of the hospital can be performed using DRGs. In such a case, DRGs are going to be followed by CQIS with the development of diseases related packages. Such packages can be defined and structured in CQIS and receive data from HIS while the packages performance evaluation can be carried by CQIS. The decisions made based upon the values of the developed packages can influence the operation of HIS. CQIS is equipped with the adequate facilities to respond to such challenges taking advantage of the registry and dictionary software modules. Thus, the hospital's administration can make decisions on which parameters of the HIS operations must be properly adjusted and which flows must be altered.

13. Conclusions and future Research

Connecting CQIS with HIS interfaces is constituted a dynamic software system. The functionality of CQIS provides the evidences for the HIS changes in order to achieve the desired quality level. In other words, the operation of CQIS defines the operational areas of HIS that must undergo changes. The decision making is still left to the dedicated personnel receiving the results of the CQIS operations. Such documented results are taken advantage by the Administration to order the HIS changes.

The existence of CQIS is necessary along with the operation of HIS since it constantly regulates the HIS operation to achieve the quality aims. The defined level of the offered quality is examined and evaluated processing the collected HIS data. The performed processing on the collected data is carried out in accordance to the instructions provided by CQIS. In

other words, the operation of CQIS is analogous to a feedback loop providing the means to make the specific changes for the performance improvement. The major benefit of the CQIS operation is accommodation, facilitation, and performance comparisons which are based on quantitative and qualitative data. The evaluations performed by CQIS lead to make decisions by the Administration in accordance to the provided policies.

The embedded evaluation methodologies and the applied strategies in CQIS can be performed either by the dedicated personnel or by the properly designed software modules. The employment of machine learning and artificial intelligence methods provide the capability to automate the CQIS functionality regulating the nature of obtaining continuous improvement on the HIS operation. However, the introduction of artificial intelligence weaves CQIS into HIS making almost transparent its operation. Moreover, the development of software interfaces controlled by artificial intelligence methods brings CQIS to become an indispensable part of HIS.

The employment of CQIS in HIS provides the means to control the operation of HIS in order to achieve the set level of quality. Thus, CQIS must be equipped with the necessary agility to respond to the HIS performance accordingly. In addition, CQIS must be capable of presenting a decision support system for the HIS empowering users to change the HIS operation to achieve the desired functionality which leads to the desired level of quality.

References

[1] O'Brien, J., & Marakas, G. M., Management Information Systems, (2008), The McGraw-Hill Companies, New York, USA.

[2] Vassiliadis, P., Simitsis, A., & Skiadopoulos, S., Conceptual modeling for ETL processes, In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, (2002), ACM, 14–21.

[3] Inmon, W. H., Building the data warehouse, (2005), John Wiley & Sons, Vol.3.

[4] Barchetti, U., Bucciero, A., Guido, A. L., Mainetti, L., & Patrono, L., Supply Chain Management and Automatic Identification Management convergence: Experiences in the Pharmaceutical Scenario, Supply Chain Coordination and Management, (2011), 978–953.

[5] Jacoby, D., Measuring sourcing performance: What's the mystery?, Purchasing, (2005), Vol.134.

[6] De Ambroggi, M., & Trucco, P., Modelling and assessment of dependent performance shaping factors through analytic network process, Reliability Engineering & System Safety, (2011), Vol.96, 7, 849–860.

[7] Production Route Card, Available from: http://www.enotes.com/americanscholar/q-and-a/what-meant-by-job-card-route-card-used-production-99397, accessed on 30. 04. 2019.

[8] Sahno, J., Shevtshenko, E., Karaulova, T. and Tahera, K., Framework for continuous improvement of production processes, Journal of Engineering Economics, (2015), Vol.26, 2, 169–180.

[9] Goldratt, E. M., Schragenheim, E., & Ptak, C. A., Necessary but not sufficient, (2011).

[10] Batalden PB, Davidoff F What is "quality improvement" and how can it transform healthcare? BMJ Quality & Safety 2007;16:2-3. (available at: https://qualitysafety.bmj.com/content/16/1/2, Accessed: 07.04.2019)

[11] Reid, R. Dan. and Sanders, Nada. R., Operations Management: An Integrated Approach, (2004), Management Science, Vol.2.

[12] ISO 9000 Standards, (available at: http://www.iso.org/iso/home/standards/management-standards/iso_9000.htm, Accessed: 07.04.2019).

[13] Karaulova, T., Kostina, M and Sahno, J., Framework of reliability estimation for manufacturing processes, Mechanics, (2012), Vol.18, 6, 713–720.

[14] ICH Q9 Quality Risk Management, European Medicines Agency Inspections, (2006)

[15] Van den Heuvel, J., Does, R. J., & De Koning, H., Lean Six Sigma in a hospital, International Journal of Six Sigma and Competitive Advantage, (2006), Vol.2, 4, 377–388.

[16] Sahno, J., & Shevtshenko, E. Quality Improvement Methodologies for Continuous Improvement of Production Processes and Product Quality and Their Evolution, In 8th International DAAAM Baltic Conference Industrial Engineering, (2014), 181–186.

[17] Hudelson P, Cléopas A, Kolly V, et al What is quality and how is it achieved? Practitioners' views versus quality models BMJ Quality & Safety 2008;17:31-36.

[18] Sabanović Z, Masić I., Computerized information system support in continuous quality improvement in hospital care, Med Arh. 2001;55(2):113-6. Croatian, PMID: 11769040.

[19] Jan Mainz, Birgitte Randrup Krog, Bodil Bjørnshave, Paul Bartels, Nationwide continuous quality improvement using clinical indicators: the Danish National Indicator Project, International Journal for Quality in Health Care, Volume 16, Issue suppl_1, April 2004, Pages i45–i50, https://doi.org/10.1093/intqhc/mzh031

[20] Klein, Donald; Motwani, Jaideep; and Cole, Beth, "Continuous Quality Improvement, Total Quality Management, and Reengineering: One Hospital's Continuous Quality Improvement Journey" (1998). Peer Reviewed Articles. 6. https://scholarworks.gvsu.edu/mgt_articles/6

[21] Stacie Petter, William DeLone & Ephraim McLean, Measuring information systems success: models, dimensions, measures, and interrelationships, European Journal of Information Systems Volume 17, 2008 - Issue 3, Pages 236-263 | Received 30 Dec 2006, Accepted 15 May 2008, Published online: 19 Dec 2017 (available at: https://doi.org/10.1057/ejis.2008.15).

[22] F. Bruder Stapleton MD, James Hendricks PhD, Patrick Hagan MHSA, Mark Del Beccaro MD, Modifying the Toyota Production System for Continuous Performance Improvement in an Academic Children's Hospital, Pediatric Clinics of North America, Volume 56, Issue 4, August 2009, Pages 799-813.

An Empirical Analysis of the Relationship among Foreign Direct Investment, Gross Domestic Product, CO_2 Emissions, Renewable Energy Contribution in the context of the Environmental Kuznets Curve and Pollution Haven Hypothesis Regarding Turkey

Mehman Karimov

Kaposvár University, Faculty of Economic Science

Abstract

This study examines the relationships between GDP per capita, CO_2 emission, Renewable Energy Contribution (REC) and Foreign Direct Investment (FDI) and evaluates the Environmental Kuznets Curve (EKC) and Pollution Haven Hypothesis (PHH) for Turkey. The EKC theory says that with increase in income per capita the pollution also increases but in a turning point when nation become richer pollution starts to decrease according to stringency of environmental regulations and implying advanced green technologies due to requirement of nation. In another hand the PHH assume that due to stringency of environmental regulations and high taxes the production become more expensive in developed countries, thus those dirty industries shifts from environmentally stricter developed countries to poor regulated developing countries. The aim of this study to analyze and investigate: which theory (EKC or PHH) does exist in Turkish economy and does FDI has positive impact on sustainable development. The time series datasets (FDI, GDP, CO2 and REC) , those were obtained from World Bank database, which covers the time period 1970-2014 were utilized in employed statistical models as the ADF Unit Root, Philips – Perron, Johansen co-integration, and the Granger Causality tests, to accomplish the empirical part of the paper. Based on the empirical results, it was approved that there wasn`t existence of the EKC theory in Turkish economy. But according to obtained empirical results it was affirmed that there was the presence of the PHH theory in Turkish economy which means the FDI has a negative impact on sustainable development of Turkish economy. Thus, the developed countries with stricter environmental regulations (mostly from Europe) relocate their heavily polluted dirty industries to Turkish economy.

Keywords: Foreign Direct Investment, CO_2 Emissions, Gross Domestic Product, Renewable Energy Contribution, Environmental Kuznets Curve, Pollution Haven Hypothesis, ADF and Philips-Perron Unit root tests, Johansen Co-integration test, Granger Causality test

1. Introduction

The protection of the World from environmental degradations has become one of the most important missions of nations. The one of the urgent cases among pollution is a carbon dioxide (CO2) emission which comes from human activities such as: cement production, deforestation as well as the burning of fossil fuels like coal, oil and natural gas. The first time this urgent issue came up on the agenda by United Nations Framework Convention on Climate Change in 1992, the international community has entrusted its secretariat with a growing responsibility to strengthen the global response to climate change and close the gap between ambition and achievement ("UN Climate Change Annual Report 2017 Website Now Live | UNFCCC," n.d.). In 1997, the world`s nations recognized the significance of environmental pollution with the increasing level of CO2 emissions, Parties to the Convention adopted the Kyoto Protocol, which created binding emission reduction targets for developed countries. During its first commitment period, 2008– 2012, 36 industrialized countries and the European Union pledged to reduce their emissions by an average of just over five per cent compared with 1990 levels ("UN Climate Change Annual Report 2017 Website Now Live | UNFCCC," n.d.). In Paris in December 2015, countries bound oneself to limit the rise of global average temperature to well below 2 °C and as close as possible to 1.5 °C above pre-industrial levels and to prevent dangerous human-induced climate change with signing the latest and ambitious Paris Agreement ("UN Climate Change Annual Report 2017 Website Now Live | UNFCCC," n.d.).

Furthermore, with globalization processes, liberalization and free movement of capital made foreign direct investment inflows to be rapidly intensified in 1980-1990s in the world arena. There is both a positive and negative impact of FDI for the host country. As a positive effect of FDI on the host country it can be considered the flow of finance, capital, new

management and new advanced technology. Thus, they can encourage the country to replace old technologies with modern environmentally friendly technologies. But the environmental pollutions can also be raised through FDI (capital intensive dirty industries). We can take in consideration 2 hypotheses in this case, Environmental Kuznets Curve (EKC) and Pollution Haven Hypothesis (PHH). The EKC theory says that, with increase in income per capita pollution, also, increases but in a turning point when nation become rich they will require clean environment and the government will strengthen environmental regulations and due to this issue (to avoid high taxes) the companies will apply clean technologies for production. The PHH says that because of stringent environmental regulations and high taxes, the developed countries transfer their heavy-polluting industries to lax environmental regulated developing countries.

The main purpose of this study reconsider the relationship between Foreign Direct Investments, CO2 emissions, Renewable Energy Contribution and Gross Domestic Product. Additionally, to investigate which theory (EKC or PHH) exists in economy of Turkey. The content of paper will be structured as follows: In section 2 literature reviews will be expounded, in section 3 data description will be shown, in section 4 methodology will be disclosed, in section 5 empirical results from employed statistical analysis will be expounded, in section 6 conclusion will be described and finally in section 7 references will be shown.

2. Literature review

2.1 Theoretical studies

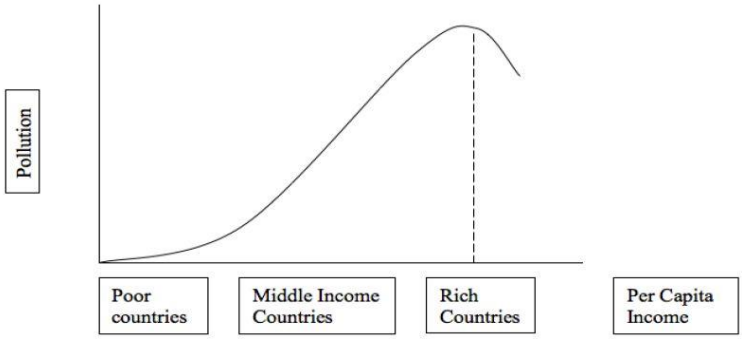
2.1.1 Environmental Kuznets Curve hypothesis (EKC) and FDI linkage

The Environmental Kuznets U-shaped Curve was invented by Simon Kuznets in 1955 (later was awarded with Nobel Prize in 1971), in which the relationship between growth in per capita income and environmental degradation illustrated by U-shaped Curve. Therefore, EKC hypothesis was introduced and became more popular with Grossman and Krueger`s publications (G. Grossman & Krueger, 1991; G. M. Grossman & Krueger, 1995) and World Bank Report (Shafik & Bandyopadhyay, 1992). According to EKC theory (opposite to PHH) the pollution is increased in poor nations with economic growth and is decreased when nation reaches higher income level, based on the relationship between income level and environmental degradation takes the inverted U-shape (Kerekes, Marjainé Szerényi, & Kocsis, 2018). This happens because developing (poor) countries don`t have enough advanced technologies or capital for implementing such kind of technologies for pollution abatement versus developed (rich) countries have enough level of technologies and capital for implementing these technologies (K. V. Murthy & Gambhir, 2018) (See Graph 1). Therefore, Foreign Direct Investment flows from developed countries which bear to uphold strict environmental standards, because of stringent regulations. Those companies can use energy-saving production which the developed countries transfer from home countries to invested country to influence for energy-efficiency in the host country. Thus, in the context of above mentioned claims, the Foreign Direct Investment can transfer new type of advanced environmentally friendly technologies and human capital, based on these the FDI can have positive effect on environmental performance and raise environmental standards through the transfer of cleaner technology and better management practices, which in turn leads to the incline in the usage of renewable energy and decreasing of CO2 emissions (Marton & Hagert, 2017).

2.1.2 Pollution Haven Hypothesis (PHH) and FDI linkage

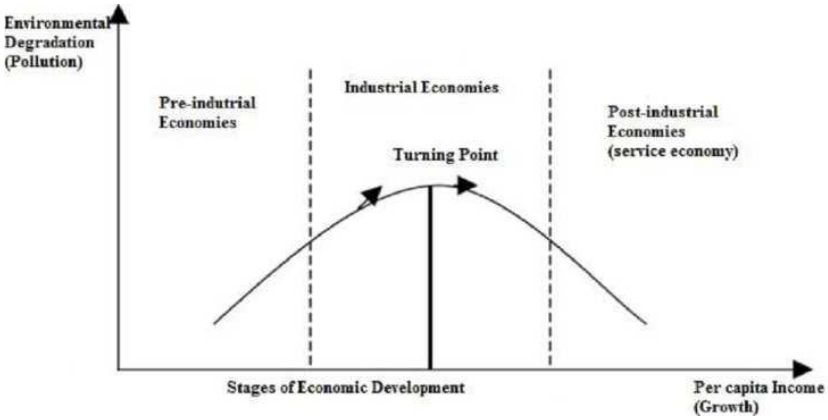
The Pollution Haven Hypothesis implies that polluting industries will relocate to environmentally poorly regulated administrations. The PHH was first postulated against of EKC theory by Copeland and Taylor (Copeland & Taylor, 1994) in the context of North-South trade under NAFTA. It was the first paper that links the environmental regulation stringency and trade patterns with the level of pollution in a country (Gill, Viswanathan, & Karim, 2018). The PHH argues that, the developed countries transfer the pollution intensive industries to less developed countries with unrestricted environmental regulations to avoid huge amount of taxes. Thus, if we will apply EKC U-curve for PHH we can see that till the developing country will reach to developed ones, the pollution will be increased parallel with economic growth, but after in a certain amount of costs of pollution the developed countries will shift pollution intensive industries to the developing countries with lax environment standards (See Graph 2). Afterwards developing countries will see increase in economic growth but parallel increase in pollution due to pollution intensive industries which were transferred from developed countries (Levinson & Taylor, 2008). The countries that apparent to become pollution havens are capital intensive, thus will attract MNCs (Foreign Direct Investment) with capital-intensive industries. In the developing countries (host countries) capital is cheaper (capital intensive industries) therefore, by definition these kind of capital is more pollution intensive. Thus, it can affect negatively in the usage of renewable energy and increase of CO2 emissions (Marton & Hagert, 2017).

Graph 1: Environmental Kuznets Curve



Source: **EKC** (K. B. Murthy & Bhasin, 2016)

Graph 2: Illustrating PHH in EKC Curve (A development- environment relationship)



Source: (Panayotou, n.d.)

