Towards Introducing Execution Tracing to Software Product Quality Frameworks

Tamás Galli, Francisco Chiclana, Jenny Carter, Helge Janicke

Centre for Computational Intelligence, Faculty of Technology, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom

p10553741@myemail.dmu.ac.uk, chiclana@dmu.ac.uk, jennyc@dmu.ac.uk, heljanic@dmu.ac.uk

Abstract: Execution tracing and logging significantly influence the time spent on localizing software errors; consequently, they have essential impact on maintainability. Moreover, in certain situations these tools are the best suited instruments to analyse the behaviour of distributed, multithreaded or embedded applications. In spite of this, software product quality frameworks do not include execution tracing or logging as a quality property. In this paper we examine the extension possibilities of the present software product quality frameworks to accommodate execution tracing. In addition, the scope of the investigation includes facilities of the frameworks to address the uncertainty involved in quality measurements.

Keywords: quality frameworks; execution tracing; logging; uncertainty

1 Introduction

Execution tracing and logging are frequently used as synonyms in software technology; however, the first one rather serves the software developers to localize errors in applications, while the second one contributes to administration tasks to check the state of software systems [12], [20], [28], [30], [35]. In the scope of this paper we also use the two phrases as synonyms.

Execution tracing dumps the data about the program state and the path of execution for developers for offline analysis, which helps to investigate error scenarios and follow changes in the state of the application. Thus, execution tracing and logging belong to dynamic analysis techniques i.e. testing, investigating live systems, which are integral parts of the maintenance activities. Dynamic analysis techniques can be applied only if the software is built and executable in contrast to static analysis techniques. However, both methods are applied to achieve the same goal of diagnosing errors, with each technique having its own particular advantages [5], [8], [11], [41].

Spillner, Linz, Schaeffer in [37] make distinction between two types of software maintenance: (1) corrective maintenance, the purpose of which is eliminating errors in the software and (2) adaptive maintenance to change the software according to new requirements. Both kinds of maintenance necessitate analysis methods to find errors but this activity dominates in corrective maintenance. The proportion of maintenance costs in the whole software life-cycle amounts to a large part [1], [25], [26], thus decreasing the time devoted to localizing errors can therefore decrease the maintenance costs.

The increasing size and complexity of software systems makes localizing software errors more difficult. This difficulty is aggravated by the enormous number of software and hardware combinations. Adding execution tracing to key places of the application can drastically reduce the time spent with debugging [3].

Utilizing a debugger is time consuming and does not offer adequate solution

- if performance problems have to be resolved because debugging the source code considerably changes the environment from point of view of execution performance. Moreover, performance is sensitive to external workload, configuration parameters, underlying hardware and software components [3].
- in case of real-time, embedded systems as it might be harmful or impossible to reproduce the error e.g. in control applications [39].
- in the case of concurrency, as it changes the race conditions for parallel running execution threads or processes. In addition, multi-core systems also need to be considered which may even have multiple clock domains [39].

A wide survey on concurrency [10], for which 10% of all Microsoft employees from development, test, and program management were selected, also supports that analysing concurrency faults makes up a significant part of their correction costs. 66% of the respondents had to deal with concurrency issues. The reproduction of these issues was classified in a five categorical scale ranging from easy to very hard. 72.9% of all responses classified reproduction of concurrency issues in the two most difficult categories. Moreover, the respondents stated that the severity of these issues, qualifying on an ordinal scale with four categories ranging from least severe to most severe, belongs to the top two: most severe, and severe. In addition, 65% of the respondents expressed the future expectation that the concurrency issues would be more problematic.

Laddad states in [27] that execution tracing is the only adequate tool to help with the analysis of run-time errors in the case of distributed systems and multithreaded applications. In the case of embedded applications, which have no user interface, by means of tracing the developer or system maintainer can answer questions such as what the application is doing [39].

Diagnosing regression test errors and finding root causes implicate major difficulties. Fault localizations can be grouped in three categories [32]: (1) dynamic dependence analysis of the failing program execution, (2) comparison of the failing program execution with a set of error free executions, (3) comparison of the failing program execution with a program execution which does not manifest the error in analysis. Variants (2) and (3) are based on execution tracing.

An experiment conducted by Karahasanovic and Thomas [21] categorized the difficulties related to the maintainability of object-oriented applications. Program logic was ranked the first in the source of difficulties. Understanding the program logic belongs to the category of software specific knowledge which can greatly be enhanced by execution tracing, offering a basis for trace visualization and program comprehension [36].

Tracing, logging or constraint checks represent significant parts of the source code of applications. Spinczyk, Urban and Lehmann [38] state that the ratio of code lines related to monitoring activities such as tracing, logging reached approximately 25% in their measurements of commercial applications. This ratio shows that a significant amount of source code is written to deal with execution tracing, which in itself is an important quality factor. However, execution tracing does not need to be tightly coupled to the application code and can be localized in separate modules [27], [31], [40].

All the above indicate that execution tracing and logging have essential impacts on the analysability of software systems. In certain cases these tools are inevitable to localize errors or investigate software behaviour. Nevertheless, present software product quality frameworks do not exhibit any property to describe execution tracing but they usually offer the potential to be extended. In this paper we analyze such extension points and articulate concrete possibilities for extension in the context of the current investigation. Software product quality frameworks form complete models to support the description and assessment of the quality of software products. As research shows [23], conformance with process quality models does not guarantee good-quality software products, motivating the application of software product quality frameworks in synergy with process quality models.

In addition, measuring software product quality is difficult. Some quality measure elements are easier to measure than others even if the quality measure element is well defined [33]. All of the software product quality frameworks reviewed (Section 2) include the description of qualitative properties in a quantitative manner and quality measure elements which cannot be measured directly but only derived from the observation of the behaviour of software developers, maintainers, operators and users. This condition introduces uncertainty in software product quality frameworks, which has recently been admitted and accepted by defining the subjective measurement method category in ISO/IEC 25021:2007:

> "Subjective measurement method - Subjective measurements are those where quantification is influenced by human judgement. Subjective measures are used when no formal objective procedures of measurement can be applied. The value of the quality measure element is influenced by human judgement as an evaluator. Therefore it is necessary to interpret the results with respect to the number of evaluators and statistical methods used for the measurement result calculation. Both should be stated while presenting the measurement results."

Manifestations of uncertainty can be classified into three broad categories: (1) objective uncertainty that refers to the future, (2) subjective uncertainty that refers to the future, and (3) subjective uncertainty that does not refer to the future but helps to categorize elements [6], [24]. Category 1 is modelled by the classical probability theory, while category 2 is considered as an application area of Bayesian statistics. Category (3) on the other hand, is modelled and studied under the name of fuzzy logic. Thus, we also aim to examine in the scope of the research how far the current quality frameworks can ensure the link to quality measures described by means of fuzzy logic to consider the above subjective uncertainty. The authors have already presented a pilot study on modelling execution tracing quality by type-1 fuzzy logic [9].

Summarising the above, the main contributions of the paper, which will be elaborated in detail in the following sections, refer to: (1) the need for execution tracing quality to be appropriately implemented within software product quality frameworks; (2) the significant differences between the current software product quality frameworks to allow such implementation; (3) the ability of the ISO/IEC software product quality frameworks to provide mathematical computations to define metrics and measures, which can be exploited in capturing and implementing subjective uncertainty within their quality models; and finally, (4) the outline of metrics and a measure for execution tracing quality for both the ISO/IEC 9126-1 and ISO/IEC 25010 software product quality frameworks.

The rest of this paper is structured as follows: Section 2 introduces the software product quality frameworks. Section 3 demonstrates the extension facilities of these frameworks, while section 4 presents a discussion on the particular changes in the frameworks required to encompass execution tracing and finally the last section closes with the conclusions.

2 Software Product Quality Frameworks

The analysis of quality frameworks was conducted, using IEEE, ACM and EBSCO databases, to discover existing alternatives or predecessors to describe software product quality. The investigation focused on software product quality models, which describe the whole set of software product quality; therefore we refer to them by the term software product quality frameworks.

In the current frameworks [13], [14], [23], [29], if the traceability property is present it refers to requirement traceability, i.e. how requirements can be followed during the development of the software. Some quality metrics and measures defined in ISO/IEC 9126-2:2003, ISO/IEC 9126-3:2003 and in ISO/IEC 25021:2007 show overlapping and similarities to execution tracing but their definitions are ambiguous and difficult to approach from a practical point of view: e.g. calculating the ratio of the number of diagnostic functions and the number of necessary diagnostic functions. Such metrics are associated with the Internal and External Analysability sub-characteristic of the characteristic Maintainability: (1) Activity Recording, (2) Readiness of Diagnostic Functions, (3) Audit Trail Capability, (4) Diagnostic Function Support, (5) Failure Analysis Capability, (6) Failure Analysis Efficiency, (7) Status Monitoring Capability [17], [18]. This means that the existing frameworks do not consider execution tracing as an aspect of software quality.

2.1. Early Frameworks

Early software product quality frameworks appeared in the second half of the 1970's to assess quality and show the way for improvements in software products [2], [29]. These frameworks had a significant influence on the recent software product quality frameworks published by the ISO standards. They kept the hierarchic nature abstraction of quality.

2.1.1. Software Product Quality Model of Boehm, Brown and Lipow

The first complete model to assess software product quality was developed by Boehm, Brown, and Lipow [2]. They established a set of quality properties, which they call characteristics, and one or more metrics to each of them. They defined the notion of metric as (1) a quantitative measure that describes the degree to which the software product possesses the given characteristic, and (2) the overall software quality must be able to be described by the function of the values of the metrics.

They came to the conclusion in their study that establishing a single overall metric for software product quality would implicate more difficulties than benefits because many of the major individual quality characteristics are conflicting; moreover, the metrics they associate to the quality characteristics are incomplete measures of the quality characteristics. Therefore, they developed a hierarchical model. The hierarchy comprises of eleven high-level characteristics representing different aspects of software product quality [2]: (1) understandability, (2) completeness, (3) conciseness, (4) portability, (5) consistency, (6) maintainability, (7) testability, (8) usability, (9) reliability, (10) structuredness, and (11) efficiency.

The model is language-independent and independent of programming paradigms, however many metrics were tested on the structured language Fortran. Additional metrics to the published ones can easily be defined and the model offers possibilities for tailoring.

2.1.2. Software Product Quality Model of McCall, Richards and Walters

McCall, Richards, and Walters published a different framework [29] to Boehm's model [2] to assess software product quality. The authors describe a global view of software product quality as a combination of three distinct activities: (1) product operation, (2) product revision, and (3) product transition i.e. the description considers also process related properties. The objective of their investigation was to provide a concept to acquisition managers to specify and measure quality in a quantitative manner in software products related to air force applications.

They established a set of software quality properties that describe the overall quality of the software product and they named these properties factors. The quality factors they associated with criteria. Criteria are attributes of the software or software development process by which the factors can be judged and defined. A criterion can have sub-criteria in a hierarchical manner and one criterion may affect more quality factors. The criteria are coupled with metrics that make possible the measurement of the criteria or sub-criteria. The separation between properties that would also qualify for being both criterion and factor the authors made the decision: user-oriented properties are quality factors while software-oriented are criteria.

Quality farctors:

(1) Product operation: (a) correctness, (b) reliability, (c) efficiency, (d) integrity, (e) usability; (2) Product revision: (a) maintainability, (b) testability, (c) flexibility; (3) Transition: (a) portability, (b) reusability, (c) interoperability;

The authors also investigated the impact of the quality factors on each other, i.e. if a particular factor is present with a high degree of quality what quality is expected for the other factors. Beside the positive relationships, there exist also negative ones between some quality factors. In those cases finding a compromise is crucial, e.g. integrity and interoperability conflict with each other, which means that the more interoperable the system is the more difficult it is to keep its integrity.

Similarly to the previous framework, this model is language-independent because its metrics are language independent and independent of programming paradigms. It leaves room for extension and tailoring.

2.2. Recent Frameworks

Recent software product quality frameworks appeared after 1990. They can be divided into three categories on the basis of their philosophy [7], [13], [14], [22]: (1) hierarchic models of the ISO/IEC standards which are strongly influenced by the early frameworks, (2) adaptations of the ISO/IEC standards, and (3) the non-hierarchic framework of Dromey. Their presentation follows in historical order.

2.2.1. Software Product Quality Model of ISO/IEC 9126 Standard Family

The quality model of the ISO/IEC 9126 standard family comprises of a hierarchic model for software product quality and quality in use. The first version of the standard was issued in 1991 which was superseded by the next version issued in 2001 [14]. The description of software evaluation was moved from the second version to the multipart standard ISO/IEC 14598 [15]. The standard ISO/IEC 25010:2011 [13] revised the quality models described by the ISO/IEC 9126 standard family.

Terminology:

- Quality characteristics: high-level quality properties which are located at the top of the hierarchy. In the terminology of ISO/IEC 14598 standard family they are called attributes.
- Sub-characteristics: Quality characteristics which are located somewhere in the hierarchy but not at the top-level. Sub-characteristics are always assigned to a higher level characteristic or sub-characteristic.
- Quality metrics: Definition of the measurement method of quality properties including the definition of the measurement scale. Quality metrics are assigned to sub-characteristics or characteristics.
- Internal quality metrics: Metrics whose inputs are formed by the intrinsic properties of the software product.
- External quality metrics: Metrics which cannot be measured directly but only derived how the software relates to its environment.
- Quality of use: The user’s view of quality.

Concepts:

The standard ISO/IEC 9126-1:2001 defines three basic views of the quality: (1) internal view, (2) external view, (3) user’s view. Internal view of the quality means the quality measured by the internal quality metrics. This reflects the quality of the source code or documentation. It is very useful if the software product is not developed as far as it could be tested. The external view of the quality is measured by the external metrics. It shows how the product relates to its environment. The user’s view of the quality is illustrated by the quality in use reflected by the quality in use metrics.

Internal and external metrics either need to be in cause-effect relationships or they need to correlate with each other. This is called predictive validity i.e. from the measurement by the internal metrics conclusions can be drawn relation to the external metrics and external quality of the software.

The software product quality model introduces six high-level characteristics: (1) functionality, (2) reliability, (3) usability, (4) efficiency, (5) maintainability, (6) portability. In addition to their sub-characteristics, each of these characteristics has an internal and external variant to form an internal and external model.

The quality in use model has four high-level characteristics without sub-characteristics: (1) effectiveness, (2) productivity, (3) safety, (4) satisfaction. External metrics need to have predictive validity for the quality in use metrics.

The model is language-independent and independent of programming paradigms.

2.2.2. Software Product Quality Model of Dromey

Software does not directly display quality properties but it shows product properties, which contribute to the quality properties in a positive or negative way. Dromey argued that the previously published software product quality models adequately addressed these particularities. He proposed a model where the main focused was on the product properties, which he calls quality-carrying properties, and on the relationship between product and quality properties in a non-hierarchic manner [7].

Terminology:

- Quality attribute: high-level quality property
- Structural form: programming language constructs
- Quality-carrying properties: binary-value variables which determine the quality

As quality attributes Dromey identified the six high-level quality characteristics of the ISO/IEC 9126:1991 standard and extended this set with the attribute reusability [7].

Concepts:

The model makes possible the specification and analysis of the relationships between quality attributes, quality-carrying properties, and structural forms. The bottom-up approach facilitates for developers to specify or investigate which quality-carrying properties be associated to the structural forms of a particular application. The top-down approach facilitates for designers to specify the quality requirements and attributes the software needs to satisfy and identify the quality-carrying properties for the structural forms to fulfil the quality needs [7].

Quality is depicted by the relations: (1) between quality-carrying properties and quality attributes; (2) between quality-carrying properties and structural forms. Dromey also proposes profiles for both relations. Quality-carrying properties and structural forms have precedence rules in such profiles. If the precedence rules are kept, the model is able to classify software quality defects.

The basic mechanism of the model can be formalized as (1) if each quality-carrying property of a structural form is satisfied, then that structural form will have no quality defect; (2) if a quality-carrying property of a structural form is violated, then it will contribute a quality defect to the software.

The definition of the model is language dependent in contrast to the previously presented software product quality frameworks because it uses programming language level constructs as structural forms and their properties as quality-carrying properties. It was prepared for supporting the procedural programming paradigm. However, the concepts can also be extended for other programming paradigms and different artefacts, including program documentation.

2.2.3. Software Product Quality Model of Kim and Lee

Kim and Lee [22] derived a model from the product quality model of the ISO/IEC 9126:2001. The authors determined the relative importance of the six high-level characteristics of the ISO standard from the point of view of the objectives of the project under examination. The order of the relative importance of the six characteristics was computed by applying the Analytic Hierarchy Process [34]. Those characteristics were kept for further investigation, the relative importance of which exceeded a defined threshold. In their case study they found three such attributes in the particular context: (1) reliability, (2) maintainability, and (3) portability.

They identified internal metrics for static code analysis and assigned these metrics to the three high-level characteristics of the ISO/IEC 9126:2001 model by considering the opinions of experts [22]. The metrics have directly been assigned to the high-level characteristics. Consequently, no intermediate level in the hierarchy with sub-characteristics was defined, i.e. the three characteristics formed categories rather than hierarchies.

The authors also presented the evaluation of a software component to illustrate the use of their model [22]. The critical places for improvement were identified in the component analysed. After performing amendments of the identified quality defects, the evaluation was carried out again, which verified the impact of the corrections.

2.2.4. Software Product Quality Model of the ISO/IEC 25010 Standard

The ISO/IEC 25000 standard family supersedes the ISO/IEC 9126 and ISO/IEC 14598 standard families. ISO/IEC 25010:2011 [13] defines a new quality in use model and a new software product quality model combining the internal and external models of the ISO/IEC 9126 standard family. However, it keeps the concepts laid down by the previous ISO/IEC 9126-1:2001 standard [13].

Terminology:

- Definition of internal, external view of quality and quality in use are taken over from the predecessor ISO/IEC 9126-1:2001 standard [14] but the internal and external software product quality models were combined to one software product quality model.
- Quality Measure Element (QME): measurable property of quality defined in ISO/IEC 25021:2007 [16].
- Quality Measure (QM): quality measure elements and a measurement function to calculate with. It is similar to the term metric in the ISO/IEC 9126-1:2001 standard. Initial list of quality measures was taken over from ISO/IEC TR 9126-2:2003 [17], ISO/IEC TR 9126-3:2003 [18] and ISO/IEC TR 9126-4:2004 [19].
- Quality attribute: low-level quality property, in contrast to the ISO/IEC 14598 standard family where the term attribute is used for the high-level quality properties of the ISO/IEC 9126 family.

Concepts:

The software product quality model introduces slight changes in the naming of the six high-level characteristics of the ISO/IEC 9126-1:2001 standard and adds two further high-level characteristics to the previous model: security, compatibility. The whole list of high-level quality characteristics: (1) functional suitability, (2) performance efficiency, (3) compatibility, (4) usability, (5) reliability, (6) security, (7) maintainability, (8) portability. In addition, the sub-characteristics were partially modified.

The quality in use model defines one additional high-level characteristic to the previous characteristics defined in ISO/IEC 9126-1:2001 and performs slight changes in the naming. The present high-level characteristics of the quality in use model: (1) effectiveness, (2) efficiency, (3) satisfaction, (3) freedom from risk, (4) context coverage. The new model also defined sub-characteristics which were not part of the previous quality in use model.

The new model description emphasizes the necessity of tailoring the model to the specific objectives of projects.

The new members of the ISO/IEC 25000 standard family: ISO/IEC 25022, 25023, 25024 are expected to be issued in the future. These standards will suspend the validity of the previous technical reports that define internal, external and quality in use metrics: ISO/IEC TR 9126-2, 9126-3, 9126-4.

These models are language-independent and independent of programming paradigms.

3 Towards Extending Present Quality Frameworks

The software product quality frameworks presented in the previous section show two basic approaches for describing software product quality: (1) the hierarchic approach depicted by the ISO/IEC 9126 and ISO/IEC 25000 standard families which have their roots in the early models; and (2) the non-hierarchic approach described by Dromey. The other frameworks tailor the first approach or its predecessors to the specific context of use. All the frameworks presented are the result of empirical research, which offers possibilities for changes and tailoring.

For this reason we will investigate the extensibility of the ISO/IEC 9126, ISO/IEC 25010 quality frameworks and the framework defined by Dromey.

This investigation includes (1) where the description of execution tracing quality could be placed in the existing models and (2) what methods the complete frameworks offer to describe execution tracing quality including the reflection of subjective uncertainty. Nevertheless, the property illustrating execution tracing quality also needs to be able to express the quality of execution tracing as a standalone model without the frameworks presented.