2.2 Empirical studies

The relationship between GDP, FDI and Environmental degradation has been the subject of intense research during the last decades. The empirical studies gave the various results due to analyzed country, amount of series and applied empirical models. Thus, some of those studies showed the existence of EKC and vice versa in analyzed region. In another hand some of those studies also showed existence of PHH and vice versa in examined country. Additionally, in some research it has been found that FDI has positive relationship with environmental degradation and vice versa. For instance, (Hitam & Borhan, 2012), analyzed the Gross Domestic Product (GDP) growth and the environmental degradation for the time period 1965 – 2010 in Malaysia. The relationship between foreign direct investment and environmental degradation has been examined by applying the non-linear model. The results indicated that environmental Kuznets curve existed and foreign direct investment increased environmental degradation. (Khandker, Amin, & Khan, 2018) explored the relationship between FDI and renewable energy consumption from 1980-2015 in the context of Bangladesh. The Johansen`s co-integration test, Granger Causality test, VECM and Cusum test were applied to analyze series. The Johansen's co-integration test confirms that variables are co-integrated in the long run and Granger causality test reveals that there is bidirectional causality between those variables of interest. Through Vector Error Correction Model (VECM), they found no causality between the variables in the short run. The CUSUM test results show that variables were stable. According to results Policies regarding attracting more FDI should be considered to increase the investment in Renewable energy sector. In the another hand, (Aslanidis & Iranzo, 2009), examined the relationship between growth in per capita income and environmental degradation from 1971-1997. The smooth transition regression (STR) model has been applied for empirical part. The results supported EKC hypothesis. (Jbara, n.d.), examined whether to find evidence that these changes over time

are consistent with the PHH or EKC for a three different year 1990, 1995, 2000 in 36 countries. The linear regression model and descriptive statistics were utilized for empirical part. According to the results, there was little evidence to suggest that an EKC existed. There was no evidence to support the PHH. (Cole, 2004), examined the extent to which the EKC inverted U relationship can be explained by trade and specifically the migration or displacement of 'dirty' industries from the developed regions to the developing regions (the pollution haven hypothesis (PHH)). According to results the PHH was existed. (Naz et al., 2019), examined the relationship between renewable energy consumption (REC), foreign direct investment (FDI) inflows, economic growth, and their resulting impact on CO2 emissions for the time period 1975-2016 in Pakistan. The results show that economic growth and FDI inflows both increased CO2 emissions, while REC substantially decreased CO2 emissions during the studied time period. The results do not support the inverted U-shaped Environmental Kuznets Curve (EKC) hypothesis for per capita income (and FDI inflows) and per capita CO2 emissions in a country. The results supported 'pollution haven hypotheses where FDI inflows damage the natural flora of the country. (Leitão, 2014) examined the correlation between economic growth, carbon dioxide emissions, renewable energy and globalization for the period 1970-2010, using time series (OLS, GMM, unit root test, VEC model, and Granger causality) in Portuguese economy. OLS estimator and GMM model demonstrated that carbon dioxide emissions and renewable energy are positively correlated with economic growth. The econometric models also show that the overall index of globalization had a positive effect on growth. The Granger causality reported a unidirectional causality between renewable energy and economic growth. (Balibey, 2015) examined the causal relationships between economic growth, carbon dioxide emission and foreign direct investment (FDI) and evaluates the Environmental Kuznets Curve (EKC) hypothesis for Turkey in 1974-2011. The causality relationships investigated by using the Johansen Co-integration test, The Granger Causality Test, Impulse-Response and Variance Decomposition Analysis of vector autoregression model (VAR) model. The causality relationships displayed that FDI (LFDI) and economic growth (LGDP) had a significant effect on carbon dioxide emissions (LCO2). Moreover, impulse-response functions and variance-decompositions of VAR model supported these relationships among LGDP, LCO2 and LFDI. The study investigated the validity of the EKC hypothesis in Turkey for the period 1974-2011 by using regression model approach for the various EKC model forms such as linear, quadratic, and cubic. Consequently, economic growth leaded to degradation of environment and depletion of natural resources. (Pazienza, 2015) examined the relationship between Foreign Direct Investment (FDI) and the environment from 1981-2005 by using panel data analysis. The result of the analysis showed the existence of negative relationships characterizing the technique, scale and cumulative effects of FDI on CO2. (Linh & Lin, 2014) examined the dynamic relationships between CO2 emissions, energy consumption, FDI and economic growth for Vietnam in the period 1980-2010 based on Environmental Kuznets Curve (EKC) approach. The empirical results obtained from co-integration, and Granger causality tests didn`t supported the EKC theory in Vietnam. However, the co-integration and Granger causality test results indicate a dynamic relationship among CO2 emissions, energy consumption, FDI and economic growth. The short-run bidirectional relationship between Vietnam's income and FDI inflows implied that the increase in Vietnam's income would attract more capital from overseas. (Mert & Bölük, 2016) examined the impact of foreign direct investment (FDI) and the potential of renewable energy consumption on carbon dioxide (CO2) emissions in 21 Kyoto countries. Panel causality tests showed that there were significant long-run causalities from the variables to carbon emissions, renewable energy consumption, and fossil fuel energy consumption and inflow foreign direct investments. The results of their model supported the pollution haloes hypothesis which stated that FDI brought clean technology and improved the environmental standards. However, an inverted U-shaped relationship (EKC) was not supported by the estimated model for the 21 Kyoto countries. Another important finding was that renewable energy consumption decreased carbon emissions.

3. Data Description

This investigation considers the secondary time series dataset, which was obtained from the World Bank Database[1] for the period span from 1970 to 2014. All variables were converted into logarithms namely LnFDI, LnGDP, LnCO2 and LnREC. The Eviws-8 has been used for empirical part of paper. These four variables were utilized in the model:

FDI – Foreign direct investment: Inward and outward flows and stock, annual (current US$)

GDP – GDP per capita (US$)

CO2 – Carbon dioxide emissions (kt) are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced

[1]World Bank Database
https://data.worldbank.org/country/turkey

REC - Renewable Energy (toe) contribution of renewables to total primary energy supply (TPES)

4. Methodology

4.1 Augmented Dickey-Fuller Unit Root Test

The Augmented Dickey-Fuller test was developed by David Dickey and Wayne Fuller, American statisticians, in 1979. The Dickey-Fuller test is used to determine whether a unit root, a feature that can cause issues in statistical inference, is present in an autoregressive model. [1] ADF test equation is[2] (1):

$$yt=c+\delta t+\phi y_{t-1}+\beta 1\Delta y_{t-1}+\ldots+\beta p\Delta y_{t-p}+\varepsilon t \ldots\ldots\ldots(1)$$

Where

Δ is the differencing operator, such that $\Delta yt=yt-yt-1$.

The number of lagged difference terms, *p*, is user specified.

εt is a mean zero innovation process.

The null hypothesis of a unit root is

H0: ϕ =1.

Under the alternative hypothesis, ϕ <1.

Variants of the model allow for different growth characteristics. The model with $\delta = 0$ has no trend component, and the model with $c = 0$ and $\delta = 0$ has no drift or trend. The test that fails to reject the null hypothesis, fails to reject the possibility of a unit root.

To estimate the significance of the coefficients in focus, the modified T (Student)-statistic (known as Dickey-Fuller statistic) is computed and compared with the relevant critical value: if the test statistic is less than the critical value then the null hypothesis is rejected. Each version of the test has its own critical value which depends on the size of the sample[3].

4.2 Philips-Perron Unit Root Test

The Phillips-Perron (PP) unit root test was developed by statisticians, Peter C.B. Phillips and Pierre Perron, in1988. Though the PP unit root test is similar to the ADF test, the primary difference is in how the tests each manage serial correlation. Where the PP test ignores any serial correlation, the ADF uses a parametric autoregression to approximate the structure of errors.[4] The mathematical equation of test is[5] (4)

$$yt = c + \delta t + a\, y_{t-1} + e(t) \ldots\ldots(4)$$

Where $e(t)$ is the innovations process.

The test assesses the null hypothesis under the model variant appropriate for series with different growth characteristics ($c = 0$ or $\delta = 0$). To estimate the significance of the coefficients in focus, the modified T (Student)-statistic (known as Phillips-Perron statistic) is computed and compared with the relevant critical value: if the test statistic is less than the critical value

[1] ThoughtCo, The Augmented Dickey-Fuller Test
https://www.thoughtco.com/the-augmented-dickey-fuller-test-1145985
[2] MathWorks, The Augmented Dickey-Fuller Test
https://www.mathworks.com/help/econ/adftest.html?s_tid=doc_ta
[3] RTMath, Mathematics experts in quantitative finance
https://rtmath.net/help/html/93a7b7b9-e3c3-4f19-8a57-49c3938d607d.htm
[4] ThoughtCo, The Augmented Dickey-Fuller Test
https://www.thoughtco.com/the-augmented-dickey-fuller-test-1145985
[5] MathWorks, Phillips-Perron test for one unit root
https://www.mathworks.com/help/econ/pptest.html

then the null hypothesis is rejected. Each version of the test has its own critical value which depends on the size of the sample.

4.3 Johansen Co-integration test

The Johansen co-integration test was developed by Danish statistician, Soren Johansen, in 1991. It is a statistical model for testing co-integration between several series those are integrated in order *I*(1) at 1st difference through trace and eigenvalue tests. The mathematical equation of test is[1] (5):

$y_t = \mu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t$ ..... (5)

H_0= there is no co-integration between analyzed series.

H_1= there is at most 1 co-integration between analyzed series.

Null hypothesis or alternative hypothesis will be accepted if p-value > 0.05.

4.4 Granger Causality Test

The Granger causality test was developed by British statistician, Sir Clive William John Granger in 1969. It is a statistical concept of causality that is based on prediction. According to Granger causality, a variable X is causal to variable Y if X is the cause of Y *or* Y is the cause of X[2]. The mathematical equation of test is (6):

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_m y_{t-m} + error_t$ .....(6)

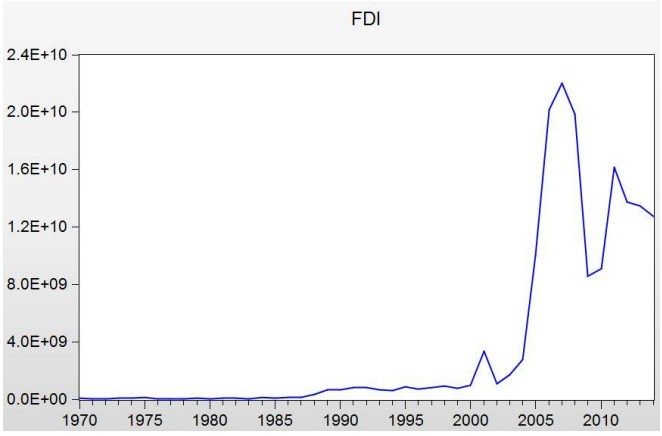
H_0= X doesn`t Granger Cause Y and Y doesn`t Granger Cause X.

Null hypothesizes will be accepted if p-values is more than 0.05.

5. Empirical Results

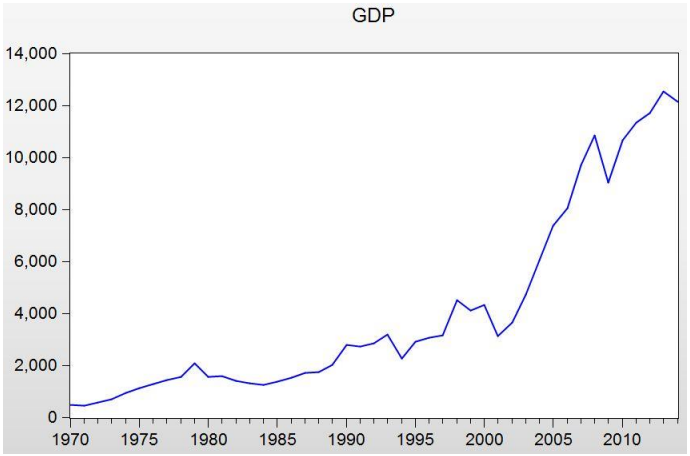
The time series plots for FDI, GDP, CO2 and REC were illustrated in Figure 1 for checking stationary of time series.

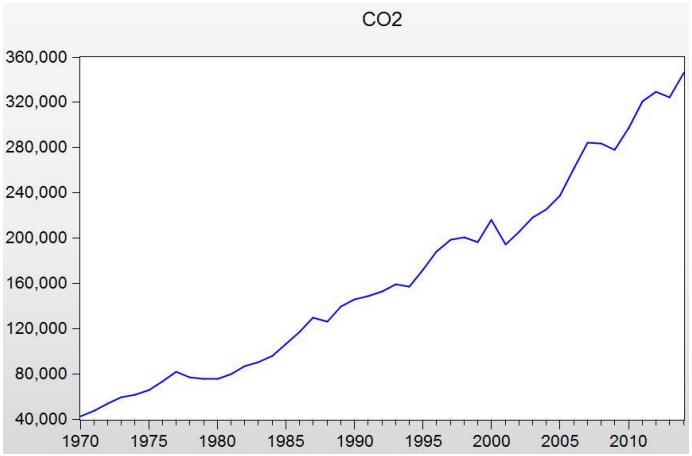
Figure 1: Time Series plots for FDI, GDP, CO2 and REC

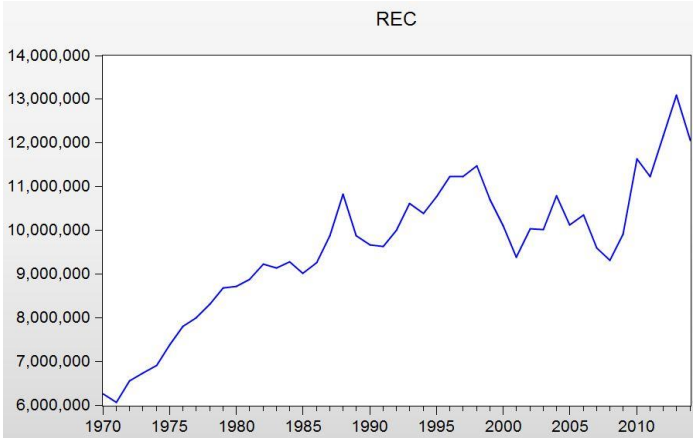


[1]IMF - International Monetary Fund, Testing for Co-integration Using the Johansen Methodology when Variables are Near-Integrated
https://www.imf.org/external/pubs/ft/wp/2007/wp07141.pdf

[2] Statistics How To, Granger Causality Test
https://www.statisticshowto.datasciencecentral.com/







Source: Author`s own invention based on World Bank Database (E-views 8)

According to graphs it has been seen that all series are non-stationary. Furthermore, the Augmented Dickey-Fuller and Phillips-Perron Unit Root Tests have been applied to determine the stationary of time series.

5.1 Augmented Dickey-Fuller Unit Root Test

As the pre-condition of Johansen co-integration test proposes, selected time-series must be non-stationary at a level and stationary at the 1st difference. Thus, the ADF test individually has been performed on the variables. According to the result of ADF test, the null hypothesis that series has a unit root at levels should be accepted, because T-statistics are less than critical values at 1% and 5% level of significance and P-values of variables are more than 0.05. Thus, after taking the first difference, the series became stationary according to these outputs: T-statistics more than critical values at 5% level of significance and P-values less than 0.05. Based on results, the null hypothesizes that series have unit root at 1st difference should be rejected. Thus, ADF results showed that the observed series appeared to be integrated of order one (I (1)) (See Table 1).

Table1: Augmented Dickey Fuller unit root test results

Null Hypothesis: (LnCO2) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnCO2 at level	-1.475990	5%	-2.931404	0.5361	Non-stationary
Null Hypothesis: (DLnCO2) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnCO2 at 1st difference:	-5.517142	5%	-2.933158	0.0000	Stationary
Null Hypothesis: (LnGDP) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnGDP at level:	-1.294479	5%	-2.931404	0.6237	Non-stationary
Null Hypothesis: (DLnGDP) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnGDP at 1st difference:	-3.962755	5%	-2.933158	0.0037	Stationary
Null Hypothesis: (LnREC) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnREC at level:	-2.484286	5%	-2.931404	0.1262	Non-stationary
Null Hypothesis: (DLnREC) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnREC at 1st difference:	-4.464795	5%	-2.933158	0.0009	Stationary
Null Hypothesis: (LnFDI) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnFDI at level:	-0.439081	5%	-2.931404	0.8930	Non-stationary
Null Hypothesis: (DLnFDI) has a unit root					
Variables	ADF Test Statistic	Level	Critical values	Prob*	Conclusion
LnFDI at 1st difference:	-5.518338	5%	-2.933158	0.0000	Stationary

Source: Author`s own calculations

5.2 Philips-Perron Unit Root Test

Additionally, Philips-Perron Unit Root Test was performed for checking stationary level of series. According to the result of PP test, the null hypothesis that series has a unit root at levels should be accepted, because T-statistics are less than critical values at 1% and 5% level of significance and P-values of variables are more than 0.05. Thus, after taking the first difference, the series became stationary according to these outputs: T-statistics more than critical values at 5% level of significance and P-values less than 0.05. Based on results, the null hypothesizes that series have unit root at 1st difference should be rejected. Thus, PP results showed that the observed series appeared to be integrated of order one (I (1)) (See Table 2).