3.1. ISO/IEC 9126 Framework

The standard allows adaptations of the software product quality model defined in the scope of the ISO/IEC 9126 standard family. The model definition in ISO/IEC 9126-1:2001 is superseded by ISO/IEC 25010:2011 but the model and the quality metrics are not superseded until ISO/IEC 25022 and 25023 are issued. Therefore we include this model in our investigation.

Following the concepts and terminology of the present software product quality framework the following steps are necessary for extension:

- Defining which characteristics and sub-characteristics can locate the execution tracing related quality description.
- Defining one or more internal and external metrics related to the quality property of execution tracing. The internal and external metrics have to correlate and the internal metrics need to have predictive validity towards the external metrics.

Extension Method

Execution tracing quality significantly influences the effort needed for error analysis. This identifies by its nature a property which belongs to maintainability or any of its sub-characteristics.

The high-level characteristic maintainability comprises of five sub-characteristics: (1) analysability, (2) changeability, (3) stability, (4) testability, (5) maintainability compliance. With regard to the goal of execution tracing the sub-characteristic analysability offers a logical point to link to because it encompasses all metrics which describe how the software or its behaviour can be analysed.

After finding the location in the hierarchy, the metrics need to be defined. As the description of execution tracing quality needs to be able to describe the quality of execution tracing as a standalone model, it is not recommended to define more metrics because it would create a dependency on the ISO product quality framework. If a new metric is introduced, the execution tracing quality model can easily be linked to the ISO framework without developing dependencies on it. For this reason we define an internal metric and an external metric keeping the naming conventions of the standard: (1) Internal Execution Tracing Capability Metric and (2) External Execution Tracing Capability Metric.

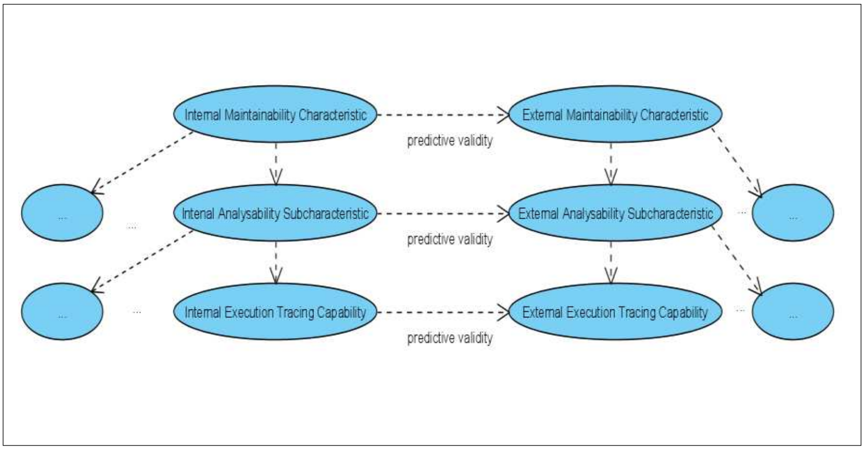


Figure 1
Extending ISO/IEC 9126 with Execution Tracing Capability

The definition of the metric also requires identification of the inputs and the method how the metric can be calculated from these inputs. The inputs of the metrics are called quality measure elements according to the terminology of the standard ISO/IEC 25021:2007.

Benefits

The expected benefit of this extension is to consider execution tracing quality when the complete software product quality is assessed. In addition, the subjective uncertainty of the inputs i.e. the quality measure elements of the metrics, can also be reflected by the mathematical calculations, which can involve fuzzy logic.

No detrimental effects of this extension on the framework are known. The standard also encourages tailoring the software product quality model to specific needs of projects. Consequently, the extension is in accordance with the philosophy of the ISO/IEC 9126 standard family.

Existing Extension Results

Carvallo and Franch [4] point out that software evaluation is necessary from a technical point of view but their examination shows that non-technical factors related to licensing and supplier characteristics are even more important in case of commercial off-the-shelf (COTS) products. The authors propose to extend the ISO/IEC 9126 standard family to include non-technical factors in a uniform way. In the proposal the authors keep the hierarchical structure of the standard and define three high-level characteristics: (1) supplier, (2) costs, (3) product, which they decompose in fifteen sub-characteristics, which on the third level of the hierarchy are even further decomposed resulting in more than two hundred non-technical quality properties. They validated the extension of the model on different projects in the telecommunication industry on which they provide a brief summary in [4].

3.2. ISO/IEC 25010 Framework

The software product quality framework defined in ISO/IEC 25010:2011 revised the software product quality framework of ISO/IEC 9126-1:2001 as described before. The new standard kept the philosophy of the previous model. The changes in the model hierarchy do not affect the node analysability below maintainability. Thus, the extension point does not change in comparison to the ISO/IEC 9126-1 framework.

Nevertheless, combination of the internal and external software product quality models in ISO/IEC 25010:2011 needs to be considered. As ISO/IEC 25022 and 25023 are not issued to supersede ISO/IEC 9126-2:2003 and 9126-3:2003, the separation of internal and external quality views is also a viable option.

Extension Method

The extension possibilities described for ISO/IEC 9126-1:2001 can be used with the revised software product quality model this new standard introduced. The measures Internal Execution Tracing Capability and External Execution Tracing Capability were merged into a single Execution Tracing Capability measure to comply with the combined internal and external model and consequently it possesses both internal and external quality measure elements.

Special attention needs to be paid to the definition of the inputs of execution tracing quality and the description of the computation by which the quality of execution tracing can be computed. Definitions of new quality measures and quality measure elements are formalised and defined in the standard ISO/IEC 25021:2007.

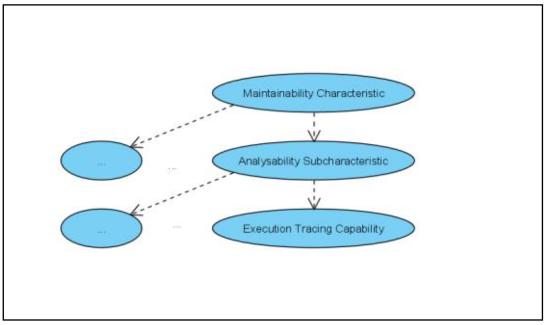


Figure 2
Extending ISO/IEC 25010 with Execution Tracing Capability

Benefits

As with the previously discussed ISO/IEC 9126-1:2001 extension, the primarily expected benefit is to consider execution tracing quality when the complete software product quality is assessed. The subjective uncertainty of the quality measure elements of the defined quality measures can also be reflected by mathematical calculations including fuzzy logic.

In comparison to extending the ISO/IEC 9126-1:2001 framework, a further advantage is to reduce the dependency on the ISO-framework to one measure which needs to be linked to it, thus supporting the standalone application of the quality model describing execution tracing.

No detrimental effects of this extension on the framework are known. The standard also declares that tailoring the software product quality model of ISO/IEC 25010 to specific needs of projects is a must i.e. tailoring is more emphasised in the revised standard than in its predecessor. Consequently, the extension is in accordance with the philosophy of the standards.

Existing Extension Results

No extension attempts of ISO/IEC 25010 were found in the literature.

3.3. Dromey's Framework

For each section the terminology of the model in question is used. Dromey's model applies the word attribute in a different way to the ISO/IEC 9126 and ISO/IEC 25000 standard families.

Dromey handles three primary sets of entities in his framework without introducing hierarchies. The relationships of these sets depict the quality requirements and the criteria for assessment. The set of high-level quality attributes contains maintainability which definitely illustrates the category to which execution tracing quality needs to be assigned. Therefore the set of high-level quality attributes need to undergo no changes. Consequently, extension possibilities for sets of quality-carrying properties and structural forms need to be examined.

Extension Method

Because all the structural forms define programming language-level constructs in the original description, higher-level structural forms are also necessary to include entities on component-level or application-level. New quality-carrying properties need to be introduced in the framework to describe the input variables of execution tracing in a binary manner to show whether the property is present in the application under investigation or not.

Execution tracing related quality-carrying properties can be linked to the new structural forms and to the high-level attribute maintainability in order to establish relationships. Then following the bottom-up approach introduced in the model description, the optimal relationships for each structural form need to be defined which guarantee the good quality; moreover, to support the top-down approach the optimal relationships need to be defined between the quality attributes and the quality-carrying properties. These profiles give a measure that can be compared to the actual software under investigation to diagnose quality defects or to set quality targets.

The original definition of the framework only considers the procedural programming paradigm, which has to be kept otherwise the present model needs to be reworked significantly to create new quality-carrying properties. The model's basic principles also facilitate the accommodation to other programming paradigms with the introduction of new quality-carrying properties and structural forms to define new relationships.

Benefits

As with the ISO/IEC 9126-1:2001 extension, the primarily expected benefit is to consider execution tracing quality when the complete software product quality is assessed.

As detrimental effect the high number of new relationships between the necessary quality-carrying properties and structural form needs to be mentioned, if not only programming language level assessment is necessary. On the other hand, the model supports the procedural programming paradigm only, so extension to further programming paradigms would implicate additional quality-carrying properties which would result in additional relationships between the quality attributes, structural forms and quality-carrying properties. The high number of possible relationships which should be processed during quality assessment can make the model unmanageable.

Existing Extension Results

No extension attempts of Dromey's framework were found in the literature.

4 Discussion

The frameworks investigated in the previous section allow extensions to include execution tracing quality but their implementations differ significantly.

Dromey's model only describes code-level constructs and their quality considering the procedural programming paradigm. The principle of the model, however, can also be applied for higher-level constructs and additional programming paradigms. From the point of view of execution tracing, procedural programming does not cause difficulties although the usability of the model would significantly be reduced if no other programming paradigms could be represented. To encompass additional programming paradigms and higher-level artefacts, Dromey's model requires considerable amounts of new quality-carrying properties and new structural forms. The high number of elements in both sets enhances the number of combinations through which relationships need to be expressed. Consequently, the execution tracing quality can be described at the cost of introducing more complexity in the model. In addition, the direct assignment of binary quality-carrying properties to high-level attributes leaves no room to uncertainty computations.

In contrast, ISO/IEC 9126 and 25010 offer an extension possibility and a sub-characteristic to which the description of execution tracing can be linked: maintainability and its analysability sub-characteristic. Linking is simple and requires considerably less effort than incorporating the illustrated changes in Dromey's model. Moreover, the quality measure or metric definitions complying with the standards allow the use of mathematical functions, by which subjective uncertainty computation can also be implemented.

If execution tracing quality were to be described by means of Dromey's framework, then it could not be used as an independent model because the framework requires a specific implementation. On the contrary, linking the description of execution tracing quality to the ISO/IEC software product quality frameworks facilitates its existence as an independent model.

ISO/IEC product quality models are more widespread than Dromey's model and they are known to a larger audience, as evidenced by the high number of publications relating to these standards, moreover by the models based on the ISO/IEC framework presented in the previous chapter. In addition, execution tracing quality can be encompassed with significantly less effort in the ISO/IEC standards than in Dromey's model.

Conclusion

Execution tracing is an important property that needs to be considered in quality frameworks to truly reflect the overall view of software product quality. Dromey's model allows extensions to include execution tracing quality although it requires significant changes in the present model. The model's philosophy does not support mathematical operations on quality-carrying properties and, therefore implementing subjective uncertainty computations is infeasible at present.

Software product quality frameworks of the ISO/IEC standards allow extensions and have a defined method to do so. Moreover, they also offer a natural linking point for execution tracing quality with the analysability sub-characteristic of maintainability. They can also allow mathematical computations that make the implementation of subjective uncertainty computations possible.

In conclusion, execution tracing quality should be linked to the software product quality framework of the ISO/IEC 25010 standard with the observation of the rules for defining quality measures and quality measure elements. This would facilitate the consideration of execution tracing quality when the whole software product quality is assessed; furthermore, it would ensure a framework for incorporating the impacts of subjective uncertainty resulting from the quality measurement process.

In summary, the findings of the paper include: (1) execution tracing quality should be reflected by the software product quality frameworks, (2) the current software product quality frameworks offer the possibility for extension but with significantly different efforts, (3) the ISO/IEC software product quality frameworks facilitate mathematical computations for defining measures or metrics, which allow to capture and implement subjective uncertainty and (4) metrics and a measure for execution tracing quality are outlined for the ISO/IEC 9126-1 and the ISO/IEC 25010 software product quality frameworks.

References

[1] Banker, R. D., and S. Slaughter. "A Field Study of Scale Economies in Software Maintenance." *Management Science* 43, No. 12 (1997): 1709-1725

[2] Boehm, B. W., J. R. Brown, and M. Lipow. "Quantitative Evaluation of Software Quality." *Proceedings of the 2nd International Conference on Software Engineering.* 1976

[3] Buch, I., and R. Park. "Improve Debugging and Performance Tuning with ETW." *MSDN Magazine, [Online], [Accessed: 01.01.2012], Avaliable from: http://msdn.microsoft.com/en-us/magazine/cc163437.aspx*, 2007

[4] Carvallo, J., and X. Franch. "Extending the ISO/IEC 9126-1 Quality Model with Non-Technical Factors for COTS Components Selection." *Proceedings of the 2006 International Workshop on Software Quality.* 2006. 9-14

[5] Csallner, C., and Y. Smaragdakis. "DSD-Crasher: A Hybrid Analysis Tool for Bug Finding." *ACM Transactions on Software Engineering and Methodology (TOSEM).* 2008

[6] Csernyak, L., G. Horvath, J. Horvath, S. Lorincz, S Molnar, and A Stern. *Matematika a Kozgazdasagi Alapkepzes Szamara (Translated Title: Mathematics for the Bachelor Curricula in Economics, Probability Theory).* Budapest: Nemzeti Tankonyvkiado, 2007

[7] Dromey, R. "A Model for Software Product Quality." *IEEE Transactions on Software Engineering.* 1995, 146-162

[8] Ernst, M. D. "Static and Dynamic Analysis: Synergy and Duality." *In Proceedings ICSE Workshop on Dynamic Analysis*, 2003: 24-27

[9] Galli, Tamas, Francisco Chiclana, Jenny Carter, and Helge Janicke. "Modelling Execution Tracing Quality by Means of Type-1 Fuzzy Logic." *Acta Polytechnica Hungarica*, Vol. 10, No. 8, 2013: 49-67

[10] Godefroid, P., and N. Nagappan. *Concurrency at Microsoft – An Exploratory Survey, [Online], 2007, [Accessed: 24.01.2012.], Available from: http://research.microsoft.com/en-us/um/people/pg/public_psfiles/ec2.pdf.* Microsoft Research

[11] Hovemeyer, D., and W. Pugh. "Finding Bugs Is Easy." *OOPSLA: Object-Oriented Programming, Systems, Languages & Applications, [Online], [Accessed: 14.02.2012], Available from: http://www.cs.nyu.edu/~lharris/papers/findbugsPaper.pdf.* 2004

[12] IBM. "Understanding Execution Traces, [Online], [Accessed: 05.02.2013], Available from: http://pic.dhe.ibm.com/infocenter/brdotnet/v7r1/index.jsp?topic=%2Fcom.ibm.websphere.ilog.brdotnet.doc%2FContent%2FBusiness_Rules%2FDocumentation%2F_pubskel%2FRules_for_DotN."

[13] International Organization for Sandardization. "ISO/IEC 25010:2011, Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models." 2011

[14] International Organization for Sandardization. "ISO/IEC 9126-1:2001, Software engineering -- Product quality -- Part 1: Quality model." 2001

[15] International Organization for Standardization. "ISO/IEC 14598:1999, Information technology -- Software product evaluation -- Part 1: General overview." 1999

[16] International Organization for Standardization. "ISO/IEC 25021:2007, Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality measure elements." 2007

[17] International Organization for Standardization. "ISO/IEC TR 9126-2:2003, Software engineering -- Product quality -- Part 2: External metrics." 2003

[18] International Organization for Standardization. "ISO/IEC TR 9126-3:2003, ." *Software engineering -- Product quality -- Part 3: Internal metrics*, 2003

[19] International Organization for Standardization. "ISO/IEC TR 9126-4:2004, Software engineering -- Product quality -- Part 4: Quality in use metrics." 2004

[20] Kahn, A. *Simulation of Message Passing Programs, [Online] http://may.cs.ucla.edu/projects/sesame/publications/sundeep_diss_html/node43.html, University of California.* 1997

[21] Karahasanovic, A., and R. Thomas. "Difficulties Experienced by Students in Maintaining Object-oriented Systems: An Empirical Study." *Proceedings of the 9th Australasian Conference on Computing Education*, 2007: 81-87

[22] Kim, C., and K. Lee. "Software Quality Model for Consumer Electronics Product." *Proceedings of the 9th International Conference on Quality Software.* 2008, 390-395

[23] Kitchenham, B., and S. Pfleeger. "Software Quality: the Elusive Target." (IEEE Software) 13, No. 1 (1996): 12-21

[24] Klir, G. J., U. H. St.Clair, and B. Yuan. *Fuzzy Set Theory Fundations and Applications.* Prentice Hall Ptr, 1997

[25] Koskinen, J. *Software Maintenance Costs.* [Online], 2010, [Accessed: 23.01.2012], Available from: http://users.jyu.fi/~koskinen/smcosts.htm

[26] Krishnan, M. S., T. Mukhopadhyay, and C. H. Kriebel. "A Decision Model for Software Maintenance." *Information Systems Research* 15, No. 4 (2004): 396-412

[27] Laddad, R. *AspectJ in Action.* Manning, MEAP, Second Edition, 2009

[28] Martin, A. "Debugger, Real-Time Trace, Logic Analyser, Long-Term Trace ETMv3, [Online], [Accessed: 05.02.2013], Available from: http://www.lauterbach.com/publications/long_term_trace_etmv3.pdf, Lauterbach GmbH." 2009

[29] McCall, J. A., P. K. Richards, and G. F. Walters. *Factors in Software Quality, Concept and Definitions of Software Quality.* Technical Report, RADC-TR-77-369, Rome Air Development Center, Air Force System Command, Griffis Air Force Base, New York 13441, [Online], [Accessed: 21.10.2011], Available from: http://handle.dtic.mil/100.2/ADA049014, 1977

[30] Microsoft Co. "Tracing WMI Activity (Windows), [Online], [Accessed: 05.02.2013], Available from: http://msdn.microsoft.com/en-us/library/windows/desktop/aa826686%28v=vs.85%29.aspx." 2012

[31] Panda, D., R. Rahman, and D. Lane. *EJB 3 in Action.* Manning Publications Co., 2007

[32] Qu, D., A. Roychoudhury, Z. Lang, and K. Vaswani. *Darwin: An Approach for Debugging Evolving Programs.* Microsoft Research, [Online], 2009, [Accessed: 24.01.2012], Availavle from: http://research.microsoft.com/apps/pubs/default.aspx?id=80898

[33] Research Triangle Institute. *RTI Project Number 7007.011, The Economic Impacts of Inadequate Infrastructure for Software Testing.* National Institute of Standards and Technology, U.S Department of Commerce, Technology Administration, National Institute of Standards and Technology, U.S Department of Commerce, Technology Administration, [Online], 2002, [Accessed: 20.01.2012], Avaliable from: http://www.nist.gov/director/planning/upload/report02-3.pdf, 2002

[34] Saaty, T. L. *The Analytic Hierarchy Process.* New York: McGraw-Hill, 1980

[35] SAP. “Using the Technical Trace and Log, [Online], [Accessed: 05.02.2013], Available from: http://help.sap.com/saphelp_nwpi71/helpdata/en/3a/63e540aa827e7fe10000000a1550b0/content.htm.”

[36] Shi, Z. “Visualizing Execution Traces, Master Thesis.” *[Online], [Accessed: 17.05.2011], Available from: http://www.mcs.vuw.ac.nz/comp/graduates/archives/mcompsc/reports/2004/Zhenyu-Shi-final-report.pdf*, 2005

[37] Spillner, A., T. Linz, and H. Schaefer. *Software Testing Foundations.* Santa Barbara, CA: Rockey Nook Inc., 2007

[38] Spinczyk, O., D. Lehmann, and M. Urban. “AspectC++: an AOP Extension for C++.” *Software Developers Journal*, 2005: 68-74

[39] Uzelac, V., A. Milenkovic, M. Burtscher, and M. Milenkovic. “Real-time Unobtrusive Program Execution Trace Compression Using Branch Predictor Events.” *CASES 2010 Proceedings of the 2010 international conference on Compilers, Architectures and Synthesis for Embedded Systems, ISBN: 978-1-60558-903-9*, 2010

[40] Winterfeldt, D. *Spring by Example.* Spring Tutorial, [Online], [Accessed: 19.12.2012], Available from: http://www.springbyexample.org/examples/aspectj-ltw-spring-config.html

[41] Young, M. “Symbiosis of Static Analysis and Program Testing.” *In Proc. 6th International Conference on Fundamental Approaches to Software Engineering*, 2003: 1-5

A Content-based Image Retrieval System Based on Polar Raster Edge Sampling Signature

Santhosh P. Mathew[1], Valentina E. Balas[2], Zachariah K. P.[1], Philip Samuel[3]

[1]Department of Computer Science, Saintgits College of Engineering, Kerala, India; santhosh.mathew@saintgits.org; zacharia.kp@saintgits.org

[2]Department of Automatics and Applied Software, Aurel Vlaicu University of Arad, Romania; valentina.balas@uav.ro

[3]Department of Information Technology, Cochin University of Science and Technology, Cochin, Kerala, India; philips@cusat.ac.in

Abstract: Content Based Image Retrieval (CBIR) is used to effectively retrieve required images from fairly large databases. CBIR extracts images that are relevant to the given query image, based on the features extracted from the contents of the image. Most of the CBIR systems available in the literature are not rotation and scale invariant. Retrieval efficiency is also poor. In this paper, shape features are extracted from the database images and the same are polar raster scanned into specified intervals in both radius and angle, using the proposed Polar Raster Edge Sampling Signature (PRESS) algorithm. Counts of edge points lying in these bins are stored in the feature library. When a query image passed on to the system, the features are extracted in the similar fashion. Subsequently, similarity measure is performed between the query image features and the database image features based on Euclidian Distance similarity measure and the database images that are relevant to the given query image are retrieved. PRESS algorithm has been successfully implemented and tested in a CBIR System developed by us. This technique preserves rotation and scale invariance. It is evaluated by querying different images. The retrieval efficiency is also evaluated by determining precision-recall values for the retrieval results.

Keywords: image retrieval; polar raster edge sampling signature; image segmentation; feature extraction; K-means clustering; Canny algorithm

1 Introduction

Content based Image Retrieval system uses content for the retrieval process. It is not easy to develop such a retrieval system because of the complexity involved in real world images that contain complex objects and detailed color information [1].

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Some of the important indications that are used to extract information from the images are Color, Shape and texture. Color histograms are widely used in content based image retrieval [2]. Though color and texture contain key information, it is possible that different images with similar color histograms represent very different things. Hence shape-describing features are to be used in an efficient content-based image retrieval system. Although much research has taken place in connection with shape description to identify the right kind of shape feature, there are no direct answers yet [3]. Most of the traditional methods of image retrieval use some method of adding metadata. Captioning, Keywords or Descriptions are added to the images so that retrieval can be performed based on the annotation words.

Image retrieval has been an area of interest for researchers during the past few decades. Database Management and Computer Vision are two main research groups who study image retrieval from different viewpoints. Database Management group primarily looks at text-based approaches and the Computer Vision group looks at visual-based approaches [4]. There is a basic difference between content-based and text-based retrieval systems. Text-based systems are depended on human interaction. Humans normally use keywords, text descriptors and similar high level features to interpret images [5]. Manual textual annotation was used initially to retrieve images. But this was observed to be a very difficult task, primarily because of the limitation in the interpretation of what we see. This resulted in the image contents like color, shape and texture gaining greater importance [6]. Also, large amount of manual effort was required in developing the annotations and in addressing the differences in interpretation of image contents and the lack of uniformity of the keyword assignments among various indexers. The keyword annotation approach becomes impractical as the size of the image database increases. To overcome these difficulties another mechanism, Content Based Image Retrieval, is used [4].

One of the important basic features in content-based image retrieval is Shape. Shape based representations are broadly divided into two: Region based and Contour based [7]. Moment descriptors - Geometrical moments, Zernike moments and Legendre moments are normally used in the case of Region based representation [8], [9]. Contour based systems normally use the boundary of the objects for representation and retrieval. There are different effective algorithms to extract the edges. Since images are usually distinguishable by their contours, this approach gives better results [11].

CBIR systems retrieve images that match the extracted content features of the query image. Identifying a suitable feature set that would ensure retrieval efficiency has been a challenge in Content Based Image Retrieval. Rotation and Scale invariance too have not been addressed effectively. The proposed CBIR system retrieves a specified number of images that match the query image. The best match will be displayed first and it also preserves Rotation and Scale Invariance.

The remaining parts of the paper are arranged as follows: In Section 2, the background of the work is described. Section 3 briefs the Content based image retrieval with proposed Polar Shape Signature extraction, Feature calculation and Image retrieval process. Experimental results and analysis of the proposed technique are discussed in Section 4. Finally, concluding remarks are provided in Section 5.

2 Background of the Work

A brief discussion of some relevant works from the available literature in the area of shape based Content Based Image Retrieval is presented in this section.

Number of trademark images around the world has increased rapidly. Trademark image retrieval (TIR) has been researched upon to ensure that new trademarks are not a replica of the trademark images that are stored in the trademark registration system. Chia-Hung Wei *et al*. [8] suggested a content-based trademark retrieval system using feature sets that could describe global shapes and local aspects of the trademarks. Color, texture and shape information are the common image descriptors in content based image retrieval systems. Xiang-Yang, Wang *et al*. [10] proposed a new combination color image retrieval scheme, making use of all these descriptors.

Shape is an important visual feature in CBIR. Shape descriptors are of two types i.e., contour-based and region-based. Contour-based shape descriptors use only the boundary information by neglecting the content within the shape, while region-based shape descriptors concentrate on the contents within the shape.

Zang and Melissa [12] proposed a method for shape representation and retrieval, taking into account the regional properties. They proposed an idea that is similar to normal raster sampling. But, instead of using the normal square grid on a shape image, a circular sampling of concentric circles and radial lines is used at the center of the shape. The binary value of the shape is sampled at the intersections of the circles and radial lines. Santhosh P. Mathew et al [13] have also tried to implement similar concept for a CBIR system. The drawback of this technique is that the shape points on the intersection are only considered. PRESS proposed by us forms an effective shape signature by taking into account all the edge points, by collecting them in specified number of radial and angular bins. Scaling and rotational invariance are also preserved in the CBIR system implemented using PRESS.

3 Content-based Image Retrieval Based on Polar Raster Edge Sampling Signature (PRESS)

The proposed effective Content Based Image Retrieval system comprises of the processes such as Image Segmentation, Feature Extraction and Image retrieval based on the query image. Here the CBIR is based on the extracted shape signature. The shape signature extraction is as follows. Initially, the image is segmented based on color, using the K – Means clustering algorithm. Canny algorithm is employed to detect the edges. The strong edges and the connected edges are identified using various techniques like double thresholding and edge tracking. The edge data in the boundary and the region are polar raster scanned in both radius and angle. Numbers of edge points identified are stored in the feature library for all the database images. When an image is queried, the system extracts shape feature for the image in the same way and then computes the similarity measure between the features of the query image and the feature existing in the feature database based on the Euclidean Distance method. Minimum distance indicates the closest match and specified number of best matched images are extracted. The proposed method is detailed in the following sections.

3.1 Extraction of Shape Features

Initially, the image in RGB color space is converted to gray scale. We make use of the K-means clustering algorithm to segment image for further processing [14] [15]. After the k means algorithm is applied, the canny algorithm is used for the detection of different edges present in all the clustered sets of the image. Then we get the different shapes that are present in the image and the features of the shaped content are extracted by employing the proposed PRESS (Polar Raster Edge Sampling Signature) algorithm. Edge data of the shape is polar raster scanned or binned into 10 intervals in both radius and angle. Counts of the edge points lying in these bins are found and stored in two vectors **r** and **t**. The entire shape feature extraction process can be represented in the following pseudo code.

In this pseudo code the proposed PRESS algorithm is used for extracting the shape features of the image. The approach is that the final edge components extracted are polar raster scanned or binned into ten intervals in the Radius and the Angle. Count of edge points lying in these bins are found and stored as two vectors **r** and **t.** It is further normalized for sum of counts. The same process is repeated for all the images in the database and the r and t values are saved in the feature database.

$BW \leftarrow edge(mask,'canny');$ // Finding the edges of the image.

$[\mathrm{imx}, \mathrm{imy}] \leftarrow size(BW);$

$B \leftarrow conv2(BW,msk);$ //Smoothing the image

$[x,y] \leftarrow find(B == 1);$

$p = [x, y];$

$[qr, qt] = polartransform(p);$

get count of edge points;

store radius bin count in vector r;

store angle bin count in vector t; //PRESS is used to extract shape features

normalize for sum of counts;

repeat for all images in the DB;

save them in feature DB;

Figure 1

Pseudo code for shape feature extraction process

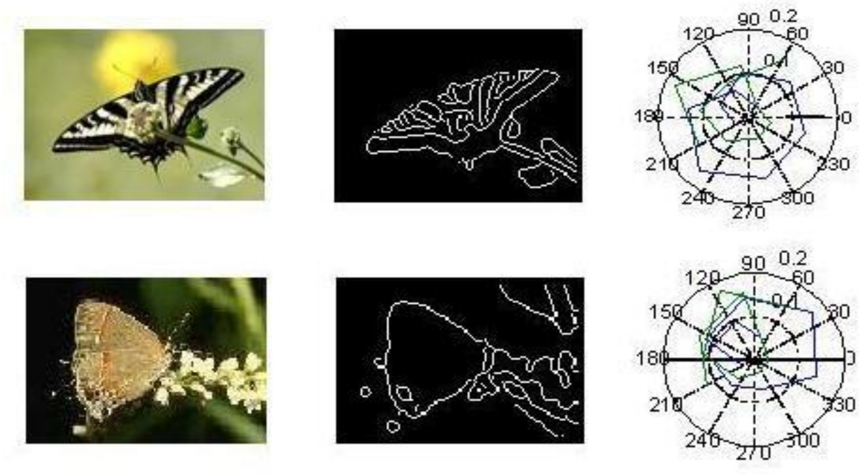


Figure 2

PRESS algorithm applied on the edges

3.2 Retrieval Process

Image features are extracted first and the appropriate match of the query image is retrieved from the database. The similarity measure between shape feature of query image Q and the features of each image in the database is calculated using Euclidean Distance (ED). The feature of query image is represented as $Q_f = [r, t]$.

$$d_i = \left\| Q_f - F_i \right\| \tag{1}$$

ED between r vectors of query and DB is found. Minimum ED between t vector of query image and circularly shifted t vector of DB image is found to account for possible rotation of query image. RMS of the two EDs is used as the distance measure between shape features. Minimum distance is the closest match. On the basis of the above similarity measure, a specified number of best matched images from the DB that are similar to the query image are retrieved.

4 Results and Discussion

We have implemented this CBIR system in MATLAB 7.10. Test images were taken from the database generated by Wang containing many images stored in the JPEG format [16]. The system extracts the shape feature for the query image and then computes the similarity measure between that and the shape features of all images existing in the database. This results in a predefined number of database images similar to the query image being retrieved.

The query image is first preprocessed to normalize the intensity levels of the input image. The output objects extracted from the input query image after the preprocessing stage is given in Row 2 of Fig. 3. Since the mean filter acts only on a single color channel, the RGB image is first converted to gray scale image, using Craig's formula. K-means clustering algorithm (k = 4) is used on the gray scale image for the segmentation of the image. Before applying the K-means clustering, the image in the form of 2D vector is rescaled to a 1D vector. The output image of the clustering process is shown in the Row 3 of Fig. 3. After the clustering process using K-Means algorithm, the 1D image is again converted back into 2D image format. Subsequently, Canny algorithm is applied to detect the edges. Edges are present in all the clustered sets of the image and Canny algorithm identifies them. The various steps involved are smoothing, finding gradients, non-maximum suppression, double thresholding and edge tracking by hysteresis. The detected edges are shown in the Row 4 of Fig. 4.

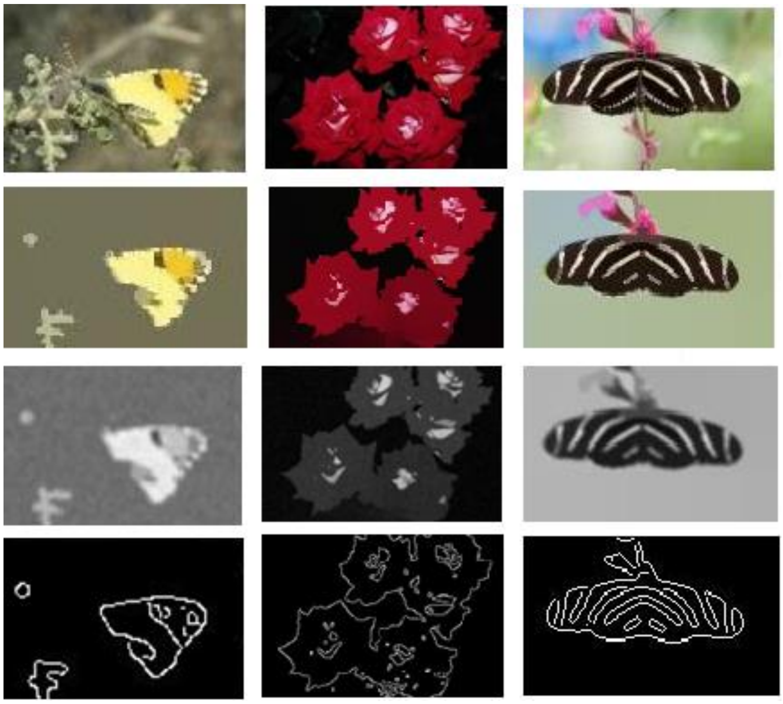


Figure 3

Row 1 – Original images, Row 2 – Objects extracted from the original images

Row 3 – Clustered gray scale images, Row 4 – Edges detected by Canny algorithm

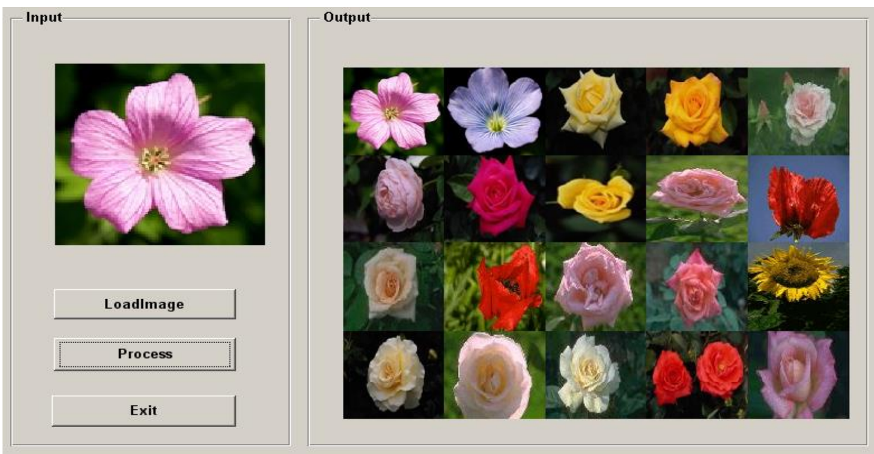


Figure 4

Sample output - Query Image and Retrieved images

Suppressing and smoothing are done on these edges to eliminate unused connected components and those edges that are not connected to strong ones. Thus the clear edges of all the segments are obtained. PRESS algorithm proposed by us is employed on the final edge components to extract the shape features. Query image shape features are matched with shape features in the DB by using RMS values of the EDs of r vectors and t vectors. Minimum distance is the closest match.

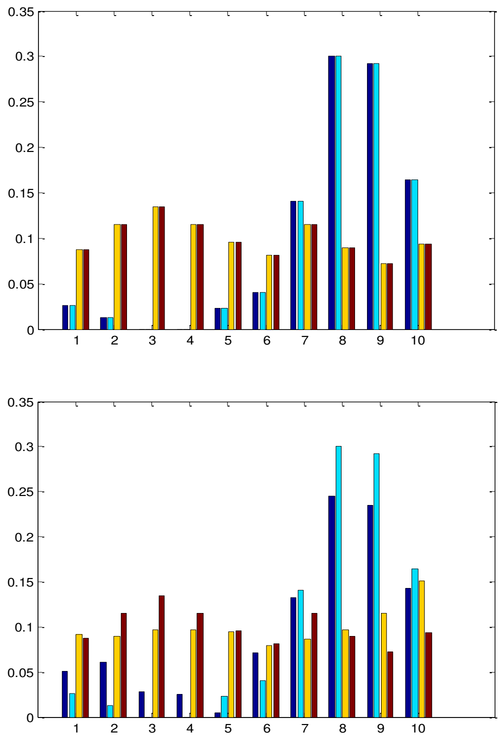


Figure 5

Bar graph indicating the exact match of the Radius bin count (rr), Theta bin count (rt), with that of the query image (qr, qt) – for the best match. Bar graph for the next best match is also shown along side.

Thus the images corresponding to the input query image are retrieved from the database. Some of the retrieved images that correspond to the input query image are shown in Fig. 4. Bar graph of the r and t values of the query and retrieved images are given in Fig. 5. The PRESS technique successfully matches the query image exactly with that in the database and can in no way miss it. Rotational Invariance and Scale Invariance were also tested with rotated and scaled images. At all instances, all the rotated and scaled images were retrieved without any problems. Sample outputs for Rotational and Scale Invariance are given in Fig. 6 and Fig. 7.

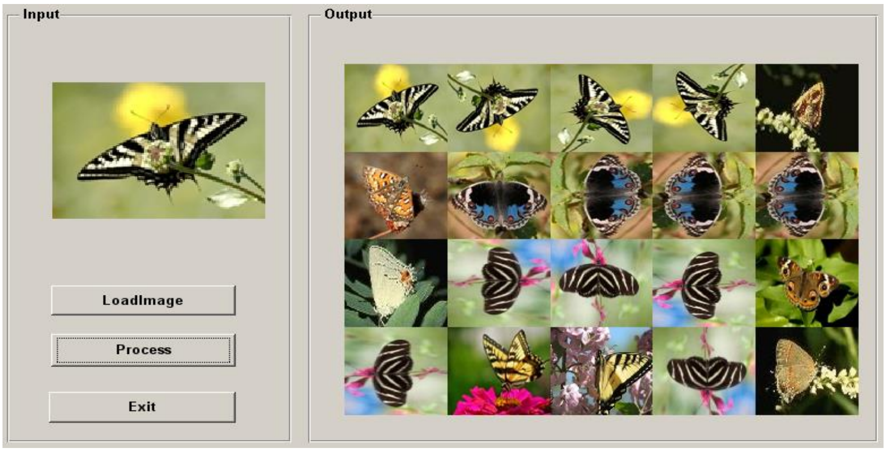


Figure 6

Rotational Invariance - Query Image and Retrieved images - that were rotated in 90, 180, 270 and 360 degrees

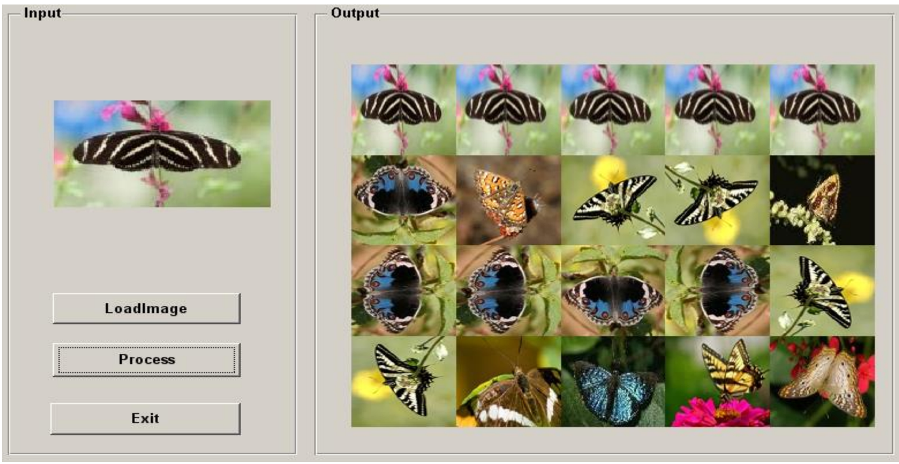


Figure 7

Scale Invariance – Query image (143×82) and Retrieved images – that were scaled to 157×90, 172×98, 200×115 and 129×74

Comparative Analysis

Precision and recall values are used to analyze the performance. Precision is the fraction of retrieved images that are relevant to the query image. Recall is the fraction of relevant images that are retrieved from the database. We obtain a measure of relevance, from these values.

$$precision = \frac{\text{Count of retrieved images relevant to the query image}}{\text{Total count of images retrieved}} \quad (2)$$

$$recall = \frac{\text{Count of retrieved images relevant to the query image}}{\text{Total count of relevant images in the database}} \quad (3)$$

Using Eq. (2) and Eq. (3), the precision and recall values for the query image are calculated for the proposed method and for some of the existing methods. The values obtained from the calculation are given in Table I. The precision values are compared with some of the existing methods and plotted in Fig. 8.