Table 2: Philips – Perron Unit Root Test results

Null Hypothesis: (LnCO2) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnCO2 at level LnCO2	-1.984542	5%	-2.929734	0.2923	Non-stationary
Null Hypothesis: (DLnCO2) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnCO2 at 1st difference:	-6.089665	5%	-2.931404	0.0000	Stationary
Null Hypothesis: (LnGDP) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnGDP at level:	-0.991199	5%	-2.929734	0.7483	Non-stationary
Null Hypothesis: (DLnGDP) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnGDP at 1st difference:	-6.833714	5%	-2.931404	0.0000	Stationary
Null Hypothesis: (LnREC) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnREC at level:	-2.050739	5%	-2.929734	0.2650	Non-stationary
Null Hypothesis: (DLnREC) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnREC at 1st difference:	-6.465150	5%	-2.931404	0.0000	Stationary
Null Hypothesis: (LnFDI) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnFDI at level:	-0.452175	5%	-2.929734	0.8907	Non-stationary
Null Hypothesis: (DLnFDI) has a unit root					
Variables	PP Test Statistic	Level	Critical values	Prob*	Conclusion
LnFDI at 1st difference:	-10.69451	5%	-2.931404	0.0000	Stationary

Source: Author`s own calculations

5.3 Johansen Co-integration Test

Based on the ADF unit root test our series are integrated of the same order, I(1) which means the Johansen co-integration test has been allowed to perform. Johansen co-integration test has been employed for LnCO2 and LnGDP to analyze the long-run relationship between these series. According to the obtained Johansen co-integration test results, those based on trace test and maximum eigenvalue test (p-values in both tests = 0.5472 and 0.7976>0.05), the null hypothesis that there is no co-integration between LnCO2 and LnGDP has been accepted. And the null hypothesis that there is at most 1 co-integration between analyzed series is rejected. It has been confirmed that there is no co-integration between analyzed series (See Table 3).

Table 3: Johansen Co-integration test for LnCO2 and LnGDP

Johansen Co-integration test: Sample (adjusted): 1972-2014, Included obs.: 43, Series: LnCO2, LnGDP, Lags interval (in first differences):1 to 1.				
Unrestricted Co-integration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.115322	7.263637	15.49471	0.5472
At most 1	0.045331	1.994801	3.841466	0.1578
Unrestricted Co-integration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.115322	5.268836	14.26460	0.7076
At most 1	0.045331	1.994801	3.841466	0.1578

Source: Author`s own calculations

The Johansen co-integration test was employed for the next step for LnREC and LnGDP to analyze the long-run relationship between them. According to the obtained Johansen co-integration test results, those based on trace test and maximum eigenvalue test(p-values in both tests = 0.1806 and 0.1793>0.05) the null hypothesis is that there is no co-integration

between LnREC and LnGDP, has been accepted. And the null hypothesis that there is at most 1 co-integration between analyzed series is rejected. It has been confirmed that there is no co-integration between analyzed series (See Table 4).

Table 4: Johansen Co-integration test for LnREC and LnGDP

Johansen Co-integration test: Sample (adjusted): 1972-2014, Included obs.: 43, Series: LnREC, LnGDP, Lags interval (in first differences):1 to 1.				
Unrestricted Co-integration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.217252	11.53565	15.49471	0.1806
At most 1	0.023056	1.003023	3.841466	0.3166
Unrestricted Co-integration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.217252	10.53263	14.26460	0.1793
At most 1	0.023056	1.003023	3.841466	0.3166

Source: Author`s own calculations

The Johansen co-integration test was employed for the further analyzing for LnCO2 and LnFDI to analyze the long-run relationship between them. According to the obtained Johansen co-integration test results, those based on trace test and maximum eigenvalue test (p-values in both tests = 0.0023 and 0.0077<0.05) the null hypothesis is that there is no co-integration between LnCO2 and LnFDI, has been rejected. It has been confirmed that there is at most 1 co-integration between analyzed series (p-values in both tests = 0.0740> 0.05) (See Table 5).

Table 5: Johansen Co-integration test for LnCO2 and LnFDI

Johansen Co-integration test: Sample (adjusted): 1972-2014, Included obs.: 43, Series: LnCO2, LnFDI, Lags interval (in first differences):1 to 1.				
Unrestricted Co-integration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.383898	29.08865	20.26184	0.0023
At most 1	0.174806	8.261897	9.164546	0.0740
Unrestricted Co-integration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.383898	20.82675	15.89210	0.0077
At most 1	0.174806	8.261897	9.164546	0.0740

Source: Author`s own calculations

The Johansen co-integration test was employed for the next step for LnREC and LnFDI to analyze the long-run relationship between them. According to the obtained Johansen co-integration test results, those based on trace test and maximum eigenvalue test (p-values in both tests = 0.2981 and 0.2940>0.05) the null hypothesis is that there is no co-integration between LnREC and LnFDI, has been accepted. And the null hypothesis that there is at most 1 co-integration between analyzed series is rejected. It has been confirmed that there is no co-integration between analyzed series (See Table 6).

Table 6: Johansen Co-integration test for LnREC and LnFDI

Johansen Co-integration test: Sample (adjusted): 1972-2014, Included obs.: 43, Series: LnREC, LnFDI, Lags interval (in first differences):1 to 1.				
Unrestricted Co-integration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.187088	9.779805	15.49471	0.2981
At most 1	0.020100	0.873119	3.841466	0.3501
Unrestricted Co-integration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None*	0.187088	8.906686	14.26460	0.2940
At most 1	0.020100	0.873119	3.841466	0.3501

Source: Author`s own calculations

5.4 Granger Causality test

In the next step the causal relationship will be checked between LnCO2 and LnGDP.

The null hypothesizes of the test:

H_0: LnGDP does not Granger Cause LnCO2, and

H_0: LnCO2 does not Granger Cause LnGDP

Null hypothesis will be rejected if the probability value is less than 0.05%.

Table 7: The Granger Causality Test Results for LnCO2 and LnGDP

Pairwise Granger causality test, Lags 2, Sample 1970-2014		
Null Hypothesis	F-statistic	Prob.
LnGDP does not Granger Cause LnCO2	1.14038	0.3304
LnCO2 does not Granger Cause LnGDP	1.87068	0.1679

Source: Author`s own calculations

According to the obtained results, the null hypothesis of no causal relationship from LnGDP to LnCO2 should be accepted (P-value=0.3304>0.05). Also, the second null hypothesis of no causal relationship from LnCO2 to LnGDP should be accepted (P-vlue=0.1679>0.05). Thus, the results of the causality test demonstrated that there is no causal relationship from LnGDP to LnCO2 and vice versa (See Table 6).

The later step is checking the causal relationship between LnREC and LnGDP.

The null hypothesizes of the test:

H_0: LnGDP does not Granger Cause LnREC, and

H_0: LnREC does not Granger Cause LnGDP

Null hypothesis will be rejected if the probability value is less than 0.05%.

Table 8: The Granger Causality Test Results for LnREC and LnGDP

Pairwise Granger causality test, Lags 2, Sample 1970-2014		
Null Hypothesis	F-statistic	Prob.
LnGDP does not Granger Cause LnREC	2.69157	0.0807
LnREC does not Granger Cause LnGDP	0.20702	0.8139

Source: Author`s own calculations

According to the obtained results, the null hypothesis of no causal relationship from LnGDP to LnREC should be accepted (P-value=0.0807>0.05). Also, the second null hypothesis of no causal relationship from LnREC to LnGDP should be accepted (P-vlue=0.8139>0.05). Thus, the results of the causality test demonstrated that there is no causal relationship from LnGDP to LnREC and vice versa (See Table 8).

The next step is checking the causal relationship between LnCO2 and LnFDI.

The null hypothesizes of the test:

H_0: LnFDI does not Granger Cause LnCO2, and

H_0: LnCO2 does not Granger Cause LnFDI

Null hypothesis will be rejected if the probability value is less than 0.05%.

Table 9: The Granger Causality Test Results for LnCO2 and LnFDI

Pairwise Granger causality test, Lags 2, Sample 1970-2014		
Null Hypothesis	F-statistic	Prob.
LnFDI does not Granger Cause LnCO2	1.86025	0.1695
LnCO2 does not Granger Cause LnFDI	4.17188	0.0230

Source: Author`s own calculations

According to the obtained results, from Granger causality test, the null hypothesis of no causal relationship from FDI to CO2 should be accepted (P-value=0.9478>0.05). But based on P-value= 0.0230<0.05%, the second null hypothesis of no causal relationship from CO2 to FDI should be rejected. Thus, the results of the causality test demonstrated the unidirectional causal relationship from CO2 to FDI (See Table 9).

The next step is checking the causal relationship between LnREC and LnFDI.

The null hypothesizes of the test:

H_0: LnFDI does not Granger Cause LnREC, and

H_0: LnREC does not Granger Cause LnFDI

Null hypothesis will be rejected if the probability value is less than 0.05%.

Table 10: The Granger Causality Test Results for LnREC and LnFDI

Pairwise Granger causality test, Lags 2, Sample 1970-2014		
Null Hypothesis	F-statistic	Prob.
LnFDI does not Granger Cause LnREC	1.01164	0.3732
LnREC does not Granger Cause LnFDI	1.61336	0.2126

Source: Author`s own calculations

According to the obtained results, the null hypothesis of no causal relationship from LnFDI to LnREC should be accepted (P-value=0.3732>0.05). Also, the second null hypothesis of no causal relationship from LnREC to LnGDP should be accepted (P-vlue=0.2126>0.05). Thus, the results of the causality test demonstrated that there is no causal relationship from LnGDP to LnREC and vice versa (See Table 10).

6. Conclusion

According to findings from empirical analysis the results can be summarized as follows. The Johansen co-integration and Causality test results indicate that, there was not co-integration and causality between GDP per capita and CO2 emissions; also there was not co-integration and causality between GDP per capita and REC. As we know from above mentioned information the EKC theory says that environmental degradation increases with increasing of GDP per capita and when nation become richer (they require healthier life), the government applies strict environmental policies and utilize advanced environmentally technologies to decrease environmental degradation. But empirical results of this paper prove that the EKC theory doesn`t exist in Turkish economy. In another hand the empirical results based on the Johansen and Causality test results shows that there was co-integration and causality between FDI and CO2 emissions; also there was not co-integration and causality between FDI and REC. Based on this results FDI which mostly comes from developed countries (United Kingdom (11.9%), Netherlands (11.6%), United States of America (9.3%), Spain (6.6%), Germany (6.5%), Austria (6.1%), Japan (2%), Switzerland (1.7%), China (1.26%), Others (44.31%)("FDI in Turkey - Invest in Turkey," n.d.)) applies heavy-polluted industries into Turkish economy. These results indicate the existence of PHH in Turkish economy. Which means, because of stringency of environmental policies the developed countries shift their capital intensive dirty industries to Turkey as a developing country.

References

[1] Aslanidis, N., & Iranzo, S. (2009). Environment and development: is there a Kuznets curve for CO2 emissions? Applied Economics, 41(6), 803–810.

[2] Balibey, M. (2015). Relationships among CO2 emissions, economic growth and foreign direct investment and the EKC hypothesis in Turkey. International Journal of Energy Economics and Policy, 5(4), 1042–1049.

[3] Cole, M. A. (2004). Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. Ecological Economics, 48(1), 71–81.

[4] Copeland, B. R., & Taylor, M. S. (1994). North-South Trade and the Environment. The Quarterly Journal of Economics, 109(3), 755–787. https://doi.org/10.2307/2118421

[5] FDI in Turkey - Invest in Turkey. (n.d.). Retrieved April 2, 2019, from http://www.invest.gov.tr/en-US/investmentguide/investorsguide/Pages/FDIinTurkey.aspx

[6] Gill, F. L., Viswanathan, K. K., & Karim, M. Z. A. (2018). The Critical Review of the Pollution Haven Hypothesis (PHH). International Journal of Energy Economics and Policy, 8(1), 167–174.

[7] Grossman, G., & Krueger, A. (1991). Environmental Impacts of a North American Free Trade Agreement (No. w3914). https://doi.org/10.3386/w3914

[8] Grossman, G. M., & Krueger, A. (1995). Economic Growth and the Environment. The Quarterly Journal of Economics, 110(2), 353–377.

[9] Hitam, M. B., & Borhan, H. B. (2012). FDI, growth and the environment: impact on quality of life in Malaysia. Procedia-Social and Behavioral Sciences, 50, 333–342.

[10] Jbara, B. W. (n.d.). Exploring the Causality between the Pollution Haven Hypothesis and the Environmental Kuznets Curve. 21.

[11] Kerekes, S., Marjainé Szerényi, Z., & Kocsis, T. (2018). Sustainability, environmental economics, welfare. Corvinus University of Budapest.

[12] Khandker, L. L., Amin, S. B., & Khan, F. (2018). Renewable Energy Consumption and Foreign Direct Investment: Reports from Bangladesh. Journal of Accounting, 8(3).

[13] Leitão, N. C. (2014). Economic growth, carbon dioxide emissions, renewable energy and globalization. International Journal of Energy Economics and Policy, 4(3), 391–399.

[14] Levinson, A., & Taylor, M. S. (2008). Unmasking the pollution haven effect. International Economic Review, 49(1), 223–254.

[15] Linh, D. H., & Lin, S. (2014). CO2 emissions, energy consumption, economic growth and FDI in Vietnam. Managing Global Transitions, 12(3), 219–232.

[16] Marton, C., & Hagert, M. (2017). The Effects of FDI on Renewable Energy Consumption.

[17] Mert, M., & Bölük, G. (2016). Do foreign direct investment and renewable energy consumption affect the CO 2 emissions? New evidence from a panel ARDL approach to Kyoto Annex countries. Environmental Science and Pollution Research, 23(21), 21669–21681.

[18] Murthy, K. B., & Bhasin, N. (2016). Environmental Kuznets Curve: CO2 emissions, pollution havens and type of economic development. Ch. X, in Ed. John R McIntyre et Al. Emerging Dynamics of Sustainability in Multinational Enterprises, Edward Elgar, Cheltenham, 209–231.

[19] Murthy, K. V., & Gambhir, S. (2018). Analyzing Environmental Kuznets Curve and Pollution Haven Hypothesis in India in the Context of Domestic and Global Policy Change. Australasian Accounting, Business and Finance Journal, 12(2), 134–156.

[20] Naz, S., Sultan, R., Zaman, K., Aldakhil, A. M., Nassani, A. A., & Abro, M. M. Q. (2019). Moderating and mediating role of renewable energy consumption, FDI inflows, and economic growth on carbon dioxide emissions: evidence from robust least square estimator. Environmental Science and Pollution Research, 26(3), 2806–2819.

[21] Panayotou, T. (n.d.). Economic Growth and the Environment. 49.

[22] Pazienza, P. (2015). The relationship between CO2 and Foreign Direct Investment in the agriculture and fishing sector of OECD countries: Evidence and policy considerations. Intellectual Economics, 9(1), 55–66.

[23] Shafik, N., & Bandyopadhyay, S. (1992). Economic Growth and Environmental Quality: Time-series and Cross-country Evidence. World Bank Publications.

[24] UN Climate Change Annual Report 2017 Website Now Live | UNFCCC. (n.d.). Retrieved April 1, 2019, from https://unfccc.int/news/un-climate-change-annual-report-2017-website-now-live

Teaching Marketing, Management and Governance in the 21st Century - Learning by Participating

Christian Zhelev

PhD, University for National and World Economy, Bulgaria

Abstract

21st century requires innovative teaching approaches, especially when it comes down to university students. In this paper it is emphasized on the importance of the applying the method "learning by participating" in order to engage the audience and improve students' performance. The focus is on the real outcome of the educational process, not only on the replicating knowledge, but also on applying it. An investigation and comparison between different approaches is provided and on that base conclusions are driven out. In the paper is also argued that this approach is especially useful when it comes down to students in Marketing and Management fields of studies.

Keywords: Marketing, Management, teaching university students, education

1. Introduction

The current paper is a result of a participation of the author in an international project[1]. Some of the data are driven from the intellectual outputs of the project, while other data could be considered as a continuation of the project findings. The main objective of the paper is to present techniques engaging student's attention and thus enhancing their performance in and out of the university hall. In the paper, it is argued that the "learning by participating" approach is rather useful and by its application among university students could enhance their performance in and out of the university hall. For the need of the paper, a research was done to establish the relevance of the proposed teaching method. The research has proven the relevance, once more – the anonymous feedback, gather from the students at the end of the educational process proved that it could be successfully implemented in various university disciplines.

The basic methodology approaches could be summarized as follows: the literature review, analyses, synthesis, observation and experimental approach.

What is more, the paper examines the traditional teaching techniques and the contemporary approaches in educating unreality students and proposes a blend between them to be used in order to boost student's performance, enhance the level of literacy, create better student – professor interaction and prepare students for real business environment.

2. Traditional and contemporary teaching techniques.

Teaching styles have changed significantly over the years. The traditional way that education was delivered was through recitation and memorisation techniques, whereas the modern way of doing things involves interactive methods[2]. It also worths mentioning, that the understanding of the author of the current paper about what teaching is that the ultimate goal of teaching a certain subject/ discipline is the high achievements of the students at the end of the course and successful application of the acquired knowledge after graduation. Countless number of techniques have been experimented, approved, rejected, and praised over the years.