Table 1

Precision (P) and Recall (R) Statistics for the proposed PRESS (Polar Raster Edge Sampling Signature) Method on the best 15, 20 and 40 retrieved images

Query	N = 15		N = 20		N = 40	
	P	R	P	R	P	R
1	0.80	0.28	0.75	0.35	0.73	0.67
2	1	0.35	0.95	0.44	0.85	0.79
3	1	0.35	0.95	0.44	0.83	0.76
4	0.80	0.28	0.75	0.35	0.70	0.65
5	1	0.35	0.95	0.44	0.93	0.86
6	0.93	0.33	0.85	0.40	0.80	0.74
7	0.80	0.28	0.75	0.35	0.73	0.67
8	0.80	0.28	0.75	0.35	0.70	0.65
9	0.93	0.33	0.80	0.37	0.78	0.72
10	0.80	0.28	0.75	0.35	0.70	0.65

Fig. 8 shows the graph for the comparative analysis between the proposed Polar Raster Edge Sampling Signature (PRESS) method and existing methods like CSS (Curvature Scale Space), ZMD (Zernike Moment Descriptor), GFD (Geometric Fourier Descriptor) and MSFD (Multiple Shape Feature Descriptor) on 5 queries for a set of best 20 images retrieved. The data for other methods are from [10], [7].

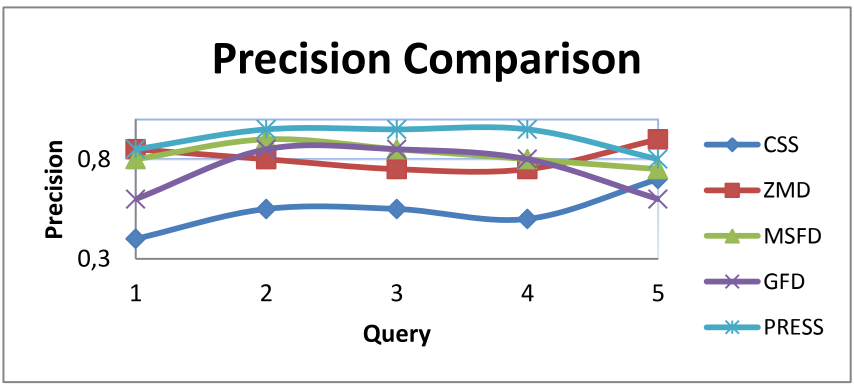


Figure 8

Comparison of Precision

Conclusion

We have proposed a CBIR system based on Shape Signature to retrieve relevant matches from the database for the query image. When an image is queried, shape feature for the image is extracted and then the similarity measure between the features of the query image and the feature existing in the feature database are calculated based on the Euclidean Distance method. Image segmentation is done using K-means clustering algorithm, which groups the image pixels under color. Canny algorithm is then used to extract edges. Proposed PRESS algorithm has been successfully applied on the edge components to extract shape features. The same is tested in a CBIR System developed by us. This technique preserves rotation and scale invariance. It is evaluated by querying different images. Precision-Recall values are used for the comparison of retrieval results. The implementation results illustrates that this novel image retrieval process effectively retrieves the images that are very close to the query image from the database. Precision - Recall table and Precision comparison plot with existing CBIR techniques prove the effectiveness of the system.

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain: Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, Dec. 2000, pp. 1349-1379

[2] S. Belongie, C. Carson, H. Greenspan and J. Malik: Color and Texture-Based Segmentation using EM and its Application to Content-based Image Retrieval, In Proc. of the Sixth International Conference on Computer Vision, Vol. 10, Jan. 1998, pp. 675-682

[3] J. Laaksonen, E. Oja and S. Brandt: Statistical Shape Features in Content-based Image Retrieval, In Proc. of the 15th International Conference on Pattern Recognition, Vol. 2, Sep. 2000, pp. 1062-1065

[4] Y. Rui and T. S. Huang: Image Retrieval: Current Techniques, Promising Directions, and Open Issues, Journal of Visual Communication and Image Representation, Vol. 10, Jan. 1999, pp. 39-62

[5] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma: A Survey of Content-based Image Retrieval with High-level Semantics, Journal of Pattern Recognition, Vol. 40, No. 1, Jan. 2007, pp. 262-282

[6] A. Jain, R. Muthuganapathy and K. Ramani: Content-based Image Retrieval Using Shape and Depth from an Engineering Database, In Proc. of the Third International Conference on Advances in Visual Computing, Vol. 2, 2007, pp. 255-264

[7] D. Zhang and G. Lu: A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-based Image Retrieval, Journal of Visual

Communication and Image Representation, Vol. 14, No. 1, Mar. 2003, pp. 39-57

[8] C.-H. Wei, Y. Li, W. Y. Chau and C.-T. Li: Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure, Journal of Pattern Recognition, Vol. 42, No. 3, Mar. 2009, pp. 386-394

[9] D. Zhang and G. Lu: Shape-based Image Retrieval using Generic Fourier Descriptor, Journal of Signal Processing: Image Communication, Vol. 17, No. 10, Nov. 2002, pp. 825-848

[10] X.-Y. Wang, Y.-J. Yu and H.-Y. Yang: An Effective Image Retrieval Scheme using Color, Texture and Shape Features, Journal of Computer Standards & Interfaces, Vol. 33, Mar. 2010, pp. 59-68

[11] Ch. S. Rao, S. S. Kumar and B. C. Mohan: Content-based Image Retrieval Using Exact Lengendre Moments and Support Vector Machine, The International Journal of Multimedia & Its Applications, Vol. 2, No. 2, May 2010, pp. 69- 79

[12] D. Zhang and M. C. Yi Lim: An Efficient and Robust Technique for Region-based Shape Representation and Retrieval, 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007) pp. 801-806

[13] S. P. Mathew and P. Samuel: A Novel Image Retrieval System using an Effective Region-based Shape Representation Technique, International Journal of Image Processing (IJIP), Vol. 4, No. 5, Dec. 2010, pp. 509-517

[14] K. S. Ravichandran and B. Ananthi, Color Skin Segmentation Using K-Means Cluster, International Journal of Computational and Applied Mathematics, Vol. 4, No. 2, 2009, pp. 153-157

[15] Szabolcs Sergyán: A New Approach of Face Detection-based Classification of Image Databases, Acta Polytechnica Hungarica, Journal of Applied Sciences, Vol. 6, No. 1, 2009, pp. 175-184

[16] R. Datta, D. Joshi, J. Li and J. Z. Wang: Image Retrieval: Ideas, Influences, and Trends of the New Age, ACM Computing Surveys, Vol. 40, No. 2, April 2008, Article 5

Thermal Dewatering of Waste Sludge in an Agitated Drum Dryer

Tibor Poós[1], Mária Örvös[1], Miklós Horváth[2]

[1]Department of Building Services and Process Engineering, Budapest University of Technology and Economics, Bertalan L. u. 4-6, H-1111 Budapest, Hungary; E-mail: poos@vegyelgep.bme.hu; orvos@vegyelgep.bme.hu

[2]Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház u. 8, H-1081 Budapest, Hungary; E-mail: horvath.miklos@bgk.uni-obuda.hu

Abstract: This study presents experiences of drying waste sludge in an agitated drum dryer. Measurements were performed on an agitated drum dryer apparatus equipped with up-to-date instrumentation. As waste sludge is prone to adhesion and balling, it is important to identify the conditions appropriate for drying. Different cases were examined; during measurements, the rotational speed of the agitator, the quantity and moisture content of the sludge fed in as well as the temperature of the wall and the drying gas were changed. The measurement data were used to represent the temperatures characteristic of drying and the mass changes of the product. Volumetric mass transfer coefficients were determined from the measurements and a modified Sherwood number was derived from them. The dimensionless Sh'-Re' equation enables to dimension agitated drum dryers.

Keywords: agitated drum dryer; Sherwood number; volumetric mass transfer coefficient; waste sludge drying

1 Introduction

Activated sludge treatment of waste water is one of the most wide-spread biological waste water treatment procedures. The apparatus used is a floating bed bioreactor fed on an on-going basis. The biomass substantially remains in the reactor: although some sludge constantly leaves the basin, most of it is fed back in order to ensure an appropriate sludge concentration [1]. The sludge increment is called excess sludge, transport and deposition of which is costly; however, after proper dewatering it can be burnt and thereby it can serve as a basic material for heat and power generation. In order for burning at an adequate degree of efficiency, the maximum moisture content of the product can be 45%, to be achieved by drying. In recent decades, a number of authors have discussed the drying of waste sludge [2, 3]. *Chen et al.* collected and presented devices suitable for dewatering and drying waste sludge in their work [4]. *Stuart et al.* summarized the benefits and drawbacks of various dryers in their work [5]. Several publications presented the application of recycled cooking oil into a solid sludge as a drying technology [6, 7, 8]. In these cases, used oil was heated up and waste sludge was directly added. By using this procedure, the product can be dried quickly and very efficiently. Drying is a highly energy-intensive process, but waste heat utilizable for this purpose can be found in infinity of cases. In absence of waste heat, an alternative solution is presented in the publications by *Slim et al.*, *Seginer et al.* and *Leon* [9, 10, 11], where drying by solar energy was highlighted. In this procedure, the constant drying parameters cannot be ensured continuously and the quantity of fluid evaporated highly depends on environmental and climatic conditions.

Due to its unfavourable physical properties, waste sludge is difficult to dry in traditional dryers. These materials stick to the wall of the dryer because of their high moisture content and they tend to cake at lower levels of moisture content. The agitated drum dryer with wall heating applied by us reduces unfavourable phenomena occurring in the case of sludge drying; drying tests were performed by in our measurements sludge from a waste water treatment plant [12]. The drying of waste sludge was tested in a similar device by *Ferrasse et al.* [13]. The initial moisture content of the waste sludge examined by us was approx. 80%, while the final moisture content was required to be 45%. We also performed measurements where the initial moisture content was reduced by backmixing dry matter, in a similar fashion to *Léonard et al.* [14]. Drying gas speed was selected for our measurements so that the dried product should not be carried away – or only negligibly – by the gas flow. In order to specify the geometry of the dryer (diameter, length), mass transfer properties are required, which can be determined by experiments. By reason of the characteristic of a mixed conglomerate, the surface of heat and mass transfer is not known, therefore a volumetric evaporation coefficient widely accepted in the literature was specified. In the course of their calculations, *Léonard et al.* referred the contact area between the gas and the product to the initial external area [15].

The aim of the article is to create a modified Sherwood number was derived from volumetric evaporation coefficients and plotted against a modified Reynolds number providing a correlation suitable for dimensioning dryers.

2 Experimantal Apparatus, Materials and Methods

2.1 Experimental Apparatus

Experimental tests were performed on the agitated drum dryer shown in *Figure 1.* Within the dryer, heat transmission is not only effected by convection through the drying gas but also by conduction from the mantle of the drum. Drying of the paste-like waste sludge is considerably accelerated by the application of electric wall heating.

The measurement station was designed for drying paste-like and granular products by simultaneous conductive and convective heat transfer. The measurement station, equipped with up-to-date instrumentation, is also suitable for studying the vary of simultaneous heat and mass transfer. Drum dryer (D-102-01) is the central component of the mechanical system. The dryer of intermittent operation consists of a 765 mm long, 250 mm wide and 275 mm high U-profile drum covered by a plain plate. The drying gas is fed in axially and is removed through the pipe end on the plain cover of the equipment. The lower, semi-cylindrical surface of the drum is equipped with controllable electric wall heating. The paste-like product in the dryer is moved by a special agitator scratcher driven by an electric motor (M-102-03). The rotational speed of the agitator can be adjusted by a continuous frequency converter within a broad range (0÷95 1/min) (SIC-102-07). The drum dryer, as well as the electric motor and driving gear to drive the agitator are mounted on the same framework structure. The apparatus stands on a scale (S-102-02), of Sartorius IS 300 IGG-H type, with a maximum capacity of 300 kg and accuracy of 2 g.

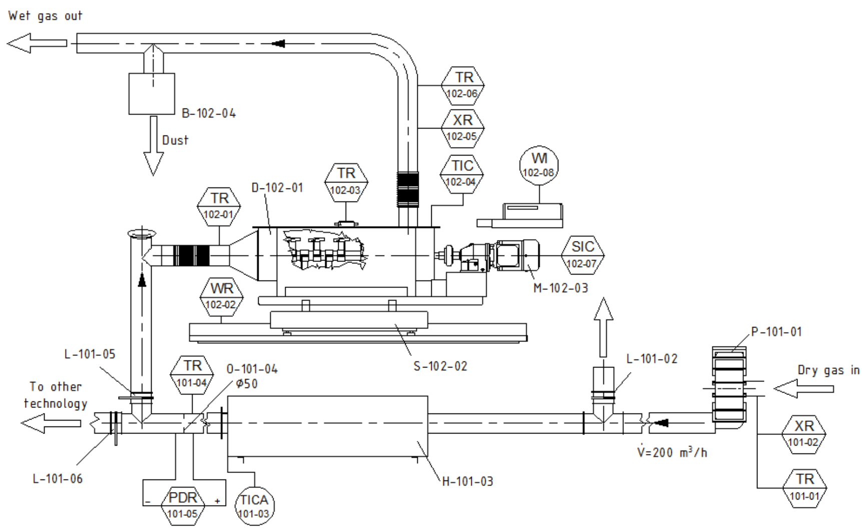


Figure 1

Instrumentation of agitated drum dryer measurement station

D-102-01: Drum dryer; M-102-03: Motor with driving gear; H-101-03: Electric air heater; P-101-01: Air blower; S-102-02: Scale; L-101-02: Volumetric flow rate controller; B-102-04: Dust separator

The drying gas is heated by an electric air heater (H-101-03) equipped with a temperature controller (TICA-101-03). The volumetric flow of the drying air, moved by an Effepizetta SCL-SH65 fan (P-101-01), can be modified by a gap controller (L-101-02).

In order to determine the volumetric flow of the drying gas, the pressure drop of the drying gas at a standard measuring orifice and its temperature were measured. Measurements included the temperature of the gas entering (TR-102-01) and exiting (TR-102-06) the drum, as well as the wet bulb temperature of the exiting gas (XR-102-05), wherefrom the exit moisture content of the gas can be determined. The surface temperature of the product to be dried was measured in the middle of the dryer longitudinally by an infrared thermometer (TR-102-03). Decrease in the wet product mass in the dryer (WR-102-02) was continuously measured and recorded by a data logger. The moisture content of the incoming gas (XR-101-02) and the environmental temperature (TR-101-01) were also measured. The dryer drum can be heated electrically from the mantle side, its temperature can be kept at a given value by an automatic controller. Wall temperature was measured at 2/3 of the length of the drum (TIC-102-04).

In order to record the parameters of drying, sensors were connected to an Almemo 2590-9 type Ahlborn data logger instrument which recorded the values measured at 10-second intervals. The data recorded by the data logger were used for plotting changes in the wet product mass and the changes of temperatures (gas at entry, gas at exit, wall, and product) in the function of drying time.

2.2 Materials and Methods

Measurements aimed to test the drying properties of sludge and to identify parameters favourable to drying. The impact of the initial moisture content of the sludge on drying, wall heating temperature and rotational speed of the agitator were examined. *Table 1* summarizes measurements with different drying parameters, where the drum loading factor (l) can be defined as follows:

$$l = \frac{V_P}{V_D} \tag{1}$$

Table 1

Values measured

No.	$T_{G,in}$ [°C]	Y_{in} [g_{wa}/kg_{dG}]	$\dot{m}_G$ [kg/h]	T_W [°C]	n [1/min]	$m_{P,in}$ [kg]	l [-]	$x_{P,in}$ [kg_{H2O}/kg_P]	$x_{P,out}$ [kg_{H2O}/kg_P]
1	100.7	5.0	239.3	80.0	38	5.19	0.16	0.831	0.507
2	100.7	4.1	185.4	67.1	50	7.84	0.21	0.810	0.509
3	110.9	9.7	150.9	70.6	28.5	5.52	0.15	0.799	0.478
4	110.2	13.4	183.5	72.7	28.5	5.47	0.15	0.796	0.465
5	110.5	9.1	168.4	80.4	38	5.07	0.14	0.612	0.216
6	110.3	10.6	167.9	87.8	38	5.56	0.17	0.821	0.284
7	110.3	9.4	166.8	65.4	38	8.33	0.22	0.782	0.212
8	110.3	8.6	166.4	65.4	38	8.28	0.21	0.778	0.284
9	110.1	9.0	166.0	66.3	38	6.19	0.15	0.350	0.185
10	110.1	9.1	166.8	65.8	38	8.10	0.19	0.650	0.288
11	108.7	8.9	166.5	65.1	38	7.82	0.18	0.534	0.221
12	109.1	10.0	167.7	66.1	38	7.09	0.20	0.786	0.476

At measurements No. 1-4, 7, 8, and 12, the product was fed in bulk (*Figure 2*), without any pre-treatment – without grinding and backmixing –, distributed evenly along the length of the apparatus. Measurements differed in the speed of the drying gas and the quantity of the dried product (load level).

Figure 2
Initial waste sludge (No. 1)

At measurement No. 5, the wet product was fed pre-treated – screened/mashed through a plastic sieve of 10x10 mm hole size (*Figure 3*) and backmixed with sludge dried earlier. The granular product was also attempted to be produced by grinding, but it was found that the wet product kneaded and made plastic during grinding compacted again before being fed into the apparatus, thereby breaking into small pieces was unnecessary. As the product tended to stick at this time as well, at measurement 6 it was already fed into the drying chamber with the original moisture content, with dry sludge powder sprinkled over the surface of the wet product screened, thereby reducing the clumping of granules. It was found at these experiments that the dry sludge powder sprinkled over the surface did not affect the process of drying and the end result perceptibly. As a result of mixing, the powder immediately got wet and its properties corresponded to those of the conglomerate. The grinding did not have a perceivable impact on the process of drying as the product gradually got compacted again as a result of mixing.

Figure 3
Sludge mashed through a sieve (No. 5)

As regards temperature curves at experiments No. 1-8, product and wall temperature curves showed irregularities after 40-50 minutes, which came to an end after a while. It can be stated on the basis of visual findings that in this interval the nature of the product changed from plastic into powdery.

At measurement 9, sludge powder dried earlier was dried further on. The measurement was performed in the phase of decreasing drying rate.

Based on the findings of earlier measurements, the initially ointment-like product, sticking both to the wall and the agitator, granulated at approx. 65% moisture content and it stopped sticking. At measurement 10, although the initial moisture content of the product was 65%, it still behaved as a smearing material. In order to avoid this, it was deemed necessary to apply further backmixing to reach lower moisture content values, in the course of which the moisture content of the mix was set lower than 65% (53.4%) at measurement 11. The product was fed into the pre-heated apparatus. First the dry product and then the wet product was filled in. As a result of agitation, the two materials got mixed quickly, and after a short while a granular load was produced. Adhesion of the product to the wall could not be detected. Initially it stuck to the agitator but this soon discontinued. At this measurement we succeeded in producing a non-adhesive load behaving like a granular conglomerate.

Figure 4
Waste sludge in the course of drying (No. 6)

The initially sticky, large-grained waste sludge was transformed into a small-grained powdery material easy to handle (*Figure 4*).

3 Results and Discussion

In the course of measurements, the following parameters were recorded continuously: entry and exit temperatures of the drying gas, the inlet volumetric flow of the drying gas, dryer wall temperature during wall heating, and the surface temperature of the product. Temperatures and product mass changes were depicted in the function of time on the diagrams included in *Figure 5* in case of four measurements optionally selected. The serial numbers of measurements are shown in the diagram title.