Prosser and Trigwell (1999) assert that teaching is about creating contexts that make learning possible. Professional institutions and the government currently seek to further similar aspects of the teaching process. Similarly, higher education institutions have emphasized in their strategies that a university education can enhance learners' lives. (Brunel University, 2001)

2.1 Traditional teaching techniques

[1] http://www.inkams.eu/

[2] https://www.stephenperse.com/blog/?pid=5&nid=45&storyid=4728

The traditional teaching techniques work to a significant extend, but it the highly competitive environment today a broader understanding of the complex approaches is needed, this is why the traditional techniques (which in this paper will be called learning by repeating and replicating) need to be implemented in the educational process, but they could not possibly be considered.

Within the next paragraph, so examples will be provided. For example, the lecturer has decided to define marketing during the first lecture, the AMA definition is considered sufficient for the beginning of educational process. In its latest revision it sounds like this: Marketing is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large.(AMA, 2017). Normally students are expected to understand what lies behind this complex definition. However, the lectures should know that few of them truly realize the link between the different element of the marketing mix and creating, delivering, exchanging, communicating.

Another example that could be provided here is within the corporate governance filed. According to OECD "Procedures and processes according to which an organisation is directed and controlled. The corporate governance structure specifies the distribution of rights and responsibilities among the different participants in the organisation – such as the board, managers, shareholders and other stakeholders – and lays down the rules and procedures for decision-making"[1]. Master students, experienced as they are in the university education, often times face difficulties understanding this definition. Further thorough explanation is needed.

Naturally, throughout the course, the lecturer would return to this basic knowledge, but often it stays not fully understood and if the students are asked to implement the gained knowledge in more complex interdisciplinary relations, they fail to do so. That is one of the many various reasons why more and more innovative teaching techniques are constantly being developed. In addition, as the business environment develops, the demand for better-qualified employees grows. By teaching students only how to "repeat" what has been learned during the semesters slowly, but certainly is turning out to be irrelevant.

Erikson (Erikson 2011) argues that this focus misses (the doctrine[2]) the real goal of education. At the end of the day, professors do not want simply to cover the material. Instead, they want to ensure that students are learning the material. If they evaluate themselves honestly, most professors would probably admit that their students are not learning as much as the professors had hoped. Professors often talk with disappointment about their students' exam answers or contributions in class. Yet, they may not be sure how to improve this situation, so they just vow to teach the doctrine more thoroughly the next time around. This strategy never answers the crucial question—if we are sure that, we taught the material, why do the students not seem to be learning it? According to the cited author (Erikson 2011), whose research lies in the field of educating law students, student who has memorized the rules but who cannot apply them in these ways does not know the law in any satisfactory way. Students do not acquire this deep understanding of the law through passive methods of instruction. This conclusion could be implemented in the field of economic studies as well. Dhillon (2009) states that business education must provide students with the communication, interpersonal, and intellectual skills that prepare them for a better understanding of the broad picture of business in today's global environment.

All of the abovementioned predetermines a shift in the understanding of the different teaching methodology. The better understanding of the students' needs lies in the basics of the provoking interest. In search of the students' engagement and influenced by the dynamic environment the lecturers should be able to implement new ways of presenting the ideas, they want to convey in the lecture hall.

2.2. Learning by doing and why it is necessary.

Gerald (Gerald, et.al, 20017) argues that the problem of students negative attitude towards learning, low academic performance and low level of acquisition of the 21st century skills, the researcher believes that by integrating contemporary teaching strategies students' attitude towards learning, academic performance and acquisition of the 21st century skills will develop. Specifically, these needed skills in today's generation is necessary to become globally competitive students. As educators, we are bound to answer these problems in order for us to cater the increasing needs of our students.

Nowadays among the academia the understanding that the case study approach is one of the bets ways of educating is widely spread. It is also considered that enhances performance and achievement. However, with the low concertation levels

[1] https://stats.oecd.org/glossary/detail.asp?ID=6778

[2] Erikson addresses traditional methods in Law Education and contrasts them to the case study approach.

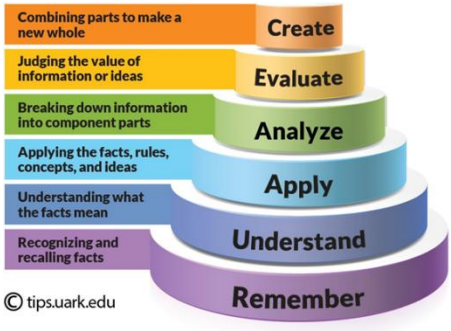
if the millennials (and even lower are expected among the digital natives) the task of the lecturer is getting even more challenging. That is why a thorough understanding of the complex process of educating the new generations is needed.

2.3. Bloom's Taxonomy and the Holistic Approach to Learning and Teaching Interaction

2.3.1. Bloom's Taxonomy

The need of a thorough understanding of the complex process of educating could be easily visualized by the Bloom's taxonomy.

Figure 1[1]



Source: https://tips.uark.edu/using-blooms-taxonomy/

Bloom's taxonomy gives the logic of the three groups of teaching approaches, as proposed in the current paper. It could be summarized that the traditional teaching techniques focus on the first two levels - remembering and understanding, while the text two levels applying and analysing reflects the better part of the learning by doing approach, and partially focuses on the evaluation process. While the learning by participating approach call for action in the university hall. Namely, the process of creation of intellectual output could enhance the student's positive experience, provide him/her with valuable insights and at the same time fill in knowledge potholes.

2.3.2. The holistic approach

The holistic approach seeks to develop people. The holistic learning and teaching approach leads to a stimulating learning environment, which influences and inspires critical learners. It provides learners with ownership of knowledge and the knowledge creation process by engaging with them. Ownership is vital in the learning processes because it provides confidence and a sense of participation in the knowledge community, leading to independence. The approach engages learners and leads to them generating peer discussion. It also makes critical learners aware of the knowledge creation process. This awareness is important because it is the first step in leading learners to question knowledge claims. The holistic approach is effective in getting students to check their own understanding of knowledge and levels of learning achieved (Patel, 2003).

The holistic approach develops students to be critical, confident and independent. It aims to make learning a process of self-improvement that explicitly recognizes the self and the social context of learning and teaching, and recognizing the needs of the individual learner in the interaction. Its premise is that the social context of the interaction is significant. It recognizes that the exchanges that take place within this social action are the foundation for developing critical learners, thus including experiential knowledge of learners and teachers to improve the quality of the teaching situation and levels of achievement of learner (Patel, 2003).

When the lecturer is armed with better understanding of the Bloom's taxonomy and the holistic approach and aware of all the pros and cons of the them, it is easier for the lecturer to flip the university hall from traditional lecturing to student-cantered approach.

[1] credit for Jessica Shabatura

3. "Learning by participating" in action or about the intellectual output of the INKAMS project and the survey results of the current paper[1]

As mentioned above, in the changing environment a blend of teaching approaches and techniques are required. Using a mix of instructional methods (lectures, discussions, video/film analysis, and tests), students are presented with specific techniques, processes, and concepts in both passive-and active learning environments.

3.1. Movie education and the INKAMS Project[2]

It the current paper, a certain intellectual output of the project is used: storytelling and move education.

The idea, communicated throughout the project, is that the students, who are involved in active learning and participating in the educational process should perform better and be more prepared for their future career development. This will boost their academic results and will help them in the process of skills development. Once the students have gained sound, theoretical preparation they are expected to produce short movies (by the mean of storytelling)[3] based on a certain topic from the curriculum. The task, simple as it may sound, proved to be quite challenging, but at the same time highly engaging and skills creating. According to the definition of the participant in the project movie education allows, thanks to narrative and a strong contextualization of personal, professional and emotional experiences and events, the combination of entertainment and learning, allowing to overcome the traditional educational model, in its classic configurations, where the learner is a passive subject called to make use of "given" (top-down) educational contents, without favouring their memory over time.

3.2. About the verification of the result. Approbation among students

For the needs of the current paper experiment as a method for testing was chosen. According to Berman and Evans (2009) we have Experiment when one or more factors are manipulated under controlled conditions.

The move education approach was offered to preliminary selected group of students 12 students (form a cohort of 180), leaning marketing in their second year of university education. Two students dropped out at the beginning of the experiment. Traditionally, for several years, the course is organised as follows: Course project (learning by doing) is offered to the students as a motivation for accelerating the hands on approach. The students sit two tests – Midterm test and Final test. Over the years, difference could be observed in the results (based on observations) but the trend was positively stable. This year (2019), the movie education was implemented as a contemporary teaching approach among the selected students. The enrolment of the students took place after the Midterm test. Within a month of work, the students managed to deal with the tasks of the movie education and by doing so to enhance their performance. The students in the selected group were asked to provide the lecturer with a constant feedback and the feedback was strictly positive. The results of the experiment are best shown in the table below.

Table 1

№	% change of the score	№	% change of the score	№	% change of the score	№	% change of the score
135	107,89%	26	6,10%	4	-5,41%	177	-18,95%
147	72,31%	120	5,00%	116	-5,83%	51	-18,97%

[1] Within this part of the paper some of the conclusions are taken directly from the intellectual outputs of the INKAMS project. The author, being part of the project, does not, bu any circumstances, claim that the results form Outputs are his only. On the contrary, I do declare, that they are collective work of all the partners, involved in the project.

[2] International Key Account Management & Sales this Erasmus+ project aims to realize, pilot, disseminate and systematize a new University-based learning program focused on International Sales & Key Account Management

[3] It is not simply the art of telling stories, but a discipline, a tool to convince, a socio-professional device.

9	68,00%	123	4,71%	131	-6,06%	13	-18,99%
29	61,76%	113	4,11%	19	-6,33%	87	-19,09%
52	55,36%	104	3,66%	71	-6,41%	72	-19,61%
144	55,10%	30	3,57%	7	-6,82%	89	-19,63%
105	52,63%	75	2,74%	139	-6,90%	12	-19,78%
47	50,00%	74	2,63%	149	-7,25%	69	-21,05%
43	48,78%	56	2,11%	38	-7,41%	78	-21,35%
98	42,19%	65	1,89%	16	-8,00%	122	-21,69%
150	38,89%	160	1,56%	81	-8,22%	39	-23,08%
134	35,94%	40	1,39%	42	-8,42%	175	-23,08%
84	33,33%	14	1,20%	157	-8,70%	103	-24,19%
18	30,14%	61	1,08%	152	-9,52%	32	-25,29%
174	25,49%	55	1,05%	50	-9,78%	88	-25,93%
93	25,00%	76	0,99%	63	-9,86%	114	-26,32%
67	21,62%	25	0,96%	110	-10,77%	130	-26,32%
126	21,54%	68	0,88%	3	-11,67%	28	-27,27%
159	20,97%	22	0,00%	91	-11,84%	112	-28,57%
24	19,75%	46	0,00%	128	-12,24%	161	-28,57%
44	19,40%	171	-1,03%	60	-12,63%	97	-28,92%
31	17,86%	118	-1,10%	102	-12,66%	145	-31,25%
137	17,65%	141	-1,27%	111	-12,70%	124	-31,40%
153	17,57%	49	-1,67%	154	-14,58%	143	-36,84%
34	17,50%	166	-1,72%	36	-15,09%	37	-42,11%
125	17,14%	59	-1,83%	117	-15,25%	90	-43,08%
11	14,46%	129	-1,96%	58	-15,48%	162	-43,21%
115	12,50%	132	-2,20%	15	-16,22%	95	-44,04%
127	11,48%	92	-2,70%	79	-16,67%	17	-46,81%
64	10,84%	96	-2,83%	100	-17,24%	173	-47,92%
77	10,53%	136	-2,99%	66	-17,43%	155	-51,95%
70	10,42%	94	-3,57%	21	-17,58%	169	-52,54%
33	8,82%	101	-3,70%	99	-17,91%	142	-53,47%
148	7,35%	27	-3,85%	54	-18,33%	163	-54,39%
121	6,98%	146	-3,98%	62	-18,33%	86	-58,49%
73	6,15%	53	-4,35%	108	-18,46%	6	-69,05%

Source: Author's data calculations, based on real students' performance

The percent change of the performance is significant among the students, who participated in the experiment: the yellow cells in the table. The clear conclusion that could be driven once more confirmed the relevance of the proposed teaching method, called in this paper "**learning by participating**".

The conclusions reached by the participants in the INKAMS project are once more validated. Namely, that though the approach of video creating (move education) / learning by participating the students not only in sales, but also in many other areas of economic education could benefit:

- Connecting knowledge and ideas explicitly to "real world" examples
- Acquiring broader skills in storyboarding, narration and digital technologies
- Enchanting their thinking, systematizing, creativity
- Integrating film into assessment offers a "eureka" moment in the student learning experience.

4. Conclusion

Teaching in the 21st century is challenging task, which should follow student centric approach. Though creating a flipped classroom, where the student is the centre. The overall conclusion is that neither of the teaching approaches work isolated. With all theirs pros and cons, a blend of teaching techiques is needed in the contemporary university hall. Yet again, after a sound preparation students in the 21st century should be able to apply the action gained throughout the educational process. The movie education, as a contemporary technique arms the students with the confidence of skills gained and leaves them with confidence for real life application of the knowledge they have received. Of course, further investigation is needed to fully confirm and extend the outcome of this research.

References

[1] Best, R.J. (2016). Market – Based Management. Pearson Education
[2] Evans, Joel R. Berman, B., (2009) Marketing in the 21st Century, Online Edition
[3] Erickson, J. (2013). Experiential Education in the Lecture Hall. Northeastern University Law Journal, 87-110.
[4] Gerald, U. P., & Gladys, S. E. (2017). Contemporary Teaching Strategies on Students' Attitude, Academic Performance and Acquisition of the 21 st Century Skills. International Journal of Scientific & Technology Research Volume 6, 322-337.
[5] Inés, B. K., & Natalia, L. V. (2007). Exploring TeachingMethods in Marketing.
[6] Juan-Antonio, M.-J. e. (2010). Case Studies as Practical Teaching in the New Marketing Courses. American Journal of Business Education, 33-38.
[7] Leppiman, A. (2009). Experience marketing from a teaching and learning perspective. Journal of Marketing Trends.
[8] Olivares, S. (2018). Business Graduate Skills: Competency-Based Model. IGI Global.
[9] Patel, N. V. (2017). A Holistic Approach to Learning and Teaching Interaction: The International Journal of Educational Management, 272-284.
[10] Renee, N. A. (2017). Contemporary Teaching Strategies:Effectively Engaging Millennials Across Effectively Engaging Millennials Across.American Customer Satisfaction Index, 2019
[11] https://www.theacsi.org/about-acsi/key-acsi-findings
[12] http://didit-extensive-reading.blogspot.com/2017/01/title-and-identity-holistic-approach-to.html
[13] https://www.stephenperse.com/blog/?pid=5&nid=45&storyid=4728
[14] http://www.inkams.eu/movie-education-material/
[15] http://www.inkams.eu/
[16] https://www.ama.org/the-definition-of-marketing-what-is-marketing/
[17] https://stats.oecd.org/glossary/detail.asp?ID=6778

Corporate Governance and Marketing Management Issues and Their Reflection on Customer Satisfaction

Christian Zhelev

PhD, University for National and World Economy, Bulgaria

Abstract

The current paper aims at shedding light on the interaction between corporate governance and marketing management. The idea of sustainable governance is widely explored nowadays by academia and business, yet again the interdisciplinary interactions are rarely the focus. Here an attempt to provide evidence that customer satisfaction is related to good corporate governance is made. The research is based on literature review and case study approach. The final output of the research offers some evidence on the matter and also leaves a wide field for further discussion.

Keywords: corporate governance, marketing, customer satisfaction, sustainable governance

1. Introduction

Business operations in the 21st century are becoming more and more sophisticated. At the same time, customers demand more and more transparency on how the goods and services are produces. They also demand the customer service and care to be on a rather high level. Due to the digitalization of the processes companies face significant challenges, which require fast and immediate reaction to the dynamic environment. All these trends and many other lead to the necessity of integrated approach, when researches are carried out. Interdisciplinary approach is needed in order to better understand the complex processes which are part of the business environment. These and other motives have predetermined the focus of this research. In the current article it is argued that there is interaction between the good corporate governance and marketing management activities in general and between the customer satisfaction in particular.

The basic methodology is the literature review, analyzes, synthesis and case study approach. The paper presents two branches of the contemporary science, which on first singth are rather different – corporate governance and marketing. The literature review showed that the research paper in the so determined field are few. Based on that, it could be said that the field in not widely explored by the academia and the humble aim of this paper is to shed some light, but mainly to serve as foundation for further discussions and research.