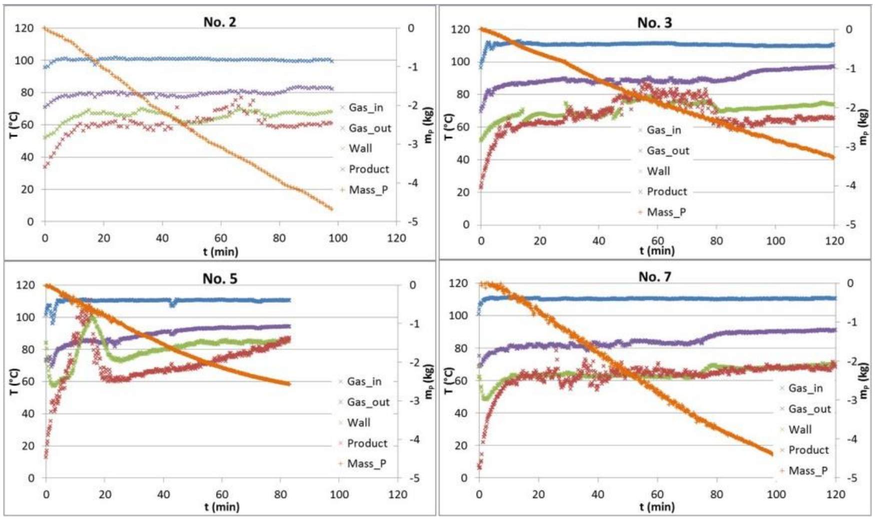


Figure 5

Temperature developments and changes in the mass of the waste sludge in the function of time (diagram titles refer to the serial number of measurements)

The entry temperature of the drying gas and the wall temperature were kept at a constant value, still, temperature fluctuations can be observed in the course of some measurements. The reason for this is that the output of the electric heating unit could not balance the high rate of heat extraction by the wet sludge. As product temperature was measured by an infrared thermometer placed on top of the dryer, it happened that the infrared thermometer would measure the temperature of the wall below it rather than product temperature because of agitation (the product stuck on the agitator or it was carried away from the below the measurement point of the thermometer). In such moments, temperature leaps larger than expected were observed (e.g. between minutes 45-80 in measurement 3). Changes in product moisture content can be calculated from its mass reduction.

Figure 5 clearly shows phases of constant drying rate. At some of the measurements (Nr. 5-9), the drying process reached levels below the critical moisture content of the product. In these cases (e.g. after 50 minutes during measurement Nr. 5), mass change and, as a consequence, the drying rate also reduced.

Drying rate can be calculated by the following correlation:

$$N = -\frac{dm_P}{dt}\frac{1}{A_{G-P}} = -\frac{dX}{dt}\frac{m_{dP}}{A_{G-P}}. \quad (2)$$

As it is highly uncertain and difficult to determine the contact area of the gas and the product in case of bulky and granular materials, the diagrams show – instead of the drying rate – the $\frac{dm_P}{dt}$ ratio proportionate therewith. *Figure 6* shows mass reduction rate in function of product moisture content in case of the four measurements selected. The diagrams show the phases of constant and falling drying rates, respectively.

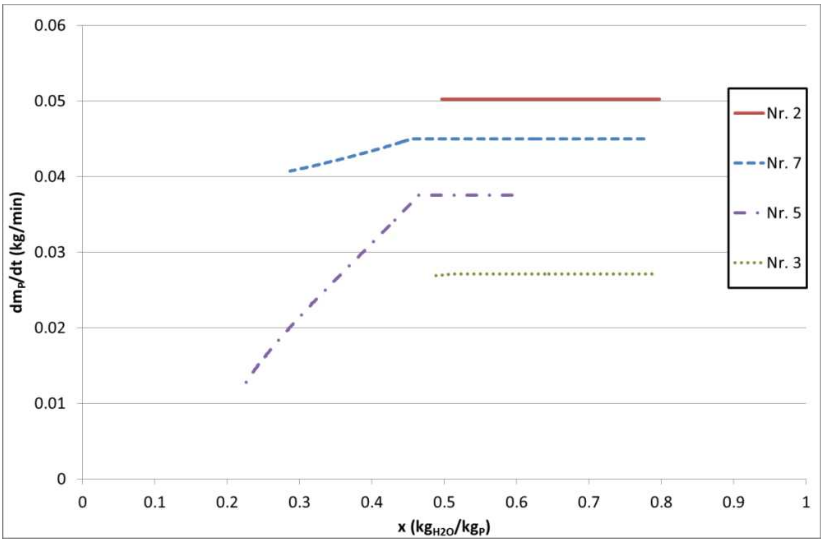


Figure 6

dm_P/dt in function of product moisture content in case of the four measurements selected

VOLUMETRIC EVAPORATION COEFFICIENT

The dryer can be dimensioned for operation by using the differential equation to describe changes in the temperature and moisture content of the drying gas [16]. In order to apply this differential equation system, it is essential to know the heat and mass transfer coefficients. By reason of the specificities of a mixed conglomerate, the actual surface to contact the drying gas is not known, therefore the volumetric heat transfer and evaporation coefficients ($\alpha a, \sigma a$) widely used in the literature are applied. The experiments performed are suitable for determining the volumetric mass transfer (evaporation) coefficient.

The following can be stated for the period of constant drying rate:

$$N = -\frac{dm_P}{A_{G-P} \cdot dt} = \sigma_{G-P}(Y_F - Y_G) \quad (3)$$

where the specific volumetric interfacial surface area between the gas and the product is:

$$a_{G-P} = \frac{A_{G-P}}{V_D} \quad (4)$$

Correlation between the evaporation coefficient and the mass transfer coefficient based on *Szentgyörgyi et al.* [17]:

$$\sigma_{G-P} = \rho_G k_c \tag{5}$$

Volumetric evaporation coefficient in the phase of constant drying rate, using equations (3) and (4):

$$\sigma a_{G-P} = -\frac{dm_P}{(Y_F - Y_G) \cdot V_D \cdot dt} \tag{6}$$

Mass transfer factor based on the Chilton-Colburn analogy, according to Treybal [18]:

$$j_M(Re) = St_M \cdot Sc_{G-P}^{2/3} \tag{7}$$

where

$$St_M = \frac{Sh_{G-P}}{Re \cdot Sc_{G-P}} \tag{8}$$

The Sherwood number can be stated in general for cases of mass transfer in plane, spherical and granular conglomerates by using equations (7) and (8) and applying the dimensionless numbers derived from the transfer coefficients:

$$Sh_{G-P} = j_M(Re) \cdot Re \cdot Sc_{G-P}^{1/3} = B \cdot Re^C \cdot Sc_{G-P}^{1/3} \tag{9}$$

where:

$$Sh_{G-P} = \frac{k_c \cdot d}{D_{v-G}} \tag{10}$$

$$Re = \frac{v_G \cdot d}{\upsilon_G} \tag{11}$$

$$Sc_{G-P} = \frac{\upsilon_G}{D_{v-G}} \tag{12}$$

This method can also be applied in case of the modified Sherwood and Reynolds numbers characterizing the gas/solid substance. The dimensionless equation can be stated in the following form:

$$Sh'_{G-P} = B \cdot Re'^C \cdot Sc_{G-P}^{1/3} \tag{13}$$

The volumetric evaporation coefficient can be determined from measurement data and by using equation (6), wherefrom the modified Sherwood number (*Sh'*) can be produced as follows by applying equations (4), (5), and (10):

$$Sh'_{G-P} = \frac{k_c \cdot \rho_G \cdot a_{G-P} \cdot d^2}{D_{v-G} \cdot \rho_G} = \frac{\sigma a_{G-P} \cdot d^2}{D_{v-G} \cdot \rho_G} \tag{14}$$

The modified Reynolds number by taking into consideration the characteristic velocities developed within the apparatus [16]:

$$Re' = \frac{\sqrt{(v_{cir}^2 + v_{ax}^2)} \cdot d}{\upsilon_G} \tag{15}$$

Constants B and C in equation (13) containing dimensionless numbers can be different depending on the way of mass transfer. In order to determine unknown constants:

$$Sh'_{G-P}/Sc_{G-P}^{1/3} = B \cdot Re'^{C} \tag{16}$$

In respect of measurements including phases of constant drying rate, the connection defined by equation (16) was determined and plotted in function of the modified Reynolds number.

On the diagram in *Figure 7*, the values yielded in the course of waste sludge drying are indicated by '♦'. The points covered a narrow range due to the fact that the drying parameter (and size of the product) was changed within a small range.

In order to extend the measurement range, earlier measurements [16] performed on other materials (millet, sunflower seed, wood block) were also used, assessed and processed in the manner described above. The values yielded are indicated by '+' in *Figure 7*.

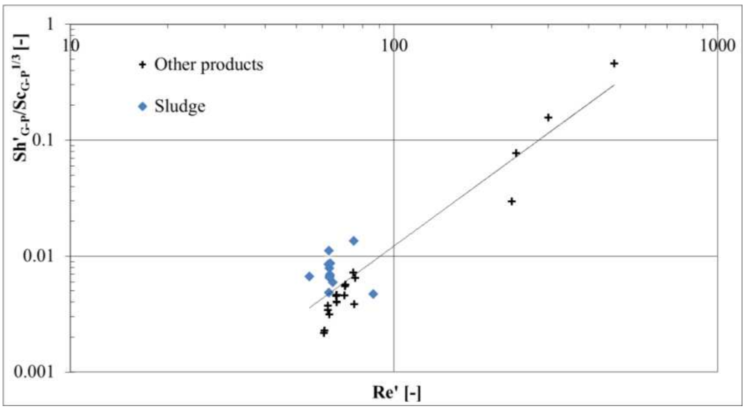


Figure 7
Correlation of $\mathrm{Sh}'_{G-P}/\mathrm{Sc}_{G-P}^{1/3}$ in function of the modified Reynolds number

The domain was considerably extended by supplementing earlier measurement point values ($55 < Re' < 480$). By fitting a straight line on the points depicted on the diagram of logarithmic scale (*Figure 7*), the equation of the straight line can be stated by the following correlation:

$$\frac{Sh'_{G-P}}{Sc_{G-P}^{1/3}} = 1.01 \cdot 10^{-6} \cdot Re'^{2,04} \tag{17}$$

Validity range of the equation to define the volumetric evaporation coefficient for conductive/convective drying in an agitated drum dryer:

- $T_P/T_W \approx 0.9$ and
- $0.68 \le T_W/T_G \le 0.83$ and
- $55 < Re' < 480$ and
- $0.1 < l < 0.25$.

Conclusion

The excess sludge to be dried, with an initial moisture content of 78-80% was a highly adhesive substance smearing along the wall of the dryer. It produced 'sausage'-type formations along the vertical wall part of the apparatus, it stuck on the agitator of the dryer, and it revolved in lumps within the drying chamber. After the moisture content of the sludge reached 65%, smearing and sticking gradually came to an end and the product granulated. Even the dried small-grained product contained some pieces of the size of a peanut, with higher moisture content than that of the main bulk. In order to reduce the initial moisture content, the product already dried (x_P=0.2÷0.4) was backmixed. Thereby the product conglomerate mixed to have an initial moisture content of approx. 65% is granular and therefore it is easy to dry. Wall heating had a favourable impact on sludge drying. No burning on the wall was observed in the temperature range applied.

At the end of these experiments, product load levels were about half or one third of the initial value. In order to prevent clogging and to realize convective drying effectively, an intermittent drum dryer should be loaded at a maximum load level of $l=0.22$.

A volumetric evaporation coefficient function was determined by using our sludge experiments and based on our earlier measurements. The results yielded can be very useful for the dimensioning of agitated drum dryers, thus providing industrial benefits to waste water treatment plants (transport cost reductions, deposition) and the renewables sector.

Acknowledgements

Special thanks to Dr. László Tömösy, retired associate professor for his assistance in the subject. TÁMOP-4.2.1.B-11/2/KMR-2011-0001 Research for critical infrastructure protection "The project is realized through the assistance of the European Union, with the co-financing of the European Social Fund."

Nomenclature

A	*Surface of heat and mass transfer*	m^2
a	*Volumetric interfacial surface area*	m^2/m^3
B	*Constant*	-
C	*Constant*	-
c	*Specific heat*	$J/kg/°C$
D	*Diffusion coefficient*	m^2/s
d	*Diameter of the product*	m
j	*Chilton-Colburn-factor*	-
k_c	*Mass transfer coefficient*	m/s
l	*Drum loading factor*	-
Le	*Lewis number*	-
m	*Mass of the product*	kg
Δm_P	*Decrease in the wet product mass*	kg

$\dot{m}$	*Mass flow*	*kg/s*
N	*Drying rate*	$kg/m^2/s$
n	*Rotational speed*	*1/s*
Re	*Reynolds number*	-
Re'	*Modified Reynolds number*	-
Sc	*Schmidt number*	-
Sh	*Sherwood number*	-
Sh'	*Modified Sherwood number*	-
St	*Stanton number*	-
T	*Temperature*	*°C*
t	*Time*	*s*
v	*Velocity*	*m/s*
V	*Volume*	m^3
x	*Moisture content of product on wet basis*	kg_{H2O}/kg_P
Y	*Absolute humidity of gas on dry basis*	kg_v/kg_{dG}
α	*Heat transfer coefficient*	$W/m^2/°C$
υ	*Kinematic viscosity*	m^2/s
ρ	*Density*	kg/m^3
σ	*Evaporation coefficient*	$kg/m^2/s$
σa	*Volumetric evaporation coefficient*	$kg/m^3/s$

Subscripts

ax	*Axial*
cir	*Circumferential*
D	*Dryer*
F	*Interface of the product*
G	*Humid drying gas*
G-P	*Between the drying gas and the product*
in	*Inlet*
H2O	*Water*
M	*Mass transfer*
out	*Outlet*
P	*Wet product*
v-G	*Between the vapour and the drying gas*
W	*Wall of the dryer*

References

[1] Joss, A.; Keller, E.; Alder, A. C.; Göbel, A.; McArdell, C. S.; Ternes, T.; Siegrist, H. Removal of Pharmaceuticals and Fragrances in Biological Wastewater Treatment. Water Research 2005, 39 (14), 3139-3152

[2] Grüter, H.; Matter, M.; Oehlmann, K. H.; Hicks, M. D. Drying of Sewage-Sludge – an Important Step in Wastedisposal. Water Science and Technology 1990, 22 (12), 57-63

[3] Vaxelaire, J.; Bongiovanni, J. M.; Mousques, P.; Puiggali, J. R. Thermal Drying of Residual Sludge. Water Research 2000, 34 (17), 4318-4323

[4] Chen, G. H.; Yue, P. L.; Mujumdar, A. S. Sludge Dewatering and Drying. Drying Technology 2002, 20 (4-5), 883-916

[5] Navaee-Ardeh, S.; Bertrand, F.; Stuart, P. R. Emerging Biodrying Technology for the Drying of Pulp and Paper Mixed Sludges. Drying Technology 2006, 24 (7), 863-878

[6] Ohm, T-I.; Chae, J-S.; Kim, J-E.; Moon, S-H. A Study on the Dewatering of Industrial Waste Sludge by Fry-DRYING technology. Journal of Hazardous Materials 2009, 168 (1), 445-450

[7] Romdhana, M. H.; Lecomte, D.; Ladevie, B. Dimensionless Formulation of Convective Heat Transfer in Fry-Drying of Sewage Sludge. Chemical Engineering Technology 2011, 34 (11), 1847-1853

[8] Peregrina, C.; Rudolph, V.; Lecomte, D.; Arlabosse, P. Immersion Frying for the Thermal Drying of Sewage Sludge: An Economic Assessment. Journal of Environmental Management 2008, 86 (1), 246-261

[9] Slim, R.; Zoughaib, A.; Clodic, D. Modeling of a Solar and Heat Pump Sludge Drying System. International Journal of Refrigeration 2008, 31 (7), 1156-1168

[10] Seginer, I.; Bux, M. Modeling Solar Drying Rate of Wastewater Sludge. Drying Technology 2006, 24 (11), 1353-1363

[11] Leon, M. A. Theoretical and Experimental Investigation of Solar-Biomass Hybrid Air Heating System for Drying Applications. Drying Technology 2009, 27 (6), 821-822

[12] Werther, J.; Ogada, T. Sewage Sludge Combustion. Progress in Energy and Combustion Science 1999, 25 (1), 55-116

[13] Ferrasse, J. H.; Arlabosse, P.; Lecomte, D. Heat, Momentum, and Mass Transfer Measurements in Indirect Agitated Sludge Dryer. Drying technology 2002, 20 (4-5), 749-769

[14] Léonard, A.; Meneses, E.; Trong, E. Le; Salmon, T.; Marchot, P.; Toye, D.; Crine, M. Influence of Back Mixing on the Convective Drying of Residual Sludges in a Fixed Bed. Water Research 2008, 42 (10-11), 2671-2677

[15] Léonard, A.; Blacher, S.; Marchot, P.; Pirard, J.-P.; Crine, M. Convective Drying of Wastewater Sludges: Influence of Air Temperature, Superficial Velocity, and Humidity on the Kinetics. Drying Technology 2005, 23 (8), 1667-1679

[16] Poós, T.; Örvös, M. Heat and Mass Transfer in Agitated, Co-, or Counter-Current Conductive-Convective Heated Drum Dryer. Drying Technology 2012, 30 (13), 1457-1468

[17] Szentgyörgyi, S.; Molnár, K.; Parti, M. Transzportfolyamatok; Tankönyvkiadó Budapest, 1986, 325-328

[18] Treybal, R. E. Mass-Transfer Operations; 3rd Ed. McGraw-Hill Company, New York, 1981

Web-based Smart Home Automation: PLC-controlled Implementation

Okan Bingol, Kubilay Tasdelen

Department of Electrical-Electronic Engineering, Suleyman Demirel University
Bati Yerleskesi, 32260 Isparta/Merkez, Turkey
okanbingol@sdu.edu.tr, kubilaytasdelen@sdu.edu.tr

Zekeriya Keskin, Yunus Emre Kocaturk

Department of Electronic and Computer Education, Suleyman Demirel University
Batı Yerleşkesi, 32260 Isparta Merkez, Turkey
l0811704029@stud.sdu.edu.tr, l0711704026@stud.sdu.edu.tr

Abstract: The most important source of motivation in continuity of technological developments is to upgrade human living standards. The technological development provides and increases human-beings' safety and comfort directly and indirectly. Developing technologies for this purpose directly affects the life standards by means of smart home systems design. It is possible to classify smart home systems into two as local and remote. In this study, a smart home automation system design was carried out by using Delta DVP28SV model PLC (Programmable Logic Controller). Smart home system can be controlled in two different ways either by any internet-connected device or an operator panel assembled on PLC. Control of the ventilation, lighting and security units in the smart home were carried out. Instantaneous states and variations of the smart home system which is feedback based are monitored and can be changed from the user interface prepared with the C# programming language. Unusual circumstances occurred in security units have been reported to the user with SMS (Short Message Service).

Keywords: Home automation; PLC; Remote control; Smart home

1 Introduction

The "smart home" concept has emerged in the early 1980's when "smart building" concepts started to be used. In those years, smart homes were designed only for user convenience [1]. "Smart homes" have smart technologies having remote or centrally controlled functions and services. In a smart home, the inhabitants' wishes and needs concerning all or some part of equipment and functionality have priority [2].

In smart house systems, the internet is also used to ensure remote control. For years, the internet has been widely used for the processes such as surfing on the pages, searching information, chatting, downloading and installation. By the rapid developments of new technologies, monitoring, controlling services have been started to be served along with internet as an instrument providing interaction with machinery and devices [3]. In smart house systems, comfort and security of houses have been enhanced, usage of energy and other resources were provided in a more rational way, and considerable savings have been achieved [4, 5, 6]. In the beginning, the “smart home” idea was aimed to raise home comfort of people who are not disabled. Nowadays, it is also convenient to satisfy personal needs of older and disabled people, to help and support those people. This application area is very important and promising in the future [1, 4]. The concept of providing proper and full control on each device and electrical equipment as well as no requirement for processes, which required manual control before, are the reasons for being under concentration.

In recent years, intelligent systems were preferred for residences, shopping malls, skyscrapers, hotels, etc. There is a large market share for intelligent building devices. Many organizations, such as MIT, Siemens, Cisco, IBM, Xerox, Microsoft, etc., have been working on this issue and these groups have set up about 20 pieces of a smart home lab. These laboratories are not a training set. Each of the buildings covers a large footprint [7, 8].

To build a superior intelligent building, there are some essential points that should be taken into consideration. In order to make the system high-qualified and successful, an experienced designer who is very suitable for environmental conditions, open to developments, and who is very well familiar with the system is needed. The term" intelligent design" includes meanings such as sustainable design, high-tech usage and user-friendly design [8].

Smart homes can be dealt with four main aspects:

- The physical structure of the building
- The system (security, air conditioning, power control)
- Services (Internet, communication)
- Management (energy, illumination, irrigation)

The physical structure of the building is important in terms of heating and natural ventilation. This feature is implemented in the project phase of the building. Engineers carry out the calculation of correct use of insulation material and the positioning of the building for energy saving in the beginnning phase of the project while drawing the project. The other three basic aspects can be applied on the foundation after the construction of the building and even for pre-made buildings [8].

In today's modern era, automation is rapidly advanced while it becomes a part completing our homes and offices [9]. Automation can be generally described as a process following pre-determined sequential steps with a very ~~a~~ little or without any human exertion. Automation is provided with the use of various sensors suitable to observe the production processes, actuators and different techniques and devices [10].

During the design process, home automation system must be effective, easy to apply and at an affordable price. PLC is considered as an alternative to such systems. PLC, security monitoring, energy consumption management and control of machines and automatic production lines that are included in almost every field of industry automation systems in particular, are commonly used [11]. PLC is an electronic device designed to be used in the field of industry that controls a system or groups of systems through analog/digital data input/output terminals, providing general control by means of inherent functions of timing, counting, data processing, comparing, sorting, data transfer and arithmetic operations. At the same time use of PLC is very advantageous for several reasons such as being able to make changes on the software and for resuming the algorithm as the energy supplied back by saving data for a long time in the case of power failure.

The traditional teaching of engineering subjects, an appropriate combination of theory, exercises and laboratory experiments should be provided [12]. In order to gain practical skills and experience, theoretical knowledge as well as intensive training in a laboratory must be given to electrical engineering students [13]. Constructivist educators propose that processes such as knowledge or learning are initiated by the individual his or herself. Training materials are needed to be presented to student both in a convenient and attractive way [14].

Many smart home systems developed for educational purposes are available in the literature. Here are some examples designed for educational purposes; set-design for smart home education [8], a smart home lab for students [14], a tool for facilitating the teaching of the smart home [15], a laboratory experiment for teaching automation inspired by the smart home [16], automatic small-scale house for teaching home automation [17], and remote control laboratory using a greenhouse scale model [18].

In this study, a web-based software and hardware application for smart home automation system has been realized. Security (gas and smoke, motion, door control) and comfort (lighting and climate control) of the smart home is provided by means of remote control. In this study, the PLC equipment called DVP28SV of Delta company with a sufficient number of digital and analog address has been used for the storage of the required information belonging to the smart home and due to ease of application over the operator panel. On the used PLC, there is 6 digital input (DI), and 12 digital output (DQ) terminals. In addition, one analog module named DVP04AD-S having 4 analog input (AI) terminals and one Ethernet module named DVPEN01-SL have been connected to the PLC. Smart home system is controlled by both PC and operator panel called DOP-AS35THTD. The user interface is written using C # programming language.

2 The Architecture of Smart Home System

In this study, a PLC controlled smart home application which can be controlled by using computer, mobile devices and operator panels via the Internet is implemented. This application is composed of two parts, including software and hardware. General block diagram of the system is shown in Figure 1.

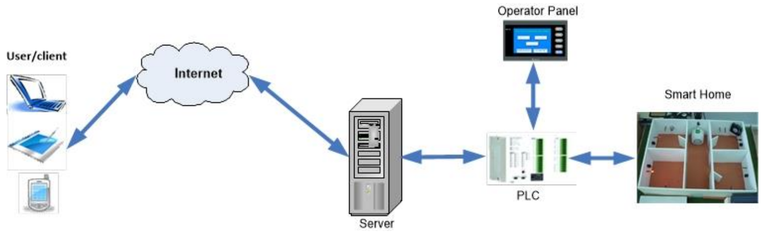


Figure 1
The block diagram of smart home system

2.1 Hardware Infrastructure

Infrastructure of the performed application of smart home hardware consists of Delta DVP28SV model PLC, DVPEN01-SL Ethernet module, DOP-AS35THTD touch operator panel, sensors, actuators and electronic circuits. In this study, 7 out of 16 digital inputs of the PLC, and 11 out of 12 digital outputs are used. Block diagram of PLC module is shown in Figure 2, while the inputs / outputs of the PLC used in the study are shown in Table 1.

It is possible to divide units controlled in the smart home system into 3 as lighting control, security control and air conditioning unit. Controls of these units are carried out by the operator panel connected to mobile devices and the PLC via the Ethernet module connected to the PLC. Via the lighting unit, living room, bedroom, kitchen and hall lights are controlled. When the system is turned on, lighting data of rooms are read and written to the database, and then transferred to the user interface so that the user can observe the current situation of rooms and changes in the state of the lights. In the energy-saving mode, thanks to the sensors, lights in unused rooms are automatically turned off.

In safety unit, gas leakage, fire, theft and door security checks have been done. When a signal is detected by gas and smoke sensors placed in the smart home, they run the alarm circuitry and information for the user is provided on the user interface. When an intruder is detected by the motion sensor placed in the hallway, pictures are taken by the camera and sent into e-mail of the user or another pre-defined address, and information is given in the user interface. With the sensor placed to the gate of the house, intrusions from the entrance door are detected and the alarm circuitry is activated.

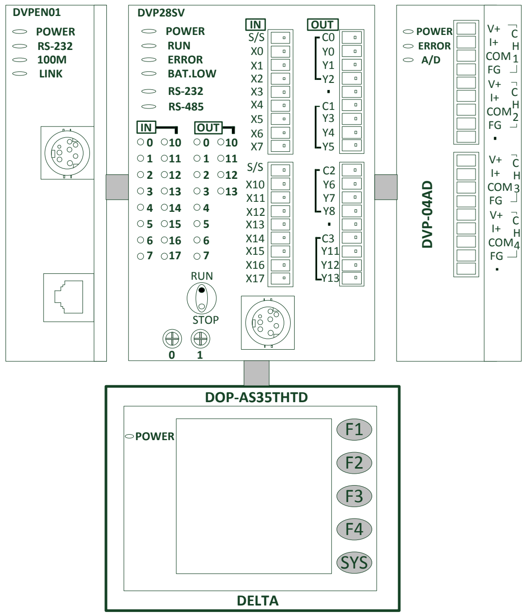


Figure 2

PLC module block diagram

Table 1

PLC Outputs and Inputs

Inputs	Tasks	Outputs	Tasks
X0	Living Room (Lighting Control)	**Y0**	Living Room (Lighting Control)
X1	Bedroom (Lighting Control)	**Y1**	Bedroom (Lighting Control)
X2	Kitchen (Lighting Control)	**Y2**	Kitchen (Lighting Control)
X3	Hall (Lighting Control)	**Y3**	Kitchen (Lighting Control)
X4	Entrance Door (Security Control)	**Y4**	Saving Mode (Lighting Control)
X5	Gas and Smoke (Security Control)	**Y5**	Alarm (Security Control)
X6	Motion Sensor (Security Control)	**Y6**	Heater (Ventilation Control)
X7	Temperature (Ventilation Control)	**Y7**	Cooler (Ventilation Control)
X9	-	**Y11**	Motion Sensor (Security Control)
X10	-	**Y12**	Gas and Smoke (Security Control)
X11	-	**Y13**	Entrance Door (Security Control)

Air-conditioning unit is a control method carried out in order to ensure the comfort of home residents. With the temperature sensor capable of measuring temperature between -55 C^0 and +150 C^0 placed on the smart home, the temperature of the house is measured in real time and transferred to the user interface. The user can set home temperature to any value. The coolant is activated when the home temperature is above the value set, and the heater is activated when it falls below it.

In order to perform controls mentioned above, active/passive settings on PLC inputs and outputs must be made.

Table 2
Equipment used in smart home system

No	Hardware
1	PC
2	Web Camera
3	Electronic circuit control cards
4	Relays
5	Supply of electronic circuit cards
6	PLC power supply
7	DVPEN01-SL Ethernet module
8	DVP28SV model PLC
9	DVP04AD-S analog module
10	Smart home prototype
11	Modem
12	DOP-AS35THTD operator board

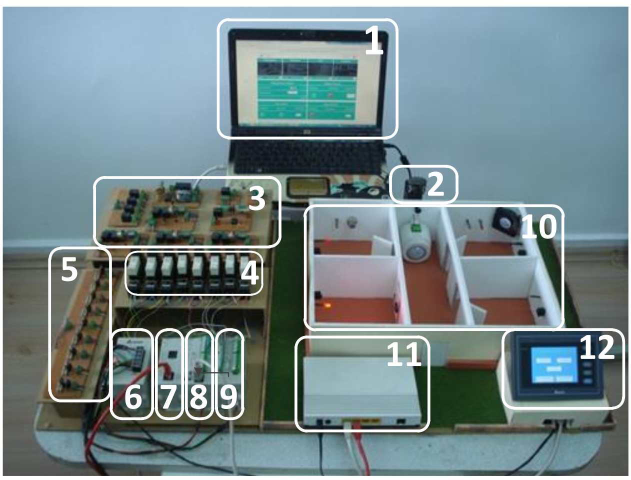


Figure 3
A general view of smart home system

2.2 Software Infrastructure

Hardware and software components of the realized smart home application are consisted of user, web, server, computer, PLC, operator panel and smart home components, and program and circuitry prepared for these components.

For the control of intelligent home via the internet in real-time, a simple and convenient infrastructure is constructed based on the client/server architecture. This infrastructure is made up of client software, server software and database section. The client software is the web interface by which users can control the smart home system through the internet. The main task of the client software running on the server is to manage lighting, ventilation and security units placed in the smart home system. The client software can perform the changes made on itself by DLL (Dynamic Link Library) is named DMT positioned essentially on Delta Modbus Library. It is possible to access inputs and outputs, registrars, auxiliary relays and memories of delta PLC through the methods reserved to user over DLL named DMT imported to client software. Users need a computer, tablet or mobile device with internet access in order to connect to the system. The control software is kept on the server by which all communications of the smart home system are provided. The server works in both directions. It records commands received from the smart home to the data base and sends recorded control commands from database to the smart home through PLC. The database comes into play in this part of the study. All adjustments and operations related to the smart home system are recorded in the database. Reports on the status of smart home are given depending on user request and the time. C# programming language has been used for the client and server software, while SQL server program has been used for the database. These three units interact constantly with each other although they are involved in different tasks and processes. The interaction between these units and the developed system architecture are shown in the block diagram in Figure 4.

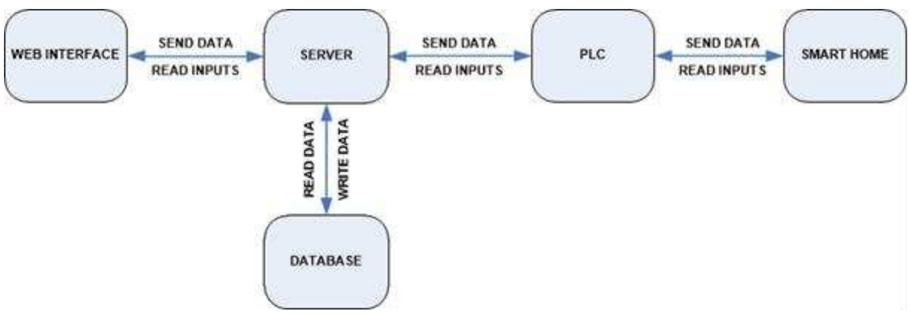


Figure 4

Software interaction

2.3 Web Interface

Part of smart home system that integrates with the user is the web interface. Here, the user can perform all the operations of a smart home system. The user gets the most up to date information about the smart home system in real-time using the web interface. Users can access all the rooms of the smart home system, the temperature value of the house, the motion control in the house (the thief), gas and smoke control and door control units through the web interface. In this way, they are able to manage smart home any way they want. The web interface is consisted of three parts as user home page, user control page and the reporting page. Login page is the part on which user logs in with the user name and password. In addition, pre-defined users on the database can log into the system. Figure 5 shows the home page.

Part of smart home system that integrates with the user is the control page. This page is working dynamically and any change that occurred in the smart home is reflected to control page. In the same way, a command given via the control page is submitted to the smart home. When the user looks at the control page he/she gets real-time information about status of the house. By using control page, the user can control lighting, security and air-conditioning units of the rooms. At the same time, adjustments related to these systems are able to make the system active/passive. The control page shown in Figure 6 faces to the user entering the user's name and password.

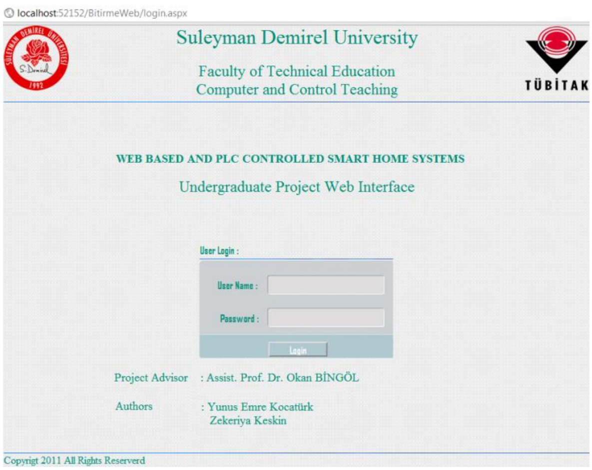


Figure 5
Home page

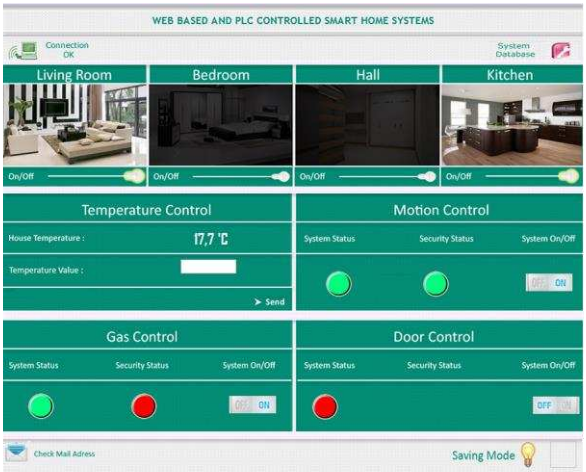


Figure 6

Control page

Reporting page brings the user all changes made by users on the internet or via the operator panel by taking from the system database. The user can see all changes about the home together with the date. Figure 7 shows the reporting page. The flow diagram describing the general operation of smart home system is shown in Figure 8.

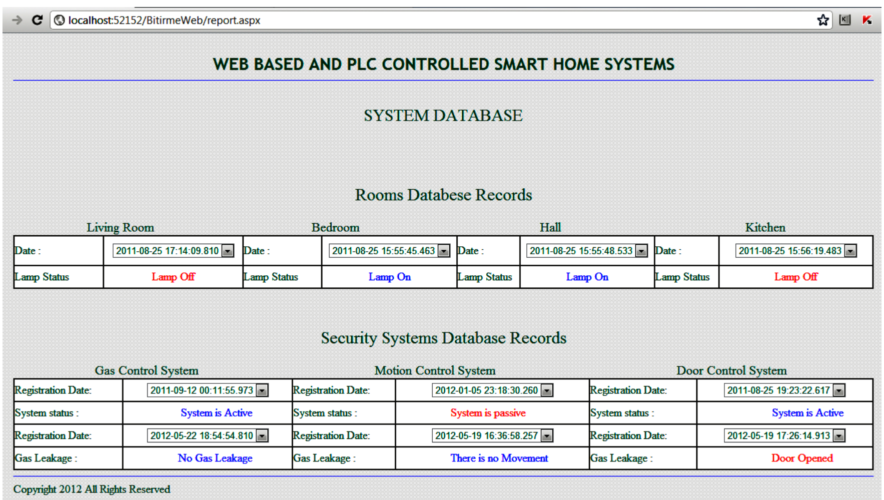


Figure 7

Reporting page

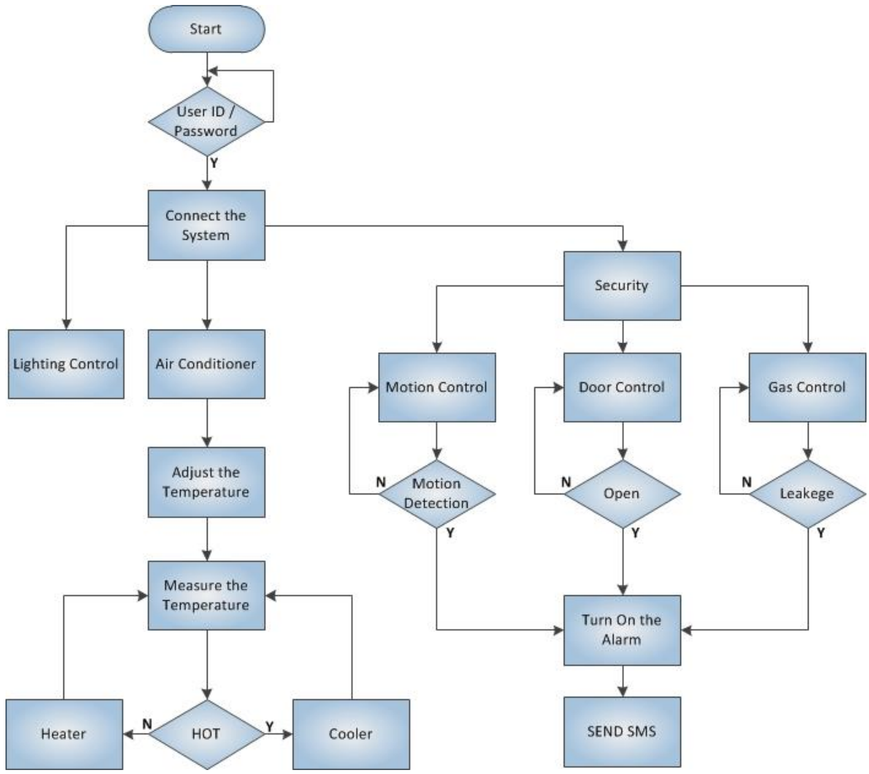


Figure 8
Flow diagram of smart home system

2.3 Operator Panel Interface

The control process of the smart home can also be carried out via the operator panel without the internet as mentioned above. Operator panel communicates directly with the PLC. All of the operations of the smart home system can be made via the operator panel. Lighting, security and air conditioning systems controlled and monitored via the internet in smart home system are also carried out via the operator panel. Four different interfaces were prepared on the operator panel for each operation process. The input interface is shown in Figure 9. The other three interfaces prepared for lighting, security, and air conditioning units can be accessed from input interface. Figures 10, 11 and 12 show lighting, security and temperature control panel interfaces respectively.

Figure 9

Operator panel home page

Figure 10

Lighting control panel

Figure 11

Security control panel

Figure 12

Temperature control panel

Conclusions

In this study, a web-based software and hardware for smart home automation system has been implemented. Smart home system was controlled by both PC and operator terminal called DOP-AS35THTD. The user interface is written using C # programming language. In the smart home system, controls of security (gas and smoke, motion, door control) and comfort (lighting and climate control) were carried out. In this study, PLC device of Delta Company having a sufficient number of digital and analog addresses was used for ease of information storage required for a smart home system and application on the operator panel. In addition, analog module with 4 analog input terminals and the Ethernet module are connected to the PLC.

Acknowledgements

The authors acknowledges TUBITAK for the realization of this study, providing financial support within the "2209-College Students Domestic/Abroad Research Projects Support Program".