2. Corporate governance – contemporary issues and sustainable knowledge

This idea of Corporate Governance[1] is widely accepted by academia and business today. It has been also implemented in the officially adopted definition of OECD in the light of modern understanding of Corporate Governance. Here it is visible that we can clearly trace the evolution of CG from a system by which companies are directed and controlled to an entire system of rules and requirements. Even though, corporate governance differs within different companies and countries, its main objective remains the same: to achieve high level of performance, profitability and to prevent the management from pursuing their own objectives at the cost of the stakeholders. Similar ideas are also shared by Boeva (Boeva, 2019) in the book Capital, melting, glaciers and 2C. In the book, Boeva widely explains the need of enlarging the scope of corporate governance; it is true that the focus of the book is sustainable governance, but the focus on all stakeholders, including customers cannot go unnoticed. It must be noted that poor corporate governance or lack of compliance to its principles

[1]The Priciples of Corporate governance, OECD/G20 reflect fully the understanding of the author about the scope of Corporate Governance. That is why main moments of theses principles serve as a foundation of the paper, without further explicit explanation. The author is aware of the main field of application of Corporate governance – public listed companies, but these, as many research papers have confirmed, the principles could be used by all types of companies.

could most likely lead to corporate abuses, frauds and poor performance of the companies, which in turn leads to poor offerings and diminishing customer satisfaction.

From a theoretical perspective, corporate governance is based on comprehensive principles that are applied as an effective and efficient framework planned to support corporations' profitability, performance, productivity, customers' satisfaction and public interest (Fawal and Mawlawi, 2018). Further more, it should be noted that the governance bodies are those who give the strategic direction of the companies.

But, at the same time, according to some authors, business strategy is a strategy that focuses on improving the competitive position of a company or a business unit (product or service) within a specific industry. The strategy and the strategic decisions have a significant effect on the overall performance of the company (Wheelen et. al., 2015).

Other well known theory is the one of Jason and Mackling. According to Jensen and Meckling (1976), the firm is not an individual but a legal entity, where conflicting objectives of individuals are brought into equilibrium within a framework of contractual relationships. These contractual relationships are not only with employees, but with customers, suppliers, and creditors. The purpose of these contracts is that all the parties acting in their self-interest are motivated to maximize the value of the organization, reducing the agency costs and adopting accounting methods that most proficiently reflect their own performance (Deegan, 2004). The development of this theory finds its reflection in the firm as a nexus of contracts and Resource dependency theory. The nexus of contracts theory depicts the firm in a web of implicit and explicit contracts with stakeholders; however, the shareholder has predominance over other stakeholders. The board and management continue to have a fiduciary responsibility to maximize shareholder value (Badu, 2012). The firm as a nexus of contracts describes the firm in relation to its environment—a view of the firm looking outward into its environment. Resource dependency theory provides a similar view. It describes power/dependence relationships between the firm and other actors in its environment and enables primary attention to those stakeholders in the environment who have the greatest impact on the profitability of the firm. These theoretical foundations, together with the Principal – agent theory and Stewardship Theory have served as a background for the establishment of the Principles of Corporate governance (OECD, 2015)

Within the next line some of the principles are presented. The idea is to better enlighten on the interaction, searched in the current paper.

- The board is not only accountable to the company and its shareholders but also has a duty to act in their best interests. In addition, boards are expected to take due regard of, and deal fairly with, other stakeholder interests including those of employees, creditors, customers, suppliers and local communities (OECD, 2015).
- The competitiveness and ultimate success of a corporation is the result of teamwork that embodies contributions from a range of different resource providers including investors, employees, creditors, customers and suppliers, and other stakeholders. Corporations should recognise that the contributions of stakeholders constitute a valuable resource for building competitive and profitable companies. It is, therefore, in the long-term interest of corporations to foster wealth-creating co-operation among stakeholders (OECD, 2015).

It is evident from the above-mentioned moments from the principles that the scope of the contemporary Corporate Governance extends far beyond the board and the shareholders (Donalds, 1995). The interdisciplinary interaction between corporate governance and marketing is obvious.

As it was mentioned above, few research papers deal with the matter. Yet again, such papers exist.

Wen-Min, et.al. (2012) point out that their study explores the relationship between operating performance and corporate governance in 30 airline companies operating in the US. First, this study applies a two-stage Data Envelopment Analysis (DEA) to evaluate the production efficiency and marketing efficiency of the airlines. Their findings indicate that, in general, there is not as much dispersion in the relative productive efficiencies of the airlines as there is in their marketing efficiencies. The low-cost airlines, on average, are more efficient carriers than the full-service ones, but less efficient marketers. Secondly, truncated regression is used to explore whether the characteristics of corporate governance affect airline performance. The results demonstrate that corporate governance influences firm performance significantly. Finally, we address the managerial decision-making matrix and make suggestions to help airline managers improve performance (Wen-Min et. al., 2012). The research reveals a specific relationship between the examined disciplines – corporate governance and marketing management.

So far it became clear that the relationship between governance and marketing is present. What is more, it is a relationship felicitated by the management of the company (Charti, 2009). Evidance of the existence of this relationship has already been established. Zhelev (2015) states that: in the search for relationships and interactions, these have been found to be two-way - on the one hand, corporate executives influence management and on the other, managers can influence corporate governance at the corporate leadership level. This interaction is important for the international business. The lack of good interaction influences the making of informed strategic decisions, which in turn leads directly to the long-term lack of development of the company.

Another angle of evidence for the interaction and relationship between the examined fields is concept of Product Market Competition. Product market competition is a significant external mechanism in as it affects management incentives, and thereby the efficiency of the firm (Donaldson, Davis, 1991). Table 1 defines the effects of product market competition on company's performance and other mechanisms in corporate governance. Even in a weak governance environment, strong product market competition can act to align managers' goals to the aim of efficient productivity (Shaba et.al., 2016). Competition encourages management to operate more efficiently and reduce costs in order to avoid bankruptcy. Evidence on the influence of product market competition on performance however, appears inconclusive (Gebhardt et. al., 2006). Strong product market competition may create other issues. Competitive pressure may worsen the moral hazard problems in the US savings and loans (S&L) industry. In order to survive in an environment of fierce competition, It has been discussed competitive pressure can lead to a variety of unethical practices, e.g., child labour, corruption, excessive executive pay, and earnings manipulation, which is a violation of the principles of governance and of marketing management.

Table 1. Impact of Product Market Competition on Company's Performance and Markeitng Management

From Product market competition to corporate performance	*Strong product market competition forces managers to focus on better financial performance, as bankruptcy and redundancy may be the ultimate result.*
From Product market competition to ownership structure	*Financial institutions that compete intensely in their own product markets may have stronger incentives to demand better financial performance on their own equity investments, than a financial owner that does not face strong competition in their own product markets.*
From Product market competition to incentive based compensation	*Product market competition heightens the incentive effect of the remuneration system, as it allows remuneration to correlate relative to the performance of close competitors rather than relative to the market.*
From Product market competition to performance monitoring system	*Higher product market competition makes it easier to measure the performance of a firm as close competitors act as a benchmark. It is also easier to ascertain how much of a firm's profit is determined by monopoly pricing. In order to find the social efficiency of the firm, this must be subtracted from net profit.*

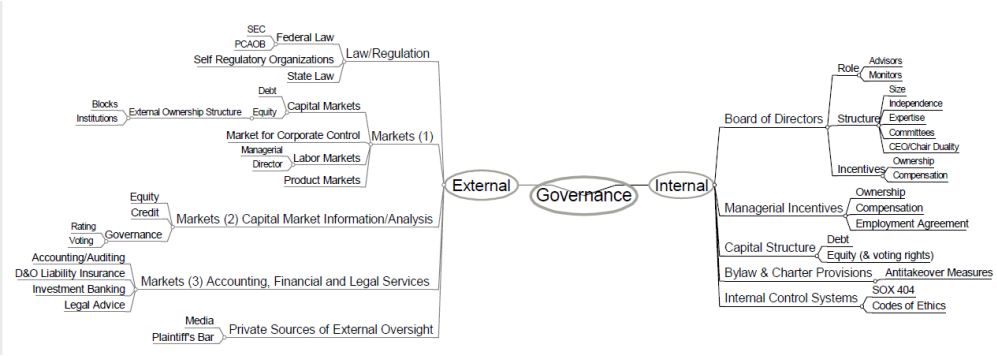
Source: Table adapted from "Encyclopaedia of Corporate Governance"

Shleifer argues that if a company treats honesty as a normal good, then the demand for honesty will be lower in more competitive environments, since competition reduces operating profits. If a firm uses immoral or illegal ways to gain competitive advantage, then competition may not lead to socially desirable outcomes.

These inconsistent conclusions may be partially due to the research ignoring the interaction between competition and the enterprise institution. This means that firms are more productive in a competitive environment. Instead, if ownership is highly concentrated, competition and governance become substitutes, meaning that firms become less productive in a competitive environment.

The final evidence about the interaction, from theoretical point of view, could be traced in the figure below. Here we can explicitly see that there is a relationship between Corporate governance and Marketing Management.

Figure 1: Corporate Governance framework.

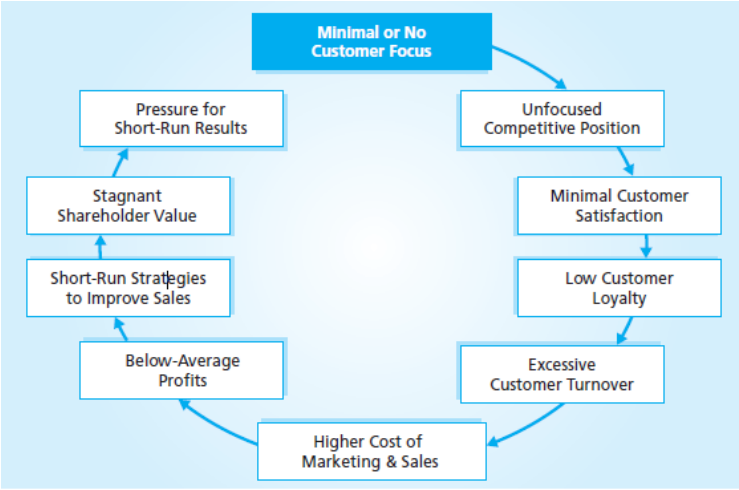


Source: Gillan S. L., 2005. Recent Developments in Corporate Governance: An Overview

3. Marketing Metrics as a part of the body of knowledge of Marketing Management.

In this part the marketing perspective of the problems is examined. The idea is even more explicitly to show the relationship between Marketing Management and Corporate Governance. For that case a certain point of view is presented – the one of Roger Best.

Figure 2. Customers and Shareholders Interactions



Soucre: Best, R.J., 2016, Market – Based Management, p.7

[1]A business's effectiveness in developing market-based strategies that deliver high levels of customer satisfaction depends on the strength of its customer focus. A few years ago, many managers considered that statement to be a nice academic philosophy that nonetheless did not have any practical application for a company trying to improve its profits. Today, however, the evidence shows that businesses with a strong customer focus deliver higher levels of customer satisfaction and above-average profits. Developing a customer focus requires customer leadership, which is evident in senior management leadership, employee customer training, and customer involvement. Letting the voice of the customer be heard is also a necessary part of customer focus; it can be achieved by striving to understand the customer experience, encouraging customer complaints, and seeking to develop customer solutions, not just products. Customer metrics, such as customer satisfaction, customer retention, and customer loyalty, are key performance metrics that measure the success of a customer-focused organization. The customer satisfaction metric is related to profitability. A high overall customer

[1] The next paragraph follows the logic of Best, 2016. Market –based management.

satisfaction score is nice, but unbundling the score into different levels of customer satisfaction allows a business to better understand its overall average and the profit impact of different levels of customer satisfaction. Efforts to better understand customer dissatisfaction can be greatly enhanced when a business encourages dissatisfied customers to complain. Allowing them to share their complaints provides both an opportunity to understand causes of customer dissatisfaction and a way to retain customers who otherwise would be lost. Various social media platforms have provided new forums for customers to complain and an opportunity for companies to respond with efforts to repair damaged customer relationships.

Customer satisfaction affects customer retention. And higher levels of customer retention deliver higher customer profits—not only annual profits but future profits as well, due to a longer customer life. It is much more expensive to acquire a new customer than to retain a current customer. Improving customer retention from 67 to 75 percent increases the customer life from 3 to 4 years. This adds one more year of sales and profits and spares the company the high acquisition cost of finding a new customer to replace each lost customer. In addition, customers generally buy more from a business the longer they stay with the business. The ultimate goal of a customer-focused business is a high level of customer loyalty. A loyal customer has a long purchase history, an above-average purchase amount, a high desire to repurchase, and a strong product preference. A loyal customer would recommend the business's products to others. Loyal customers are highly profitable and have an above-average customer lifetime value. Repeat customers could have an above-average lifetime value and long customer history but are less committed to the business's product and less likely to recommend it to others. Nevertheless, they are very important to the business's profitability and need to be carefully managed with respect to customer retention. The goal is to move them to a higher level of customer loyalty. Many businesses have captive customers but do not know it. These customers may have a long customer history and average or better lifetime value. But they are not happy and would switch to a different company if they could. As one might guess, they also share their frustrations with others and are more likely to advise others not to buy from the company. New customers may be first-time customers or returning customers. A customer focused business knows the customer history and purchase behavior of the returning customers and should manage them accordingly (i.e., differently from the first-time customers). Both types of new customers offer a business the opportunity to move them up in customer loyalty over time and improve profits with a higher customer profit and customer lifetime value.

All the above-mentioned facts, stated by Best (2016) are focused on the customers mainly, but what cannot go unnoticed in the role of the management and governance, in turn. Because the paragraph above deals with the strategic decisions and the strategic decisions are taken and approved on the board meeting and are executed by the management of the company.

In order to prove once more the relationship a careful examination of the Customer customer satisfaction is needed. The most reliable instrument, widely recognized by business and academia is the American Customer Satisfaction Index.

4. The American customer satisfaction index. The case of Nestle and BMW[1].

The American Customer Satisfaction Index (ACSI) is a new type of market-based performance measure for firms, industries, economic sectors, and national economies (Fornell et.al, 1996). Highlights of the findings include that (1) customization is more important than reliability in determining customer satisfaction, (2) customer expectations play a greater role in sectors in which variance in production and consumption is relatively low, and (3) customer satisfaction is more quality-driven than value- or price-driven. The authors conclude with a discussion of the implications of ACSI for public policymakers, managers, consumers, and marketing in general (Fornell et.al, 1996). Research using this measure of customer satisfaction revealed that businesses scoring in the top 25 percent in customer satisfaction produced an average shareholder value three times greater than businesses in the bottom 25 percent (Best, 2016)

According to the official website of the index:

- Customer satisfaction is a leading indicator of company financial performance. Stocks of companies with high ACSI scores tend to do better than those of companies with low scores.
- Changes in customer satisfaction affect the general willingness of households to buy. As such, price-adjusted ACSI is a leading indicator of consumer spending growth and has accounted for more of the variation in future spending growth than any other single factor (https://www.theacsi.org/about-acsi/key-acsi-findings).

[1] The two companies were chosen based on preliminary determined criteria. After a though examination of the index, it was concluded that their cases reflect in the best way the link: strong companies, strong governance – satisfied customers.

Table 2. Comparison BMW/Nestle

Company Background BMW Group is a German multinational company which produces automobiles and motorcycles. The company was founded in 1916 as a manufacturer of aircraft engines, which it produced from 1917 until 1918 and again from 1933 to 1945. Automobiles are marketed under the brands BMW, Mini and Rolls-Royce, and motorcycles are marketed under the brand BMW Motorrad. In 2015, BMW was the world's twelfth-largest producer of motor vehicles, with 2,279,503 vehicles produced. BMW is headquartered in Munich and produces motor vehicles in Germany, Brazil, China, India, South Africa, the United Kingdom, the United States and Mexico	**Company Background** Netsle AG is a Swiss multinational food and drink processing conglomerate corporation headquartered in Vevey, Vaud, Switzerland. It is the largest food company in the world, measured by revenues and other metrics, since 2014. It ranked No. 64 on the Fortune Global 500 in 2017 and No. 33 on the 2016 edition of the Forbes Global 2000 list of largest public companies. Nestlé was formed in 1905 by the merger of the Anglo-Swiss Milk Company, established in 1866 by brothers George and Charles Page, and Farine Lactée Henri Nestlé, founded in 1866 by Henri Nestlé
Corporate Governance The BMW Group's corporate culture is characterized by clear responsibility, mutual respect, and trust. Lawful conduct, fair competition, and respect for human rights are integral to our business activities and an important condition for securing the long-term success of our company. Our primary goal is to avoid risks that could jeopardize the trust our customers, shareholders, business partners, and the general public place in the BMW Group. For this purpose, the BMW Group has established a Compliance Management System equipped with instruments and measures to help associates handle legal risks.	**Corporate Governance** Our Board of Directors is highly engaged and dedicated to creating long-term, sustainable value based on strong principles of governance and an appropriate tone from the top. Our corporate governance framework is carefully constructed, and continually evaluated and updated, to ensure that it promotes accountability and supports our strategy to foster long-term value and sustainable growth for the benefit of all shareholders. We further integrated our public reporting on our financial and non-financial performance by including the highlights from our Nestlé in society report in our Annual Review. We recognize that for our company to be successful over time and create sustainable value for shareholders, we must also create value for society. We do this through our more than 2000 brands that enhance quality of life and contribute to a healthier future.
Strategy and marketing aspects We offer the best customer experience. We have a unique customer understanding through our ongoing, direct dialogue with our customers. We focus solely on premium products and services. We anticipate customer needs and desires. And we bring these to life - quickly and precisely - in innovative and emotional offerings and experiences.	**Strategy and marketing aspects** Nestlé has many distinctive strengths that keep us at the top of our industry. Our people are our greatest strength. We have an attractive product portfolio in growing categories with leading market positions. We are a global company with deep local roots, which gives us a unique ability to understand local consumers and adapt fast to their preferences. We have powerful, valuable brands, which consumers trust. Our products reach more than 1 billion consumers every day across the world. We also have industry-leading R&D capabilities that support our Nutrition, Health and Wellness strategy and our innovation initiatives. As the 'Good Food, Good Life' company, we enhance quality of life and contribute to a healthier future. Winning with consumers is the source of our sustainable financial performance and our way to earning trust and maintain

	our market leadership. Based on a compelling Nutrition, Health and Wellness strategy, our company delivers sustainable value over the short term and the long term.