References

[1] D. H. Stefanov and Z. Bien, The Smart House for Older Persons and Persons With Physical Disabilities: Structure, Technology Arrangements, and Perspectives, IEEE Transactions On Neural Systems And Rehabilitation Engineering, Vol. 12, No. 2, June 2004, pp. 228-250

[2] J. Lertlakkhanakul, J. W. Choi and M. Y. Kim, Building Data Model and Simulation Platform for Spatial Interaction Management in Smart Home, Automation in Construction, Vol. 17, Issue 8, November 2008, pp. 948-957

[3] A. R. Al-Ali and M. AL-Rousan, Java-based Home Automation System, IEEE Transactions on Consumer Electronics, Vol. 50, No. 2, May 2004, pp. 498-504

[4] R. J. C. Nunes and J. C. M. Delgado, An Internet Application for Home Automation, 10th Mediterranean Electrotechnical Conference, MeleCon 2000, Vol. I. pp. 298-301

[5] C. Douligeris, Intelligent Home Systems, IEEE Communications Magazine, Vol. 31, Issue 10, October 1993, pp. 52-61

[6] Y.-J. Mon, C.-M. Lin and I. J. Rudas, Wireless Sensor Network (WSN) Control for Indoor Temperature Monitoring, Acta Polytechnica Hungarica, Vol. 9, No. 6, 2012, pp. 17-28

[7] L. Jiang, D.-Y. Liu and B. Yang, Smart Home Research, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Vol. 2, August 2004, pp. 659-663

[8] E. N. Yılmaz, Education Set Design for Smart Home Applications, Computer Applications in Engineering Education, Vol. 19, Issue 4, December 2006, pp. 631-638

[9] İ. Coşkun and H. Ardam, A Remote Controller for Home and Office Appliances by Telephone, IEEE Transactions on Consumer Electronics, Vol. 44, Issue 4, Nov. 1998, pp. 1291-1297

[10] A. K. Gupta, S. K. Arora, Industrial Automation and Robotics, Dec 1 2007, 348 pages, Laxmi Publications, ISBN-10: 8131801810

[11] S.-J. (‘‘Tony’’) Hsieh, P. Y. Hsieh, Web-based Modules for Programmable Logic Controller Education, Computer Applications in Engineering Education, Vol. 13, Issue 4, December 2005, pp. 266-279

[12] T. Yigit and Ç. Elmas, An Educational Tool for Controlling SRM, Computer Applications in Engineering Education, Vol 16, Issue 4, 2008, pp. 268-279

[13] A. Keyhani, M. N. Marwali, L. E. Higuera, G. Athalye and G. Baumgartner, An Integrated Virtual Learning System for the Development of Motor Drive Systems, IEEE Transactions on Power Systems, Vol. 17, Issue 1, 2002, pp. 1-6

[14] L. Burnell, A. Sanchez, J. Priest, and C. Hannon, The Crescent Lab: A Smart Home Lab for Students, Computer Science, ENC ’06. Seventh Mexican International Conference, 2006, pp. 55-61

[15] M. Jimenez, P. Sanchez, F. Rosique, B. Alvarez and A. Iborra, A Tool for Facilitating the Teaching of Smart Home Applications, Computer Applications in Engineering Education, 8 FEB 2011, DOI: 10.1002/cae.20521

[16] S. Alayon, C. Gonzalez and P. Toledo, A Laboratory Experiment for Teaching Automation Inspired by the Smart Home, Computer Applications in Engineering Education, 24 Feb 2011, DOI: 10.1002/cae.20530

[17] F. Mateos, V. M. Gonza´lez, R. Poo, M. Garcı´a, and R. Olaiz, Design and Development of an Automatic Small-Scale House for Teaching Domotics, Proceedings of 31st ASEE/IEEE Frontiers in Education Conference, Reno, Nevada, 2001, Vol. 01

[18] J. L. Guzman, M. Berenguer, F. Rodriguez, and S. Dormido, Web-based Remote Control Laboratory using a Greenhouse Scale Model, Computer Applications in Engineering Education Vol. 13, Issue 2, Jul 2005, pp. 111-124

The Influence of Students' Personality Traits on their Perception of a Good Teacher within the Five-Factor Model of Personality

Anna Göncz, Lajos Göncz, Jasmina Pekić

Department of Psychology, Faculty of Philosophy, University of Novi Sad, Dr. Zorana Djindjića 2, 21000 Novi Sad, Serbia
agenc@ff.uns.ac.rs, goncz@ff.uns.ac.rs, jpekic@ff.uns.ac.rs

Abstract: This paper investigates the extent to which it is possible to predict the influence of personality traits of students of computer sciences (i.e. natural and technical sciences, n = 188) and of humanities (i.e. social and humanistic sciences, n = 250) on their perception of a good teacher. The five-factor model of personality provided the theoretical framework of the research, in which personality traits are classified into five domains: neuroticism, extraversion, openness to experience, agreeableness and conscientiousness. Statistical analysis on the total sample (n = 443) at the domain level showed that about 20% of the criteria variance (assessment of personality characteristics ascribed to a good teacher using the Big Five Inventory questionnaire) was attributable to the predictors (self-assessment of personality dimensions using the same questionnaire). For the analysis of individual questionnaire items, the common variance was 15%. Self-reports of openness, agreeableness and neuroticism were better predictors of evaluations of a good teacher than self-reports of extraversion and conscientiousness. The basic expectation of a good teacher was openness to experience, and the aspects of agreeableness and neuroticism that provide good interpersonal relations. Students of social and humanistic sciences favoured more extravert, open and conscientious teachers than natural and technical sciences students. The prediction of good teacher expectations (on the basis of personality characteristics) was more successful with the students of natural and technical sciences than with social and humanistic sciences students.

Keywords: personality traits of students; assessment of teachers; five-factor model of personality

1 Introduction

Research in which the personality of the teacher is considered in the context of a certain theory of personality represents a significant step forward. It provides a theoretical framework for the formulation of more specific hypotheses that are suitable for the empirical verification, and aids in the interpretation of findings. In addition, such an approach resolves certain methodological uncertainties regarding the manner of data collection. It makes replication of research easier, and allows generalizations of research findings. In this paper we report a study in which a variant of this approach was applied.

This paper sought to answer the following question: to what extent can students' assessments of personality (using the five-factor model) attributed to a good teacher be predicted based on the personality characteristics of the students? This paper also examined whether the prediction of a good teacher was more successful based on the personality characteristics of students from different disciplines, namely computer sciences and humanities. This represents a novel aspect of the current research.

1.1 Methodological Dilemmas in Assessing the Personality of Teachers

A review of the research related to the assessment of teachers' personality reveals that the issue of data collection has been resolved in various ways. Feldman [1] gave an overview of 72 studies: some researchers asked respondents to write a term paper on the characteristics of their most or least favoured teacher, while others opted for the technique of free guidance of characteristics or offered respondents a choice from a list of characteristics. The classification of responses and limited opportunities for in-depth statistical analysis of the data are weak points in most such studies. Although these procedures are still used [2], some progress has been made.

Lamke [3] was the first to introduce an approach that applied a questionnaire based on Cattell's theory of personality [4] for the (self) description of the characteristics and behaviour of successful and unsuccessful teachers. A similar approach was more recently employed by Mount, Barrick and Stewart [5] and Aidla and Vadi [6] using questionnaires based on the five-factor model of personality [7]. Rushton, Morgan and Richard [8] used a questionnaire based on Jung's theory [9, 10]), while Li and Wu [11] relied on Eysenck's [12] theory of personality. This approach is essentially a modified version of previous approaches in which respondents choose from a list of characteristics. However, a significant advantage of a technique that uses standardized questionnaires of personality is that the same list of descriptions is always offered. This facilitates the replication of studies and enables comparisons with results obtained from different samples. It can also enable generalization of the results.

As in Aidla and Vadi's study [6], the current study adopted the five-factor model of personality as the theoretical framework, using the Big Five Inventory (44-item version).

1.2 The Five-Factor Model of Personality as a Framework for Research on the Personality of the Teachers

The five-factor model of personality describes the personality structure in terms of personality traits arranged along five dimensions, namely: neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A) and conscientiousness (C).

The dimension of neuroticism represents emotional lability or stability. Emotional lability is characterized by: the dominance of negative feelings; anxiety; depression and perturbation; difficulty in coping with stressful situations; poor impulse control; irrational reactions; and a tendency towards mental disorders, uncertainty and changeable moods. At the other end of the dimension are emotionally stable personalities who are rarely anxious, inoffensive, easily adapt to changes, not irritable or depressed, of balanced mood, calm, cold-blooded, and have good capacities for the prevalence of stress.

Extraversion is the dimension of sociability, and is, sometimes also considered a dimension of energy. The dominant characteristic of an extravert person is positive emotions. These individuals are confident, dynamic, active, talkative, excitement-seeking, ambitious, friendly, warm and enthusiastic. They have highly developed social skills and diverse interests. At the other extreme are introverts, who are aloof, calm, shy, thoughtful, and prefer to work independently. These individuals have developed an ability to concentrate and analyse, and they tolerate monotony well.

Openness to experience is described as receptivity to new experiences and culture, and is associated with a rich inner life, imaginativeness and creativity. Open people are willing to experiment. They are curious, tolerant and accept new values and ideas in a non-dogmatic manner. They are not slaves of the authorities, but neither are they inconsistent nor unprincipled. They have an expressed interest in art and science. At the other extreme of this dimension are closed people that are not inclined to change and innovation, and instead prefer the conventional and familiar. They are traditional, with conservative views, though they may not necessarily be authoritarian and intolerant.

Agreeableness represents another sociability dimension. Agreeable people are generous, sensitive to the needs of others, willing to help and forgive, mild-mannered, modest, attentive, benevolent, and humane. Personal interests are not at the forefront. At the other end of the dimension are those disagreeable, distrustful, cynical, suspicious and insensitive, who do not consider the consequences of their actions.

Conscientiousness is the dimension of reliability, acceptance of responsibility, accuracy, durability, and adherence to policies and planning. Conscientious persons have developed work habits; they organize and systematically carry out their duties in a self-disciplined and deliberately planned manner. They are strong willed and are determined to achieve goals. If it is overly expressed, conscientiousness leads to compulsive neatness and pettiness. At the other end of the dimension are people who are more casual. Such individuals approach their duties in a disorganized manner and without an elaborate plan. They are less interested in achieving objectives and in this regard are less guided by ethical principles, though not necessarily immoral.

The most common application of the five-factor model is in the field of organizational behaviour, particularly in professional selection and occupation choice. Liao and Lee [13] found a significant positive effect of all dimensions on engagement in the workplace except for neuroticism, which had a negative effect. In addition, Barrick and Mount [14] found that conscientiousness was a good predictor of success in very different occupational groups, and extraversion may predict performance in occupations where social interaction is important. With regard to job satisfaction, a meta-analysis by Judge et al. [15] revealed a significant positive correlation with conscientiousness, extraversion and agreeableness, and a negative correlation with neuroticism, but only the relations with neuroticism and extraversion were found in all studies.

The dimensions of the five-factor model of personality can provide a basis on which to formulate certain expectations about preferred and undesirable personality characteristics of teachers. For instance, it seems reasonable to assume that the desirable traits of a good teacher would include low neuroticism, with pronounced extraversion, openness to experience, agreeableness (pleasantness) and (adaptive) conscientiousness. The extreme values in all domains are likely to be considered undesirable. Some of these assumptions are confirmed by empirical findings. Sanchez et al. [16] concluded that social science students at universities in Andalusia expected most of their teachers to have respect in their relations and be understanding (agreeableness), be open to cooperation without regard to personality characteristics (openness to experience), and have good teaching skills (conscientiousness). Aidla and Vadi [6] found that male teachers from Estonia were more agreeable and conscientious, but also less neurotic than the general population. The study of Genc et al. [17] confirmed that a good teacher is expected to have less emotional lability, with a prominence in extraversion, openness, agreeableness and conscientiousness.

Investigations of the determinants of students' assessment of a good teacher have focused primarily on gender. Studies have shown that expectations among females are more uniform [17]. Males are stricter, and females are milder in their evaluation of female teachers, while the assessment of male teachers does not depend on the assessor's gender [18].

2 Method

2.1 Sample and Procedure

Data were collected from 443 students at the University of Novi Sad, who were studying either natural and technical (or computer) sciences (Faculty of Science and Faculty of Engineering, subsample A, n = 188) or social and human sciences (or humanities) (Faculty of Philosophy and Arts Academy, subsample B, n = 250). In subsample A, 60 % of the respondents were female, with an average age of 21.85 years (SD = 1.53, range from 19 to 27 years). Their average mark in secondary school was 4.64 (SD = 1.53) on a scale from one to five, and at the university they had, on a scale from five to ten, an average mark of 8.34 (SD = 0.57). In subgroup B, 77% were female and their mean age was 21.30 years (SD = 0.72, age range 21 to 25 years). The mean mark for achievement in secondary school was 4.80 (SD = 0.46), and at university their average success was 8.46 (SD = 0.66).

Students' self-assessment on the Big Five Inventory revealed that those studying natural and technical sciences were significantly more agreeable and conscientious than the students of social and humanistic sciences, while the students of humanistic orientation were marginally more open than the natural and technical sciences students.

All respondents provided information regarding their faculty and department, gender, age, and success both in high school and at the university. During completion of the questionnaire, for each statement students rated to what extent the statement related to themselves. They then put themselves in the position of the teachers and answered in terms of how they believed a good teacher would respond. Data were collected during classes and by using the "snowball" method. Participation in the survey was voluntary and was rewarded with additional points for the students' course. The respondents were informed about the research and the manner of protection of data anonymity. Ninety-eight (20%) of respondents attended classes given in Hungarian and provided data in Hungarian.

2.2 Instrument

The current study employed the 44-item version of the Big Five Inventory (BFI) [19], using both the Serbian [20] and a Hungarian language version. The latter version was compiled with a back translation technique, and as a control, it utilized items both from the English version (available at http://www.outofservice.com/bigfive/) and a 132-item Hungarian version [21]. The questionnaire contains brief descriptions of personality, and the respondent rates the degree to which he/she agrees with every statement on a five-point scale (from 1 = strongly disagree to 5 = strongly agree). The items were selected on the basis of expert opinion regarding which were the best descriptors of the dimensions. Čolović [20] reported satisfactory indicators of convergent validity and different measures of the questionnaire reliability of the BFI (e.g. Cronbach's alpha coefficients of reliability of internal consistence for individual scales ranged from 0.72 to 0.80 for a Serbian population). In our study, Cronbach's alpha was 0.72 for neuroticism, 0.80 for extraversion, 0.88 for openness, 0.81 for agreeableness and 0.81 for conscientiousness. Given the small number of items in the subscales (eight for neuroticism and extraversion, nine for agreeableness and conscientiousness, and ten for openness) these are satisfactory indicators of reliability. The alpha coefficient for the entire scale was 0.88, and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.84, which exceeded the recommended value of 0.60.

To obtain a more complete picture, principal components analysis (PCA) was performed. Table 1 shows the correlations for students' self-assessment on the BFI. All correlations are significant, most of which have a low to moderate positive correlation, with the exception of the N scale which shows a low negative correlation. The scree_plot indicated that ten factors could be reasonably extracted from the items, i.e. that each dimension could be broken down into two sub-dimensions. This is consistent with earlier findings associated with the construction of various versions of the BFI [21, 22].

Table 1

Correlations between the BFI dimensions in our sample

Dimension	E	O	A	C
N	−0.241**	−0.202**	−0.241**	−0.212**
E		0.285**	0.271**	0.252**
O			0.175**	0.180**
A/C				0.403**

*Notes: N = neuroticism; E = extraversion; O = openness to experience; A = agreeableness; C = conscientiousness. ** = p < 0.01*

The instrument is brief and easily applicable, but does not provide sufficient nuanced description of personality.

2.3 Hypotheses

It was assumed that the appraisers' personality characteristics would be an important determinant of rated characteristics of a (good) teacher. Therefore, this study tested the following hypotheses. (1) We expected a significant positive correlation between self-assessed personality characteristics and assessed characteristics of a good teacher for dimensions (more robust traits of higher order). (2) A significant positive correlation was also predicted at the level of individual items (specific features of lower order) by which the dimensions are defined. (3) It was also assumed that in both groups of students the variance in ratings of a good teacher would be explained by personality characteristics to the same extent, i.e. self-reports of personality traits have the same predictive power with respect to their expectations of a good teacher.

2.4 Data Processing

Data were analysed using statistical packages SPSS (version 19) and Statistics (version 10). In addition to descriptive indicators, t-tests and analysis of variance were used for group comparisons. To examine the relationships between correlation coefficients, canonical correlation and regression analysis techniques were conducted.

3 Results and Discussion

3.1 Descriptive Indicators and Correlations Among Variables

Table 2 shows the descriptive statistics for predictor variables (self-evaluation of BFI dimensions) and the criterion variables (evaluation of a good teacher using BFI dimensions) for the total sample and subsamples. Table 3 shows the correlations between these variables.

Table 2

Descriptive statistics for predictor and criterion variables

Variables	N	Min	Max	M	SD	Sk	Ku
Predictor							
BFI self-assessment: total sample							
Dimensions							
N	420	9	38	22.79	5.13	0.04	0.00
E	433	9	40	29.62	5.77	−0.47	−0.02
O	428	12	50	38.70	6.92	−0.78	0.87
A	432	17	45	35.97	5.32	−0.50	−0.20
C	432	13	45	31.56	5.78	−0.22	−0.01
BFI self-assessment: subsamples							
Subsample A*							
Dimensions							
N	171	9	37	22.56	5.65	0.20	−0.27
E	174	12	40	29.59	5.63	−0.36	−0.,39
O	173	16	50	37.92	7.12	−0.74	0.19
C	173	17	45	36.56	5.39	−0.71	0.37
A	173	19	45	31.71	5.65	0.03	−0.70
Subsample B**							
Dimensions							
N	244	9	38	22.99	4.78	−0.10	0.27
E	243	12	40	29.74	5.86	−0.44	−0.18
O	240	12	50	39.41	6.54	−0.86	1.52
A	243	20	45	35.56	5.21	−0.30	−0.58
C	243	13	45	31.53	5.87	−0.40	0.31
Criterion							
BFI ratings of good teacher: total sample							

Dimensions							
N	430	8	38	19.43	4.54	−0.10	0.12
E	428	16	40	32.72	3.76	−0.63	0.97
O	426	17	50	42.35	5.48	−0.99	0.97
C	427	15	45	38.77	4.32	−0.76	0.43
A	429	18	45	38.07	4.66	−0.82	0.86
BFI ratings of good teacher: subsamples							
Subsample A*							
Dimensions							
N	171	10	38	19.23	4.84	0.28	0.22
E	172	16	40	31.80	3.83	−0.60	0.83
O	171	17	50	40.47	6.25	−0.98	0.19
A	172	15	45	38.20	4.69	−1.15	2.58
C	171	18	45	36.36	5.21	-−0.61	0.20
Subsample B**							
Dimensions							
N	244	8	30	19.60	4.28	−0.43	0.09
E	240	16	40	33.23	3.59	−0.66	1.26
O	240	28	50	43.64	4.49	−0.86	0.65
A	241	19	45	39.10	4.04	−1.16	2.86
C	242	23	45	39.13	3.88	−0.64	0.72

*Notes: N = number of respondents; Min = minimal row score; Max = maximal row score; M = mean; SD = standard deviation; Sk = skewness; Ku = -Kurtosis; N = neuroticism; E = extraversion; O = openness to experience; A = agreeableness; C = conscientiousness. * indicates students of computer sciences; ** indicates students of humanities*

The indicators in Table 2 show that the skewness and kurtosis scores for the total sample are very good (less than 1), and are generally acceptable for the subsamples. The subsample comparisons for BFI self-assessment showed that the students of natural and technical sciences were significantly more agreeable ($p < 0.05$) and more conscientious ($p < 0.10$) than the students of social and humanistic sciences, whereas the latter students were marginally more open ($p < 0.10$) than the students of natural and technical sciences.

Table 3

Correlations between self-assessed BFI dimensions and general characteristics of the student sample with the students' BFI evaluations of a good teacher (Pearson r): total sample

Student variables	Good teacher evaluation N	E	O	A	C
Self-assessment					
N	0.441**	−0.160**	−0.074	−0.156**	−0.128**
E		−0.370**	0.118*	0.203**	0.197**
O			0.495**	0.249**	0.295**
A				0.484**	0.217**
C					0.313**
Sample characteristics					
Gender	−0.054	0.118*	0.092	0.118*	0.074
Faculty	0.054	0.174*	0.271**	0.098*	0.291**
Secondary school achievement[1]	−0.077	0.196**	0.141**	0.134**	0.156**
University achievement	−0.037	0.111*	0.116*	0.063	0.100*

Testing language (S or H)	−0.353**	0.124*	0.002	0.162**	−0.37
Age (years)	0.016	−0.065	0.248**	0.068	−0.312**

*Notes: N = neuroticism; E = extraversion; O = openness to experience; A = agreeableness; C = conscientiousness; S = Serbian; H = Hungarian.[1]The grades for secondary school did not have a normal distribution; hence, Spearman's rho was calculated. *p < 0.05; **p < 0.01. (Underlining indicates a positive correlation of medium intensity.)*

The data shown in the upper part of Table 3 refer to the validation of the first hypothesis and are commented on in the next section. Here we are just noting that the results support the hypothesis, which suggests that the students' personality characteristics played a significant role in their evaluation of a good teacher (in terms of the five-factor model dimensions). The lower part of Table 3 shows a low but significant correlation between the sample characteristics and evaluations of a good teacher. The implications of these relations were analyzed in Table 6.