Source: the official websites of the companies

A survey on of the capital market reveals that actually the two companies have sustainable trend of share prices. A check of the website American Customer Satisfaction Index shows that the two companies, over the years, are among the leaders in their industries[1].

The presented cases clearly state the relationship between governance and marketing management. The other important conclusion is that there is interdependence between the governance and marketing which in this case could be summed us as follows: better governance, better companies, profitable and satisfied customers.

They show something more, the big multinational companies have already understood the need of complex approach towards governance and marketing. The academia is lagging in this field.

5. The case of Bulgaria /Instead of conclusion/

Bulgarian companies do not measure customer satisfaction, but this index in conjunction with other marketing performance metrics could be implemented in order to create better understandings/ insights not only of customers, but also of company's performance. As discussed in the current paper these metrics interact with the level of corporate governance of the company, as well. This interaction is directly correlated to the level of corporate governance, because, after all, the fundamental idea of corporate governance is not only how the companies are directed and controlled, but also to take care of all the stakeholders (internal and external). As it was shown, the interaction is present. Of course further investigation is needed to fully confirm and extend the field of this research paper.

References

[1] Babu, J., 2012. Role Of Corporate Governance In Strategic Management. KMM Institute of Post Graduate Studies.

[2] Best, R.J., 2016. Market – Based Management. Pearson Education

[3] Boeva B., 2019. Capital, melting, glaciers and 2C. Iztok-Zapad publishing company

[4] Charti B., 2009. The Case for Subsidiary Corporate Governance. RBC Financial Group.

[5] Deegan C., 2004. Environmental disclosures and share prices - a discussion about efforts to study this relationship. Accounting Forum, vol. 28, no. 1, pp. 87-9

[6] Donaldson T. 1995. The stakeholder theory of the corporation: concepts, evidence and implications. Academy of Management Review, 1995: 65-91.

[7] Donaldson L., Davis J., 1991. Stewardship Theory or Agency Theory:CEO Governance and Shareholder Returns. Australian journal of management

[8] Fawal A., Mawlawi A., 2018. The Corporate Governance and Its Effect on the Marketing Performance Strategy in the Lebanese Banking Sector. Management 2018, 8(3): 86-93. DOI: 10.5923/j.mm.20180803.03

[9] Fornell C., Johnson M., Anderson E., Cha J., Bryant B., 1996. The American Customer Satisfaction Index: Nature, Purpose, and The American Customer Satisfaction Index: Nature, Purpose, and Findings. Cornell University, School of Hotel Administration, The Scholarly Commons

[10] Gebhardt G., Carpenter G.S., Sherry J.F., 2006. Creating a Market Orientation: A Longitudinal, Multifirm, Grounded Analysis of Cultural Transformation. Journal of Marketing, October 2006: 37–55.

[11] Gillan S. L., 2005. Recent Developments in Corporate Governance: An Overview. Journal of Corporate Finance, Forthcoming. Available at SSRN: https://ssrn.com/abstract=855444

[1] https://www.theacsi.org/news-and-resources/customer-satisfaction-reports/reports-2019

[12] Jensen M., Meckling W., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, Vol. 3, Issue 4, October 1976, pp. 305-360

[13] Shaba Y., Yaaba B., Abubakar I., 2016. Corporate Governance and Market Performance: Empirical Evidence from Nigerian Banks. International Journal of Research in Management, Economics and Commerce, Vol. 06 Issue 12, December 2016, Page 9-20

[14] Shleifer, Andrei, and Robert W. Vishny. 1997. A Survey of CorporateGovernance. The Journal of Finance 52, no. 2: 737–783. Portico

[15] Wen-Min L., Wei-Kang W., Shiu-Wan H., En-Tzu L., 2012. The effects of corporate governance on airline performance: Production and marketing efficiency perspectives. Transportation Research Part E: Logistics and Transportation Review, Vol. 48, Issue 2, March 2012, pp. 529-544

[16] Wheelen T. L., Hunger J. D., Hoffman A. N., Bamford C. E., 2015. Strategic Management and Business Policy. Pearson Education.

[17] Zhelev Ch., 2015. Corporate governance and international management – some issues concerning board members and senior managers in the transnational public listed corporations. Horizons in The Development of Human Resources and Knowledge. Volume 1.

[18] OECD, Principles of Corporate Governance, OECD, Paris, 2015

[19] American Customer Satisfaction Index, 2019, https://www.theacsi.org/about-acsi/key-acsi-findings

[20] Nestle AG, 2020, https://www.nestle.com/

[21] BMW, 2020, https://www.bmwgroup.com/en.html

Requalification of Residential Space in Tirana - Methodologies and Intervention Strategies

Etleva Dobjani

Arch. PhD, University POLIS, Tirana, Albania

Abstract

The city of Tirana is subject to constant physical and spatial metamorphosis. In its urban territory, the different residential typologies are well distinguished by socio-economic conditions and the technical-constructive characteristics which have been influenced by political development of the country. The multi-family residential buildings from the post-war period up to the 1990s, in addition to the problems that accompanied them from their initial construction, are found today in front of a physical degradation derived from the years they have. Problematic residential spaces are also most of the buildings constructed in the first decade after 1990, characterized by a low technological and housing quality, derived from an uncontrolled and informal development of the building sector. The building quality in this research, is focused on the applied architectural standards, the technological solutions adopted and the energy consumption derived from them. From the analysis made it has been reached in the conclusions that a renewal and requalification process is necessary to reduce the energy waste and to increase the quality of housing within residential spaces. The main objective of this research is to contribute to the sustainable development of the residential area of Tirana, referring to both the architectural and technological scale. Sustainable development in this paper is closely linked to the quality of residential spaces, which is directly related to the quality of life of the inhabitants. Due to the complexity of the urban environment and its transformations over time, the identification of light regeneration and redevelopment strategies is fundamental, minimizing demolition works. This article includes numerous European examples and a theoretical part, from which proposals for the future development of the residential areas of the city of Tirana can be extrapolated.

Keywords: Requalification, regeneration strategies, sustainable development, residential space, energy efficiency

Introduction

The vast stock of multifamily residential buildings in Tirana is very dynamic. A part of this building stock is inherited from the regime period that means that they have fulfilled their life cycle of 50 years or that they're near to this process. Another part of the building stock has been constructed in the first years of 90s and for more than 15 years, the buildings were raised in a turbulent political situation that led to the poor quality of buildings in terms of their technology and energy efficiency parameters. Due to these phenomena renovation and requalification is needed in order to respect new energy savings in the buildings and to create more qualitative living spaces. Over the last decade, some improvements are made in this field, but there is nevertheless much to do in this area. It is valuable to find through the research some possible strategies for its future development.

The objective of this research is the definition of some possible interventions in order to improve the residential living spaces in Tirana neighborhoods. Due to the complexity of the urban environment and its transformations in time, the urban regeneration processes are the subject of an urban and architectural multi-scale study. The strategy proposals will be in this paper both in the architecture scale and in small scales accompanied with technology details. By defining some requalification strategies inside the existing urban pattern, some sustainable strategies for future development residential forms can come up, in order to be applicable in the analogue existing residential neighborhoods or in the new one. The sustainable development is here linked with the quality of the living spaces, which is closely linked to the quality of life of the inhabitants in these spaces.

Being in an existing urban pattern, drastic changes cannot be done and regeneration strategies can be identified in situ. Regeneration concept has been used from two centuries up to now in Europe but it is still an un-explored and a very new concept in Albania and in Balkans. In England, Germany, Netherland, France etc, there are projects applied in different social neighborhoods or peripheral ones and many academic studies has been conducted in Italy regarding to this topic.

Important reference book has been used here as, "*Rigenerare le aree periferiche*" of A. de Cesaris and D. Mandolesi; "*Linee guida per la riqualificazione dei quartieri innovative nell'Italia centromeridionale*" edited by F. Dematteis and B. Todaro; *Sustainable design. A Critical Guide, Princeton Architectural Press,* D. Bergman, (2012); *The building envelope. Applications of new technology cladding,* Brookes, A. J., Grech, Ch.; the collection of researches in *A HANDBOOK OF SUSTAINABLE BUILDING DESIGN AND ENGINEERING. An Integrated Approach to Energy, Health and Operational Performance,* edited by D. Mumovic and M. Santamouris, etc.

Existing Buildings Structure.

Various constructive and technological systems of the buildings before '90 and their conditions today

Residential neighborhods with multifamily houses started to be constructed in Tirana in the 1940s. The first buildings were constructed by italians as the 'Aeronautics buildings", "La casa degli impiegati", "Blloku Miqesia", etc. After the first years of forties, with the political regime changes new building strategy were applyed. New neighborhoods of residential blocks raised up. Some of them are *50 Vjetori*, *1 Maji*, *Shallvaret*, *21 Dhjetori*, etc. After 1945, in Albania, as in other countries of East Europe, was a centralized economy and a constraint city planning. In this period there, follows a reconstruction post war process and an urban evolution based on the principles of the *modern architecture.*[1] *«The modern architecture was never officially accepted by the regime of the Albanian state and it was secretly smuggled by the architects. Modern buildings were constructed not only in Tirana like Hotel Tirana (V. Pistoli), Flora residence (arch. M.Pepa and arch. I.Prushi), those in Dibra street (arch. M.Velo) but also in the city of Korça (arch. P. Kolevica) or in the buildings constructed in the residence center of Shkodra, etc.»*[2] Elements of traditional or classical architecture were substituted by pure forms and volumes. This was particularly noted not only in the residential areas, but also in the administrative and tertiary buildings.

The first '*social realism' buildings* in Albania realized in classical style were mainly inspired by the Russian Academy which itself is based on ancient Greek style, Roman Renaissance architecture and European neoclassicism. Some social, administrative and cultural buildings which belongs to this category are Kinostudio, the Ministry of Internal Affairs, residential buildings in Kombinati i Tekstileve "Lenin", the Lenin-Stalin museum, the Central Comity of the Albanian Communist Party, the theater of Shkoder, *the Parliament in Tirana*,[3] Shallvare palaces, Agimi palaces (arch. A. Strazimiri), etc. The architecture language of these buildings facades was a mix between classic European elements and local architecture elements.

After these first constructions in the end of fifties and the beginning of sixties due to the weakening of relations between Albania and the Soviet Union, having also a low economic level in the country, the architecture itself was oriented through certain standards that guaranteed minor costs of their constructions. The key factor of this '*new trend era*' was the influence of many architects that had studied abroad: in Poland (Enver Faja, Vasilica Silco), Bulgaria (Valentina Pistoli), Hungary (Mergim Çano), Romania (Kristaq Sahatçiu) and many others who brought many modern elements influencing the way that the projects were implemented. The classic trend of Russian Academy was left aside and the orientation was more on the modern *l'Existence Minimum.* An example of this new way of planning was the residence area in the "Miqesia" block. The local architecture art had to be free of any oriental influence and without any decoration elements. The architecture had to be functional, minimal and built for the masses usually using volunteering labor. The prisoners constructed the Agimi Palaces (in the end of 50's) through volunteering labor.

Architecture, as any other form of art, was developed through a strong control of the state. The projects were adjusted to the rules and approved by a very strict commission. The minimum standard of the apartment was 40-42 meter square, which included a number of four or five beds confirming the necessity to increase illumination and ventilation of each building. The control of the state in the architectonic development, actions against the western influences and the drastic reduction of the state investments led to poor architectonic solutions in terms of esthetics and quality. The development of the construction materials sector led to the standardization of buildings and their constant repetition through the entire city firstly with bricks and after 1965 with prefabricated panels. In this period, the four-five floor residential buildings were mostly developed. Analog plans were replicated through all the country. «The phenomenon of the standardization of the apartment

[1] *The modern influence in Albania was clearly seen in the architecture of Albanian architects such as Q. Butka, arch. S. Luarasi, arch. A. Lufi, E.* Faja, etc. Some administrative buildings in Tirana which can have elements of Modern architecture are the National Gallery of Arts, the National Historical Museum, etc

[2] Faja, E., *Gjeneza e Arkitekturës moderne dhe e realizmit socialist në Shqipëri, në vitet 1945-1980*, 16/12/2010 (in 55 Magazine Online, http://gazeta55.al/gjeneza-e-arkitektures-moderne-dhe-e-realizmit-socialist-ne-shqiperi-ne-vitet-1945-1980/)

[3] The foundations of the Albanian Parliament in Tirana were built based in the project of Bosio, but during the communist regime, the project was adapted according to the tendency of the time by the Albanian architect *Anton Lufi*.

space's design also included the standardization of dimensions and was called '*tipization*'. »[1] Consequently, in this way there were implemented standard coded projects (60.1; 60.2; 60.3 etc., referred to the years 1960, 70.1; 70.2 for the 70's and 80.1, 80.2 etc., for the 80's) used in all the new collective residential blocks. Rarely any modification would occur based on the exact location and in that case they had to be approved by the "the Central Comity"[2]. The construction material of the buildings in this period of time was realised with traditional brick walls of 38 cm and 25 cm for buildings of 3 and 4 floors and sometimes 50 cm in the first floors of the buildings with 5 and 6 storeys. The buildings were plastered in both sides mainly or only in the inside part. This technology was simple withoiut any exterior layer. Another construction typology was with concrete panel walls which was a technology coming from China. The concrete panels were ready to be installed in construction site and the plans were very rigid because of their construction typology, differently from the brick walls typology which allowed for greater flexibility. These multifamily buildings come today, after more than 40 years of life, with the need to repair facade elements, heating plants, doors and windows in order to improve indoor living conditions, minimize energy losses and improve aesthetic quality of the facades.

Constructive and technological systems after '90s and the main problem issues

The city grew up from a single function to a multi-functional center and developed in vertical level and in horizontal too. The rapid urban growth and the lack of capacities of public control because of the political transitions, influenced in the creation of an irregular urban landscape and in the low quality of the residential space. The increasing demand for new homes, created an unpleasant situation of informal parasites interventions in the facades of the buildings, both existing and new ones. The intrusion of new volumes on the existing building facades and in the urban pattern, implies a critical situation due to the lack of open collective spaces and public services. Consequently, not only the quality of the building facades that surround the open space has deteriorated, but in the meantime there were also created low aesthetic, environmental, compositional, architectural and technological quality open spaces. The presence of informal and individual interventions in the facades, without homogeneous material, dimensions or colures, has created a diversified pattern creating so an uneven façade interrupting the façade continuity by punctual interventions.

Figure 1. Building constructed after '90. Facade degradation.

Figure 2. Building constructed before '90. Facade trasformation

The new buildings have different typologies: tower typology or linear typology. Usually more buildings volume are joint together creating long buildings with more than a hundred meters. The stairwells are mostly closed and their relationship with the public space is missing. The façade development follows a division between the base, the body and the building crown. The ground floor is intended for commercial functions. The upper floors are used for houses function repeating the plan on multiple levels. The top floor is the roof that usually houses a smaller housing plan, so a different typology.

[1] Dobjani, E., Barandovski,l., Nelkovska, O., *Quality of life changes the quality of space*, Habitat 3, Tirana 2015, p.29

[2] The Central Comity in the Communist state is the term used to designate the most important political body of the state, directing the organization of congressional sessions and conducting deliberative functions electing the executive bodies. It takes on a greater relief over the parliament duties, which ratifies the decisions taken by the central committee.

The new buildings are built with load-bearing concrete structure (concrete beams and columns). The technology used for the building envelope is a very low quality. The exterior walls are made of 20 cm hollow bricks and in the most of the cases are only plastered. This poor technology influences the low quality of life in the interior spaces of the house as a derivative of the great heat dispersions. The lack of thermoinsulation or the presenze of thermal bridges shows problems as mold and humidity inside the apartaments and plaster deterioration in the façade.

Requalification and regeneration strategies of residential building

Building adaptation with new requirements of nowadays

«If new buildings must respond to the changing needs of society, then we must also consider how to adapt the vastly greater number of existing buildings.»[1]

The urban fabrics are considered as organisms constantly evolving that should be changed and adapted to the new needs of contemporary life. As Rogers states in his book Cities for a small planet, new buildings as also existing ones should satisfy the daily needs of their inhabitants and their needs change over time. The permanent high demand in Tirana for new housings and the lack of government institutions control in the partly modification of the building volume by inhabitants, matching with their apartment position in facade, led to the chaotic development we have today in multifamily housing shells and urban shape. Furthermore, in Albania there is no normative framework that regulates individual informally interventions in façade and neither do we have any requalification program how to adapt existent buildings to the new needs of the residents. Related to all these observations, the regeneration strategies in Tirana buildings, are focused in the process of sustainable development, energy efficiency and life quality, identifying possible strategies of interventions. Some important issues of intervention have been identified referring to the building regeneration and envelope transformation in terms of energy efficiency.

The façade of the building is considered an important element that relates and divide the interior inhabited space of a building and the outside space. It is not only an important element because of its aesthetic function but is also an important element because of its function as a protection layer against outside weather conditions. It is important because of the comfort that should insure to the inside space. In the first Venice Architecture Biennale in 1980, with the title 'The presence of the past'directed by Paolo Portoghesi, the attention was brought toward the façade and among 13 other installations, Dardi participated with the installation of Strada Novissima. In his text, Behind The Façade, he see the facade «as a meeting layer between internal structure and urban dimension, filter and diaphragm, a layer that function as a fence between public space and private one and relates them.»[2]

Germany, with sensitive policies towards the environment and energy saving, has organized during the past two decades, programs for restructuring and renewal operations to increase the quality of the internal space of the houses through improved thermal and acoustic performance by intervening in external envelope and improving the quality of the spaces in-between at the neighborhood level. The surgery helps retrain these neighborhoods and use new technologies or renewable materials.

In other post-socialist countries, as Serbia, Macedonia, Albania, etc., requalification programs have been missing totally. In Albania there are different laws related with the preservation of the "heat in the buildings" (Nr. 8937, date 12.9.2002) or the law on "approval rates, terms and conditions of design and construction, production and retention of heat in the buildings (DECISION no. 38 dated 16.1.2003), the LAW no. 9379, dated 28.4.2005, On energy efficiency or the Decision Nr.584, of 2.11.2000, on Energy savings and warmth keeping jobs, but these laws first of all have no "power" in the process of executive control in the field application leading to new buildings with low quality conditions; second, there is not any low related with the regeneration or requalification process of existing building stock constructed during the regime period.

A well-designed building enclosure, according to Magwood 2017, does four things: Keeps water out; Controls air flow into and out of the building; Keeps heat energy in or out, as desired; Manages vapor migration. A building that effectively controls water, air, heat, and vapor according to the demands of the climate and the needs of the occupant is a successful,

[1] Rogers, R., *Cities for a small planet*, ed. by Philip Gumuchdjian Faber & Faber, 1997, p.79

[2] *Catalogo della I motsra Internazionale di Architettura, La Biennale di venezia, 1980* oggi in Costanzo, M., (a cura di) *Costantino Dardi. Architettura in forma di parole*, Macerata, Quodlibet, 2009, e in Giancotti, A., *Trasformare l'involucro. Conservazione e riscrittura dell'imagine nel patrimonio dell'edilizia residenziale pubblica*, in Todaro, B., De Matteis, F., (a cura di) *Interventi sull'abitare. Linee guida per la riqualificazione dei quartieri innovativi nell'Italia centromeridionale*, Propsettive Edizioni, Roma, 2012, p. 151

comfortable, efficient, and durable building. Your basic understand of building science begins with identifying each of these four control layers in your building[1]

The current situation of the residential buildings in Tirana is categorized according to three construction periods, which also indicates their constructive character and façade quality:

_ Buildings constructed before 1944, which are mainly villas of one, two or three floors. Referring to INSTAT data, the stock of housing built before 1945, consists in 215 thousand apartments, mainly for individual family units, or more households, depending on the generation.

_ The period of the communist regime, from 1944-1990, led to the construction of public residential buildings, characterized by their height from three to six stores. During this period, only residential condominiums were built. The architectural and spatial composition of the condominiums and apartments was standardized in only a few typologies. Public residential buildings were owned by the state and were built with public funds using very often the volunteer work. Referring to the INSTAT data in this period are built almost 450 thousand houses.

_ After nineties, in the first decade of the capitalist period, it prevails the construction of private houses in the suburbs of Tirana in comparison with residential buildings of 6-12 floors, more widely spread in the second decade.

Buildings constructed during this period, increased the density in the central urban areas and spread to the outskirts of the city, creating a current housing stock of poor technological quality and rigid architecture.

	Up to 1960	1961-1980	1981-1990	1991-2000	2001-2005	2006-2011	Inhabited buildings*	Non-inhabited buildings*	Total
Tirana	6,066	8,087	8,204	34,259	16,175	12,424	13,895	11,173	110,283

Figure 3. The quantity of multi-family residential buildings in Tirana in different periods after 1960. (font: INSTANT)

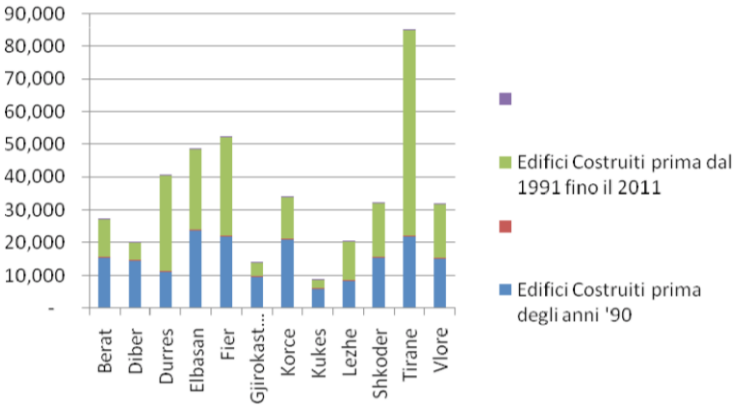


Figure 4. Edifici residenziali plurifamiliari costruiti prima e dopo gli anni '90 a Tirana. (fonte dei dati: INSTAT, grafica: l'autore)

Due to the low level of architectonic quality in houses and apartments, 25% of the total energy spent in Albania goes in residential sector. These parameters reflect on ways how to build and how the architects can improve the housing conditions. If regeneration and revitalization programs will be developed referring to the building envelope, the cost of energy for each housing unit, will be lower than the present. This intervention will improve consequently the thermal comfort inside the apartment, and at the same time will reduce the energy costs in the residential sector at the state level.

[1] Magwood, Ch., (2017) *Essential Sustainable Home Design_ A Complete Guide to Goals, Options, and the Design Process,* New Society Publishers, pg. 62

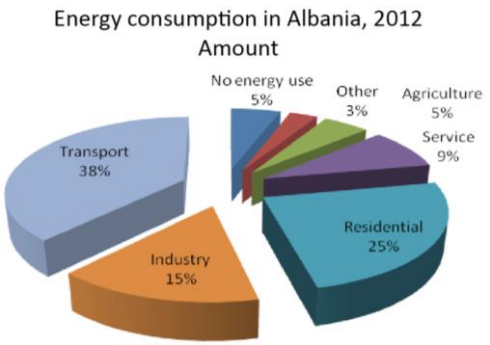


Figure 5. Energy consumption in Albania on various sectors, 2012. 25% of total budget goes for residential sector. (font: INSTANT)

As Rogers also states in his book Cities for a small planet, «the construction of our habitat continues to be dominated by market forces and short term financial imperatives. Not surprisingly, this has produced spectacularly chaotic results. It astounds me that the built environment in so many places remains an incidental political issue. »[1] This phenomenon is also very actual in Albania, because due to political change and the very long transition period the quality of residential spaces was very low. The total number of dwellings in Tirana in 2011, according to INSTAT data, is 1,012,400, and only 545.001 of these houses are in urban areas, the rest are rural housing. The biggest problems on the present situation of residential buildings, are their poor insulation of the walls and the roof, presence of thermal bridges and high energy consumption.

The current situation led to the necessity to improve the inside living conditions by improving the building facades or give suggestions for a residential building renewal. With all the efforts that Albania is making to join the European Communities, a very important area to invest in research is the new building stock, very problematic for its poor current conditions, which in 1950 will be considered at the end of their rehabilitation cycle (30-50 years).

Directive 2010/31 / EU requires all States Member or candidate countries to set minimum energy performance requirements for new and existing buildings, to ensure the energy certification and disciplinary controls on air conditioning systems and provides that, by 2021, all new buildings are "Nearly Zero Energy Buildings". State's Member should establish a national system of mandatory energy efficiency with the aims to achieve a cumulative goal of the final energy savings by 31 December 2020, at least 1.5% of the total energy sold to final consumers, for the period 2014-2020. These goals are not fulfilled by Albanian development and a lot of investments should be done in this field in the architectonic and technological conditions of the house, in order to reduce energy costs and at the same time to increase the quality of the inside living conditions of the inhabitants. In recent years, Albania has taken important steps in the approval of energy efficiency laws in particular with Law No.124/2015 (Energy Efficiency Law) and Law No.116/2016 (Law on the Energy Performance of Buildings), which comply with Directive 2012/27/EU and Directive 2010/31/EU, respectively. The Albanian legislation has not yet defined the minimum requirements of envelope, plant solutions, including those using renewable energy sources for new buildings or energy refurbishment.

Are analyzed below two multifamily residential buildings in Tirana referring to their architectonic details and their energy consumptions. These buildings belong to different periods of construction.

The first one has four floors and has been constructed in 1959. It has a constructive systems made of with retaining brick walls. It is located near the Blloku area, in Abdyl Frasheri street and Wilson ring.

[1] Rogers, R., *Cities for a small planet*, ed. by Philip Gumuchdjian Faber & Faber, 1997, p.17

Figure 6. The first multifamily house location, taken as case study

The building analyzed is a multi-family residential building with two staircases from which access four apartments per each floor. It was initially designed for four floors and then a fifth floor was added over time. It has a depth of 9 meters, a length of 26 meters and a height of 16.6 meters, with a load-bearing brick structure of 25 centimeters in width.

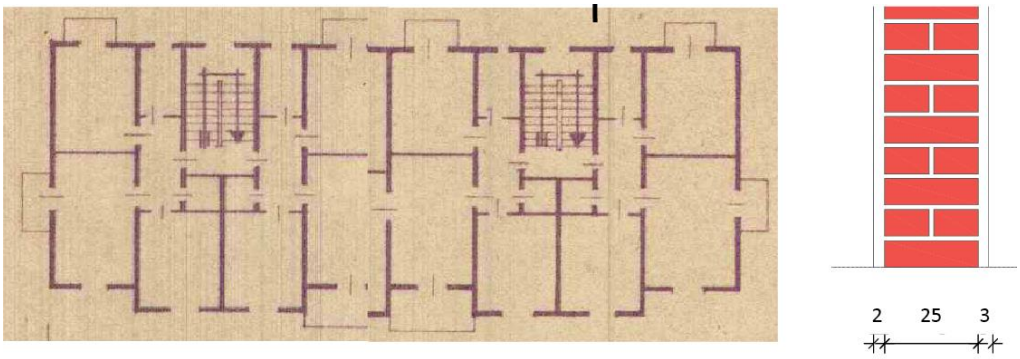
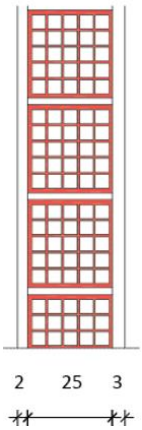


Figure 7. The first multifamily house plan and detail of the perimetral wall.

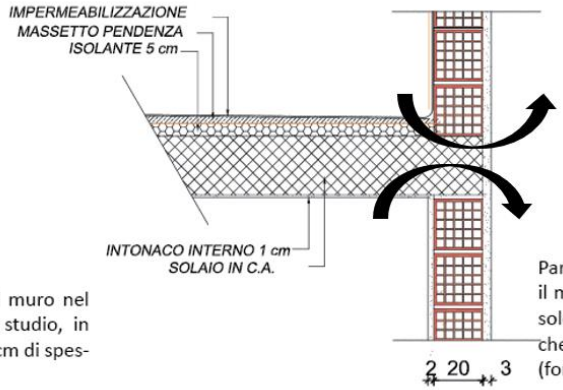
The second building has been constructed in the last two decades. It is a multifamily house designed for ten floors and it's located in a new neighborhood near Kombinat.

Figure 8. The second multifamily house taken as case study. Its location in the left and the plan in the right.

The building has been constructed in 2013. The type of building is an open U-shaped block with an internal courtyard. There are in total four stairwells from which four families enter per each floor. In total there are sixteen apartments per floor. It has a depth of 14.5 meters, a length of 37 meters and a height of 37 meters. The total area of the plant is 683 meters and the supporting structure is made with columns and beams in concrete material. The perimeter walls are 20 cm width, realized with hollow bricks and plastered in both outside and inside part.



Particolare del muro nel secondo caso studio, in laterizio di 20 cm di spessore. (fonte: l'autore)



Particolare del nodo tra il muro perimetrale e la soletta dell'interpiano che crea ponti termici. (fonte: l'autore)

Figure 9. Details of the perimetral wall and the node between the perimeter wall and the intermediate floor.

Possible interventions in the façade can be designed *adding a layer of thermal insulation; adding a layer of thermal insulation plus a layer of perforated bricks; making green facades, green roofs, and eliminate thermal bridges nodes.*

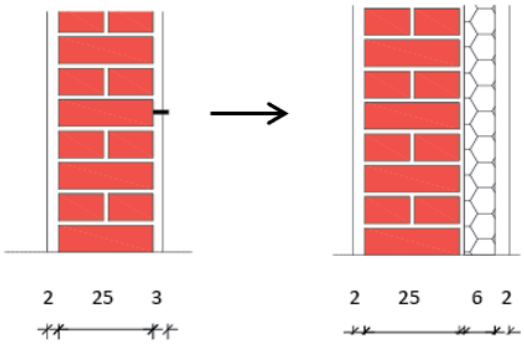


Figure 10. Possible interventions in the façade with traditional brick walls, by adding a layer of thermal insulation.

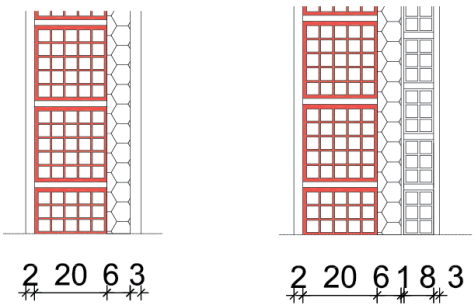


Figure 11. Possible interventions in the façade with hollow brick walls, by adding a layer of thermal insulation only, or a layer of thermal insulation plus a layer of external perforated bricks to improve thermal insulation of the external wall

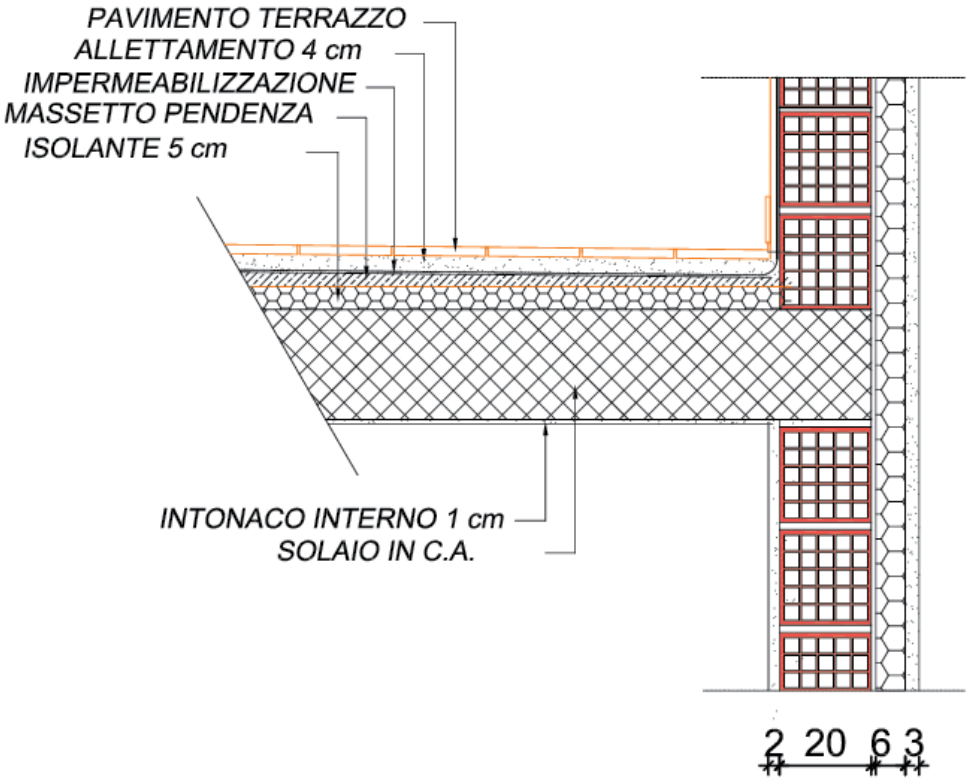


Figure 12. Possible interventions in the façade in order to eliminate the thermal bridges in problematic nodes.

Reconceptualization of the shell in residential buildings with new energy efficiency requirements

ENERGY PERFORMANCE CALCULATIONS

The two buildings taken as a key study in the previous paragraph will be analyzed here by their physical state, such as the building width, length, height and the surface number of floors, the coefficient S / V ratio. After all these data, will be calculated the value of U (W/m^2K), thermal transmittance of the architectural elements as external wall, roof, floors and windows or external doors. Will be calculated also the building losses caused by various factors. These calculations help us understand how much each of these buildings is spending today on energy to heat the inner environment and how much energy will be spent after the regeneration interventions. The ratio between the total investments in regenerating facades interventions and the savings gives us the time of the return of the investment.

These calculations help us understand how we can intervene in existing buildings, how much time do we need for the return of the investment costs and the benefit in Euro that we have due to electricity costs for each family per year.

Case 1. Multifamily residential building constructed in 1959

No.	Case study	year of construction
1	Multifamily residential house	1959

Width	Length	Height	Floor Area	Volume	No. Floors	Envelope Area (5 facades)	S/V	Perimeter
[m]	[m]	[m]	[m2]	[m3]		[m2]		
9	26	16.6	234	3884.4	5	1362	0.06	70

Architectonic Element	Area	U	Tin - T out	Losses because of transimittion (Qt)	Thermal bridges (Qt)	Losses because of orientation (Qt)		Losses because of Infiltration	Qtot
	[m2]	[W/m²K]	°C	W	W				[W/°C]
wall without win	1120.0	1.83	18.5	37917.6	3791.8	1895.9			43605.2
roof	200.0	1.39	18.5	5143.0	514.3	257.2			5914.5
floor	234.0	1.20	18.5	5194.8	519.5	259.7			5974.0
doors	18.9	5.80	18.5	2028.0	202.8	101.4			2332.2
windows	23.1	5.80	18.5	2478.6	247.9	123.9		35571.4	2850.4
								Qtotatle	96247.7 W
								GV=Q/(V*DT)	0.80764082 [W/m3ºC]

Performance Energetica	493.5861333	(kWh/m2 vit)
Bisogno Energetico per il riscaldamento del piano	115499.1552	kWh/vit
Costi annuali per il riscaldamento per l'intero edificio	1616988.173	Leke/vit
	11549.91552	euro / year

Annual Costs for a family	20212.35216	Leke/vit
Costs for a family in a month	144.373944	euro/anno
Annual Costs for a family	7.2186972	euro mese

Energy costs of the building are calculated here as the amount of total Energy needs for heating for the floor area in kWh/year and the Annual costs for heating in Leke(Albanian coins) /year.

Energy needs for heating are Initially Energy performance * floor area and "Initially Energy performance" is the [GV=Q/(V* ΔT)]/ /(1000*floor area in m2)

Q=Total losses

V= Volume

EXISTING SITUATION

Initially Energy performance	493.59	(kWh/m2 year)
Energy needs for heating for the floor	115,499	kWh/year
Annual costs for heating	1,616,988.17	Leke/year
	11,549.92	Euro/year

INVESTMENTS

Interventions	Surface	Costs	
External layer of 6 cm of thermoinsulation	3847	5,924,380	Leke
Serramenti	42.0	529,200	Leke

total investiments		6,453,580	Leke
		46,097.00	Euro

	U [W/m²K]	Energy Performance of the building [kWh/m2 year]	Costs Before interventions [euro/year]	Costs After Interventions [euro/year]	Savings [euro/year]	Expenses Costs	Investments return [year]
Muro perimetrale di 25 cm	1.83	493.59	11549.92				
Muro perimetrale di 25 cm + termisolamento di 3 cm	0.77	288.1213867		6742.04	4807.88	11200.00	2.3
Muro perimetrale di 25 cm + termisolamento di 6 cm	0.49	233.6538453		5467.50	6082.42	16800.00	2.8
Muro perimetrale di 25 cm + termisolamento di 3 cm + mattone forato di 8 cm	0.32	201.0896213		4705.50	6844.42	24640.00	3.6
Muro perimetrale di 25 cm + termisolamento di 6 cm + mattone forato di 8 cm	0.26	167.2460885		3913.56	7636.36	33600.00	4.4

Case 2. Multifamily residential house constructed in 2013

No.	Case study	year of construction
1	Multifamily residential house	2013

Width	Length	Height	Floor Area	Volume	No. Floors	Envelope Area (5 facades)	S/V	perimeter
[m]	[m]	[m]	[m2]	[m3]		[m2]		
14.5	37	32	**683**	**21856**	**10**	**4619**	**0.03**	123

case 3

Architectonic Element	Area	U	Tin - T out	Losses because of transimittion (Qt)	Thermal bridges (Qt)	Losses because of orientation (Qt)	Losses because of Infiltration	Qtot
	[m2]	[W/m²K]	°C	W	W			[W/°C]
wall without windows	3847.0	1.72	18.5	122411.5	12241.2	6120.6		140773.3
roof	600.0	0.37	18.5	4107.0	410.7	205.4		4723.1
floor	695.0	1.20	18.5	15429.0	1542.9	771.5		17743.4
doors	52.0	2.80	18.5	2693.6	269.4	134.7		3097.6
windows	37.0	2.80	18.5	1916.6	191.7	95.8	200146.3	2204.1

Qtotatle	368687.7	W
GV=Q/(V*DT)	0.39871175	[W/m3ºC]

Energy performance	469.7270968	(kWh/m2 year)
Energy needs for heating for the floor	326460.3323	kWh/year
Annual costs for heating for the whole building	4570444.652	Leke/year
	32646.03323	euro/year

EXISTING SITUATION

Initially Energy performance	469.73	(kWh/m2 year)
Energy needs for heating for the floor	326,460	kWh/year
Annual costs for heating	4,570,444.65	Leke/year
	32,646.03	Euro/year

INVESTMENTS

Interventions	Surface	Costs	
External layer of 6 cm of thermoinsulation	3847	5,924,380	Leke
Serramenti	89.0	1,121,400	Leke
total investiments		7,045,780	Leke
		50,327.00	Euro

	U [W/m²K]	Energy Performance of the building [kWh/m2 year]	Costs Before interventions [euro/year]	Costs After Interventions [euro/year]	Savings [euro/year]	Expenses Costs	Investments return [year]
Muro perimetrale di 20 cm, con mattoni perforati	1.72	469.7270968	32646.03				
Muro perimetrale di 20 cm, con mattoni perforati + termisolamento di 3 cm	0.75	248.467303		17268.48	15377.56	11200.00	0.7
Muro perimetrale di 20 cm, con mattoni perforati + termisolamento di 6 cm	0.48	186.8795253		12988.13	19657.91	16800.00	0.9
Muro perimetrale di 20 cm, con mattoni perforati+ termisolamento di 3 cm + mattone forato di 8 cm	0.321	150.6111673		10467.48	22178.56	24640.00	1.1
Muro perimetrale di 20 cm, con mattoni perforati+ termisolamento di 6 cm + mattone forato di 8 cm	0.31	148.1020357		10293.09	22352.94	33600.00	1.5

Due to energy performance calculations made in two buildings, we can see that the energy performance of the buildings if we add a layer of six centimeters in both buildings, improve a lot in comparison with the existing situation. In the first case, the costs for energy heating are reduced from 12080.6 euro per year in 5557.65 euro per year and the time of return of the investment is seven years. In the second building the costs for energy heating are reduced from 34231.35 euro per year in 17585.54 euro per year and the time of return of the investment is three years. That means that the energy losses in the second case are bigger and the savings with the new interventions are bigger in the energy savings.

Conclusions

The main goal of this research is giving a contribution in the sustainable development of the residential space in Tirana and defining some possible strategies to be applicable in the existing residential buildings constructed before and after nineties. The quality of facades in the existing stock of the buildings, except the architectural perspective has an emergent need for interventions to improve the degradation over time and it is an important architectonic element, which is closely associated with the thermal losses of the building and the creation of residential warm environments, thus affecting also in the energy consumption costs. Residential blocks built before nineties, have established need for reconstruction and interventions to adapt the volume to the new energy requirements. New buildings constructed in the last three decades with facades made of bricks of twenty centimeter, shows high thermal losses and in the presence of thermal bridges is created mold and moisture inside the living space, thereby leading to the need for interventions in the façade, roofs and openings to reduce thermal loss, creating so warm environments with less energy expenses. The big energy losses because of the bad quality of the perimeter wall, the presence of the moisture and mold inside the living space, the plastering degradation in the major stock of new and old buildings, became a focal point of the research. Possible interventions are proposed reflecting in terms of not only international references and examples how to improve the outside wall but also referring to some calculation in situ to define better the problem. There has been made some calculations of thermal transmittance of the wall, or the U-value, to see the rate of the heat transfer from the inside living space and the energy performance of the wall as it is now and how can it improve by adding other layers.

In conclusion, there have been proposed some interventions to improve the residential living space. In the building scale, there are proposed intervention in the façade by adding other layers of thermal insulations and second layers to increase the energy performance as bricks or panels of different materials, or green element that can improve building performance in terms of architectural point of view and also the energy consumption of the building as green facades, green roofs, and the elimination of thermal bridges.

Bibliography

[1] Bergman, D. (2012) *Sustainable design. A Critical Guide,* Princeton Architectural Press, New York
[2] Brookes, A. J., Grech, Ch. (1990) *The building envelope. Applications of new technology cladding,* Butterworth Architecture, London
[3] Dobjani, E., Barandovski,I., Nelkovska, O., *Quality of life changes the quality of space*, (2015) Habitat 3, Tirana
[4] Dobjani, E., (2010) *Città e Abitazione. Tirana 1920-2006,* A+P 5, Tirane
[5] Faja, E., (2010) *Gjeneza e Arkitekturës moderne dhe e realizmit socialist në Shqipëri, në vitet 1945-1980*, (available in 55 Magazine Online, http://gazeta55.al/gjeneza-e-arkitektures-moderne-dhe-e-realizmit-socialist-ne-shqiperi-ne-vitet-1945-1980/)
[6] Rogers, R., (1997) *Cities for a small planet*, ed. by Philip Gumuchdjian Faber & Faber
[7] Giancotti, A., (2012) *Trasformare l'involucro. Conservazione e riscrittura dell'imagine nel patrimonio dell'edilizia residenziale pubblica*, in Todaro, B., De Matteis, F., (a cura di) *Interventi sull'abitare. Linee guida per la riqualificazione dei quartieri innovativi nell'Italia centromeridionale*, Propsettive Edizioni, Roma
[8] Magwood, Ch., (2017) *Essential Sustainable Home Design_ A Complete Guide to Goals, Options, and the Design Process,* New Society Publishers
[9] Haddock, S. V., Moulaert, F., (2009) *Rigenerare la città. Pratiche di innovazione sociale nelle città europee,* il Mulino, Bologna
[10] Hadjri, K., UCLan, (ed.)(2004) *Readings on contemporary housing research,* in Oikonet, A global multidisciplinary network on housing research and learning
[11] L. Reale, (2015) *La Residenza collettiva,* SE Sistemi Editoriali, Napoli
[12] Magwood, Ch., (2017) *Essential Sustainable Home Design. A complete guide to goals, options, and the design process,* Canada
[13] Misja, V., Misja, A., (2004) *Veshtrim mbi situaten e banesave ne Shqiperi*, Akademia e Shkencave te Shqiperise, Instituti Ekonomise, . Tirane
[14] Riza, E., (2009) *Kapitalizmi, si i ndryshon banesat tiranase,* Gazeta Shqipetare, Tirane
[15] Sergi, G., (2008) *Tirana, una città emergente. Politiche urbane, Piani e Progetti*, COEDIT Cormagi Editore, Genova
[16] Smith M., Whitelegg, J., Williams N. J., (1998) Greening the Built Environment, earthscan, London
[17] Todaro, B., De Matteis, F., (a cura di) (2012) *Interventi sull'abitare. Linee guida per la riqualificazione dei quartieri innovativi nell'Italia centromeridionale*, Propsettive Edizioni, Roma
[18] Tsenkova, S., (2005) *Trends and Progress in Housing Reforms In South Eastern Europe*, Paris
[19] Zucchi, V., (2011) *La qualità ambientale dello spazio residenziale,* Franco Angeli
[20] Dastbaz, M., Strange, I., Selkowitz, S., (Ed.2016), *Building Sustainable Futures Design and the Built Environment*, Springer International Publishing Switzerland
[21] Mumovic, D., Santamouris, M., ed. (2019) A Handbook of Sustainable Building Design and Engineering *An Integrated Approach to Energy, Health and Operational Performance,* Routledge, New York
[22] SABNIS G. M., ed. (2016) *Green building with concrete. Sustainable design and construction,* CRC Press, Taylor & Francis Group, NW

A Study of Gas Diffusion Characteristics on Nano-Structured Ceramic Membranes

Priscilla Ogunlude

Robert Gordon University, Centre for Process Integration and Membrane Technology School of Engineering

Ofasa Abunumah

Robert Gordon University, Centre for Process Integration and Membrane Technology School of Engineering

Edward Gobina

Robert Gordon University, Centre for Process Integration and Membrane Technology School of Engineering

Abstract

The use of membranes for gas upgrading has increasingly become of interest as it has shown great potential for efficient and affective gas purification and a pathway to green energy. The emission of greenhouse gases to the atmosphere has detrimental effects on the economy in terms of global warming which has led to many natural disasters, heat waves, food shortage, loss of life and property. To combat this, studies of capturing and utilizing greenhouse gases are ongoing. In this paper, the study of biogas components (methane and carbon dioxide) diffusion through membranes are studied to employ its use as a solution for the challenge. The study involved the use of membranes of different pore sizes (15, 200 and 6000nm) to ascertain the flow characteristics and regime of the gases under different operating conditions. Single gas permeation tests were conducted, and the results show the flow of gases is dependent on factors including molecular weight, kinematic diameter and viscosity of the gas components. It was observed that pressure has a greater influence on the gas flow through membranes compared to temperature with the effect of pore size having the greatest impact. The flux of methane through the membrane is greater than that of carbon dioxide in regular pore geometry and depicts a greater potential for upgrading of biogas.

Keywords: Biogas, upgrading, emissions, nano-structured, membrane, carbon capture

The Dimensional Stability and Durability of Acrylic Resins for the Injection of Cementitious Systems

C. S. Paglia

University of Applied Sciences of Southern Switzerland, DACD, Institute of Materials and Constructions, Trevano, CP 12, 6952 Canobbio, Switzerland

A. Krattiger

Drytech SA, Impermeabilizzazioni, via Industrie 12, 6930 Bedano, Switzerland,

Abstract

The dimensional stability and durability of Acrylamide- and Methacrylate-based acrylic resins have been studied. The dimensional stability was characterized by measuring the volume expansion of samples immersed in water for a period up to 240 days or by exposing the samples to 23°C and 50 %, 90 % relative humidity and by monitoring the shrinkage. The durability was investigated by exposing the resins to cyclic variations of temperature in air and in water. The resins generally exhibit a significant volume change up to 160 % of the initial volume when immersed in water or exposed to a relatively dry atmosphere (23°C and 50 %). A general increase in the material stiffness and/or crack formation on the surface of the resin is observed. On a long-term basis, the durability of the resins may significantly vary with occasionally a partial or complete deterioration of the some resins. A general better dimensional stability and durability is observed for the Methacrylate-based resins as compared to the Acrylamide-based resins.

Keywords: dimensional stability, durability, acrylic resins, injection

Simpler Machine Learning Using Spreadsheets: Neural Network Predict

Leong Thin-Yin

Singapore University of Social Sciences (SUSS)

Leong Yonghui Jonathan

Singapore University of Social Sciences (SUSS)

Abstract

Machine Learning as a phenomenon has gone viral, with many technologists and software vendors promoting it. However, offered tools remain highly technical and not accessible to those without rigorous training in Computer Science or Business Analytics. It would be more useful if end-users can understand it beyond the sales pitch or blind application, and perhaps, even work from scratch to build simple models without much additional training. With better assimilation and acceptance of this AI methodology as an acquired skill and not just head knowledge, many more may want to invest the intensive effort to learn the required tough mathematics and cryptic programming. Or, after simple trial explorations, be willing to put aside substantial budgets to employ skilled professionals for full-scale business application. With simplicity and accessibility in mind, this paper renders Neural Network, a key machine learning methodology, on the ubiquitous and easily comprehensible spreadsheet without macros or add-ins, employing only elementary operations and if so desired, optionally leveraging on its built-in Solver. We will show that backpropagation can be achieved using the elegant though obscure recursive computation feature, with no need for Solver. We will demonstrate the application of neural network on a familiar problem: early and prior prediction of students' graduation GPA. The paper can be used to form the core content for introducing machine learning to non-technical audiences, particularly those majoring in Business and the Social Sciences.

Keywords: machine learning, neural network, business analytics, spreadsheets, education, higher education, experiential learning