3.2 Validation of the First Hypothesis: the Relationship Between Self-Assessment and the Evaluation of a Good Teacher (in Terms of BFI Dimensions) for the Total Sample

The first hypothesis assumed a significant positive correlation (at the level of BFI dimensions) between self-assessment and the evaluation of a good teacher.

The upper part of Table 3 shows that students' self-assessment and their evaluation of a good teacher for the same dimensions of the BFI have a positive correlation of medium intensity (underlined in the table), according to the recommendations by Cohen [23] regarding interpretation of the strength of the relationship between the variables. Openness to experience (r = 0.495) showed the most pronounced covariation, followed by agreeableness (r = 0.484), neuroticism (r =441), extraversion (r = 0.370) and conscientiousness (r = 0.313). Preliminary analysis confirmed that the assumptions of normality, linearity and homogeneity of variance were satisfied; hence, it can be concluded that, for openness to experience, 24.5% of the variance of one variable was determined by the variance of the other. The coefficient of determination was 23.4% for agreeableness, 19.4% for neuroticism, 13.7% for extraversion and 9.8% for conscientiousness. These results suggest that self-assessment of openness to experience had the greatest influence on the students' evaluation of a good teacher, followed by agreeableness. For the questionnaire as a whole, the correlation was of medium intensity (r = 0.443, and the coefficient of determination r^2 = 0,196 suggested about 20% of the common variance), which supports our first hypothesis.

Table 3 also shows low but significant correlations between most self-assessments on a specific BFI dimension with assessments of a good teacher for the other dimensions. However, the correlation between the self-assessments of neuroticism and evaluations of teacher openness was not significant. Overall, this indicates that self-assessments of personality dimensions are to some extent a predictor of evaluations of a good teacher on some other dimension.

Canonical correlation analysis (and standard and hierarchical multiple regression analyses) was conducted to enable a more detailed interpretation (Tables 4–6). The analysis showed a significant association between predictor (BFI self-assessments) and criterion variables (evaluations of a good teacher) ($r = 53$, $p < 0.001$), which explained 28.53% of the common variance (about 9% higher than when the common variance was determined by coefficients of determination using a bivariate correlation, probably due to shared disruptive variables). As many as five pairs of significant canonical factors were extracted, of which from the first three specific relevant conclusions for our research could be deduced. Table 4 shows the matrix structure for these pairs of canonical factors.

Table 4
The factor structure of the three canonical factors for the correlations between predictor and criterion variables

BFI dimensions	Student personality (predictor variables): correlations with canonical factor	Good teacher personality (criterion variables): correlations with canonical factor
First factor		
N	0.456	0.628
E	0.407	0.198
O	−0.659	−0.927
A	0.649	0.863
C	0.043	0.103
Second factor		
N	−0.486	0.593
E	0.209	0.198
O	−0.310	0.038
A	−0.631	−0.691
C	−0.141	−0.059
Third factor		
N	−0.757	−0.704
E	−0.244	−0.080
O	−0.743	−0.777
A	−0.121	−0.172
C	−0.049	−0.132

Notes: N = neuroticism; E = extraversion; O = openness to experience; A = agreeableness; C = conscientiousness

Correlations with the first canonical factor suggest that the evaluations of a good teacher were primarily determined by the students' self-assessments on dimensions of openness to experience, agreeableness and neuroticism, and less so by extraversion. Conscientiousness had no significant impact on this factor. Very similar relationships were obtained for the linear correlation. However, the structure of relationships affected by the first factor further indicates that conservative students (negative correlation for openness to experience) expected a good teacher to be more conservative, but also more agreeable and more emotionally stable, and somewhat less extraverted. The second pair of canonical factors revealed that self-assessed agreeableness, and even neuroticism, were better determinants of the characteristics of a good teacher than openness to experience, and the weakly expressed role of extraversion remained slightly more prominent than conscientiousness. The third pair of factors provided more information about the role of neuroticism and openness to experience as predictors of good teacher evaluations. Emotionally stable students also expected emotional stability from a good teacher, and those who were open to experience had the same evaluations of a good teacher. Overall, this analysis revealed that self-assessments on the dimensions of openness to experience, agreeableness and neuroticism were better predictors of perceptions of a good teacher than self-assessments in the domain of extraversion and conscientiousness. In addition, students' self-assessments and evaluations of teachers for these dimensions were similar in terms of direction and were of greater intensity. If, for example, the evaluator was ranked highly on emotional stability or agreeableness, he/she expected even more pronounced emotional stability and agreeableness from a good teacher. For openness to experience the relationship was more complex: it seems that an open estimator wants an open teacher as well, but more pronounced conservatism (i.e. low openness) in students was related with their expectation to even greater conservatism in teachers.

Table 5

Standard multiple regression analysis: estimation of the model in terms of the prediction of personality traits of a good teacher and the partial contributions of predictors

M	r	r^2	Corrected r^2	Standard error of estimate	F	p
1	0.456[a]	0.208	0.198	11.962	21.118	0.0001
Predictors				Beta	t	p
N				0.139	2.944	0.003
E				0.134	2.771	0.006
O				0.304	6.468	0.000
A				0.163	3.258	0.001
C				0.102	2.055	0.040

Notes: M = model. Predictors: self-assessment of neuroticism (N), extraversion (E), openness (O), agreeableness (A) and conscientiousness (C). Dependent variable: BFI assessments of a good teacher

Table 5 shows the results of the standard multiple regression analysis. The regression model, in which evaluations of a good teacher (in terms of the BFI dimensions) were the criterion variables and the raw scores for BFI self-assessment were regarded as the predictor variables, was statistically significant ($F_{(5,403)} = 21.118$, $p = 0.0001$). The predictor variables explained 19.8% of the criteria, which is almost identical to the common variance obtained by calculating the bivariate correlations. Both the regression model and bivariate correlation indicated that each predictor significantly contributed to the prediction, but with different intensity. Openness to experience had the largest contribution (beta = 0.304, p = 0.000). This was followed by agreeableness (beta = 0.163 p = 0.001), neuroticism (beta = 0.139, p = 0.003) extraversion (beta = 0.134, p = 0.006) and conscientiousness (beta = 0.102, p = 0.040). Pearson's r indicated the same order. However, regression analysis using this more restrictive regression model (which removes part of the common variance attributable to overlapping between predictor variables), showed that the intensity of those relationships was far weaker than on the basis of linear correlation. The coefficients of determination in the linear correlations ranged from 9.8% - 24.5%, but for this regression analysis the range was from 1% (for conscientiousness) to 9.2% (for openness to experience).

This analysis confirmed that the self-assessment scores for the BFI dimensions as a group or block explain about 20% of the variance of the dependent variable, and that each individual dimension significantly contributes to explaining the unique variance of the dependent variable. Additional information obtained from this analysis that is of theoretical importance includes the fact that the individual contributions of each domain, if overlapping within them is excluded, are of significantly lower intensity. With such analysis, openness to experience explains about 9% of the variability in the assessment of a good teacher.

Relations between students' self-assessment and their evaluation of a good teacher on the BFI dimensions were also tested by hierarchical (sequential) multiple regression analysis with the disruptive effects of sample characteristics and testing language removed (the effects of such factors can be seen in the lower part of Table 3). This analysis sought to examine whether BFI self-assessments of the dimensions could still predict a significant part of the variance in the evaluation of a good teacher when the possible influence of such factors were removed. The main results of this analysis, which have more theoretical than practical importance for the current research, are presented in Table 6.

Table 6

Hierarchical multiple regression analysis: estimation of the model in terms of the prediction of BFI personality traits of a good teacher and the partial contributions of predictors

M	r	r^2	Corrected r^2	Standard error of the estimate	r^2 changes	F changes	df_1	df_2	p
1	0.379[a]	0.144	0.073	12.856	0.144	2.043	6	73	0.071
2	0.577[b]	0.333	0.225	11.756	0.189	3.860	5	68	0.004
1	**Predictors**[a]						Beta	t	p
	Sample characteristics								
	Gender						−0.010	−0.086	0.932
	Faculty						0.266	2.287	0.025
	Success at secondary school						0.107	0.881	0.381
	Success at the faculty						0.066	0.269	0.789

	Testing language/ ethnic affiliation	0.040	0.351	0.727
	Age/year of study	−0.166	−1.461	0.148
2	**Predictors**[b]			
	Sample characteristics			
	Gender	−0.086	−0.787	0.434
	Faculty	0.276	2.548	0.013
	Success at secondary school	0.063	0.562	0.576
	Success at the faculty	0.031	0.269	0.789
	Testing language/ethnic affiliation	0.056	0.537	0.593
	Age/year of study	−0.179	−1.658	0.102
	BFI self-assessment dimensions			
	N	0.171	1.554	0.125
	E	0.151	1.374	0.174
	O	0.262	2.472	0.016
	A	0.195	1.712	0.092
	C	0.095	0.827	0.411

Notes: M = model. [a]Predictors: gender, faculty, success at secondary school, success at the faculty, testing language/ethnic affiliation, age/years of study (characteristics of the sample). [b]Predictors: gender, faculty, success at secondary school, success at the faculty, testing language/ethnic affiliation, age/years of study, and self-assessment of neuroticism (N), extraversion (E), openness (O), agreeableness (A) and conscientiousness (C). Dependent variable: BFI - good teacher

The sample characteristics entered in the first step explained 14.4% of the variance of evaluations of a good teacher (previous analyses showed that the requirements of normality, linearity, multicollinearity and homogeneity of variance were met). The model was marginally significant, and only the faculty affiliation had a notable beta coefficient. This suggests that predictions of good teacher assessments were somewhat more successful for the students of natural and technical sciences than for the students' of social and humanistic sciences. In the second block the self-assessment scores in the BFI domains were entered, and the model as a whole explained 33.3% of the total variance of the dependent variable. After removing the influence of sample characteristics, the self-assessments of students explained an additional 18.9% of the variance in the assessment of a good teacher. In the final model, only two indicators were important: the student’s faculty and self-assessments of openness to experience. There was also a slight tendency for the dimension of agreeableness. The common variance was only slightly different (less than 1%) from that indicated by the standard multiple regression. However, the partial contributions of self-assessments as the predictors of criterion variables were far less pronounced than indicated by the bivariate correlation and standard regression analysis. Only openness to experience had an effect of moderate intensity (about 7%, compared with 9% in the standard regression analysis), and for the other predictors ranged from 0.9 to 3.8%. With the exception of openness to experience, the prominence of the predictors varied.

3.3 Validation of the Second Hypothesis: the Relationship Between Self-Assessment and the Evaluation of a Good Teacher for Individual Questionnaire Items (Total Sample)

In the second hypothesis, we assumed that the students' self-assessments and their assessments of a good teacher would be significantly positively correlated in terms of individual BFI questionnaire items, which would also suggest that their perceptions of a good teacher were significantly determined by personality characteristics. As shown in Table 7, the average correlation was of moderate intensity ($r = 0.387$), and the coefficient of determination ($r^2 = 14.98$) indicated that the common variance was about 15%. As the dimensions contain 8–10 items, it was expected that the prediction of the properties of a good teacher based on individual items would on average be poorer than prediction based on the dimensions when the common variance is 19.8%.

Table 7
Correlation matrix of individual BFI questionnaire items* between students' self-assessment and their evaluation of a good teacher (Pearson's r)

Average correlation: $r = 0.387$ ($SD = 0.099$), $r^2 = 14.98$

I	M_1 M_2	r	I	M_1 M_2	r	I	M_1 M_2	r	I	M_1 M_2	r
1	3.80 4.12	0.253**	**12**	4.05 4.29	0.317**	**23**	2.43 3.50	0.350**	**34**	2.91 2.98	0.454**
2	3.37 3.90	0.364**	**13**	4.32 4.63	0.406**	**24**	2.61 1.94	0.390**	**35**	3.17 3.33	0.536**
3	3.99 4.66	0.345**		3.55 2.91	0.510**	**25**	4.00 4.52	0.260**	**36**	4.30 4.47	0.370**
4	1.69 1.60	0.334**	**15**	3.92 4.36	0.413**	**26**	4.02 4.46	0.329**	**37**	3.32	0.387**
5	3.84 4.44	0.319**	**16**	3.99 4.53	0.390**	**27**	4.14 4.31	0.447**	**38**	3.62 4.39	0.312**
6	3.20 3.21	0.339**	**17**	4.23 4.44	0.436**	**28**	3.91 4.47	0.511**	**39**	2.98 2.02	0.387**
7	4.24 4.54	0.385**	**18**	3.53 4.28	0.275**	**29**	3.01 3.01	0.832**	**40**	4.21 4.45	0.385**
8	2.55 3.44	0.317**	**19**	3.34 3.06	0.296**	**30**	4.19 4.34	0.493**	**41**	3.79 4.13	0.381**
9	2.66 1.88	0.338**	**20**	4.07 4.43	0.283**	**31**	2.95 3.69	0.400*	**42**	4.18 4.59	0.382**
10	4.25 4.47	0.462**	**21**	3.44 3.77	0.292**	**32**	4.05 4.38	0.372**	**43**	3.11 4.03	0.295**
11	4.00 4.47	0.344**	**22**	4.36 4.39	0.481**	**33**	4.08 4.64	0.335**	**44**	3.31 3.85	0.420**

Notes: * *= As noted before, the description of items is available at http://www.outofservice.com/bigfive/;*
I = item number; M_1 *= mean of self-assessment;* M_2 *= mean of good teacher evaluation. ** correlation significant at* $p < 0.01$

Table 7 shows that all correlations were significant at the 0.01 level. They were divided into two groups according to the size of the effect based on the suggestions given for the interpretation of effect size by Tenjović and Smederevac [24]. A correlation of r = 0.364 or higher indicated a strong effect, and lower correlations were interpreted as an effect of medium intensity. Strong effects were indicated for 25 of the questionnaire items (having a coefficient of determination higher than 0.13), while the remaining 19 items had an effect of medium intensity (with a coefficient of determination greater than 0.05).

When interpreting the results, it should be noted that items 2, 6, 8, 9, 12, 18, 21, 23, 24, 27, 29, 31, 34, 35, 37 and 43 were recoded such that the lower arithmetic mean indicated a greater severity of the dimension. Taking this into account, it is apparent that the best predictor of the evaluation of a good teacher is item 29: *I see myself as a person who is mostly in a good mood* (r = 0.832). There was also a strong effect for another 24 items. In fact, for more than half of the items there was a relatively pronounced effect of self-assessment on the evaluation of a good teacher.

If we take the effect size as a criterion, among the 25 items with a strong effect, eight related to openness to experience, seven to agreeableness, five to neuroticism, three to extraversion and two to conscientiousness. This means that at the level of individual items, a good teacher was evaluated primarily on the basis of statements relating to aspects of openness to experience and agreeableness. The exception to this is (good) mood which is classified within the domain of neuroticism, but it is certainly related to pleasant (agreeable) and open (non-conventional) behaviour. It seems that the good mood of a teacher was the most important characteristic on which evaluations of a "good teacher” were made.

Among the remaining 19 items with medium influence, seven related to conscientiousness, five to extraversion, three to neuroticism, three to openness to experience and one to agreeableness. This again confirms that descriptions of behaviour related to conscientiousness and extraversion played a smaller role in evaluations of a good teacher. In addition, the results are in line with the findings of the previous statistical analyses, wherein openness, agreeableness and neuroticism were better predictors of evaluations of a good teacher than extraversion and conscientiousness. Student evaluations of a teacher were primarily formed based on the degree to which the teachers were open to experience, agreeable and emotionally stable.

If these results are viewed in relation to teachers’ responsibilities in contemporary schools, which need more student-centred situations, and teachers who know their students’ learning characteristics [25], it seems that the expectations of students toward personality traits of a good teacher and the responsibilities are in line: it is expected from a good teacher to have the obtained personality structure to incite the students self-actualization.

3.4 Validation of the Third Hypothesis: Differences in the Prediction of Good Teacher Evaluations Based on Personality Characteristics of Students from Different Faculties

We assumed that there would be no differences between the groups of students from different faculties, and that self-assessments of personality would explain the same percentage of the common variance in the evaluation of a good teacher within both groups of students. To test this hypothesis we compared the predictive power of self-assessments at the domain level for the two subsamples by determining the bivariate correlation between BFI self-assessments and evaluations of a good teacher. In addition, we compared the group results using standard regression analysis. The results are shown in Tables 8 and 9.

Table 8

Correlations between self-assessments and evaluations of a good teacher on BFI dimensions for students of natural and technical sciences (subsample A) and social and humanistic sciences (subsample B) (Pearson's r)

BFI self-assessment	BFI evaluation of a good teacher									
	Subsample A*					Subsample B**				
	N	E	O	A	C	N	E	O	A	C
N	0.458**					0.433**				
E		0.463**					0.345**			
O			0.704**					0.299**		
A				0.568**					0.451**	
C					0.435**					0.273**
	Average correlation: 0.649, $r^2 = 0.421$**					Medium correlation: 0.370, $r^2 = 0.137$**				

*Notes: N = neuroticism; E = extraversion; O = openness to experience; A = agreeableness; C = conscientiousness. *p < 0.05; **p < 0.01. * indicates students of computer sciences; ** indicates students of humanities*

Comparisons of correlations for the two subsamples (Table 8) show some significant differences. In subsample A the correlation between all self-assessments and evaluations of a good teacher (r = 0.649) was significantly higher than in subsample B (r = 0.370). The difference of r = 0.279 was highly significant (p <0.0001).The medium correlation for subsample A explains more than 42% of the common variance of the evaluation, whereas in subgroup B the self-assessments explain less than 14% of the common variance. This means that personality traits are far better predictors of perceptions of a good teacher for the students of natural and technical sciences compared with students of social and humanistic sciences. In addition, for the students of natural and technical sciences, all dimensions contributed to the common variance of the evaluation of a good teacher above 18%; the effects of openness to experience (49.56%) and agreeableness (32.26%) were particularly pronounced. For the students of social and humanistic sciences, only agreeableness and neuroticism explained about 20% of the common variance of the evaluation of teachers. In addition, the order of the impact of individual dimensions for the two subsamples was different.

Table 9

Standard multiple regression analysis for the students of natural and technical sciences (subsample A) and social and humanistic sciences (subsample B): evaluation of the model in terms of the prediction of BFI dimensions of a good teacher and the partial contributions of predictors

M	r	r^2	Corrected r^2	Standard error of estimate	F	p
1	0.688	0.473	0.456	10.634	28.029	0.000
1	0.456	0.208	0.198	11.962	21.118	0.0001
1	*Subsample A predictors*			Beta	t	p
	N			0.219	3.488	0.001
	E			0.095	1.483	0.140
	O			0.516	8.149	0.000
	A			0.222	3.363	0.001
	C			0.102	2.055	0.001
1	*Subsample B predictors*					
	N			0.146	2.203	0.029
	E			0.223	3.319	0.001
	O			0.146	2.297	0.023
	A			0.209	3.015	0.003
	C			0.050	0.713	0.447

Notes: M = model. Predictors: self-assessment of neuroticism (N), extraversion (E), openness (O), agreeableness (A) and conscientiousness (C). Dependent variable: BFI - good teacher

Within the regression model, the evaluation of a good teacher was the criterion, and raw scores for the BFI dimensions were the predictors. The model for the whole of subsample A was statistically significant ($F_{(5, 156)} = 28.029$, $p < 0.001$) and explained 47.3% of the criteria. Similar results were obtained for the bivariate correlation. The regression model shows that each predictor, with the exception of neutoticism, significantly contributed to the prediction (openness to experience had the greatest partial contribution), but the intensity of the relationship was weaker than the linear correlation has shown. In subsample B the model was also statistically significant ($F_{(5, 226)} = 7.939$, $p < 0.001$) and explained about 15% of the common variance of the evaluation. Each predictor, except conscientiousness, contributed significantly to explaining the criterion (but only up to 4.97% for extraversion). If we compare the prediction effectiveness in the two subsamples, the difference is obvious. F values differ for 20.09, which is highly significant (p

<0.001), and individual predictors explained a greater percentage of the criteria in subsample A compared with subsample B. This means that predictions based on the self-assessment of BFI dimensions were more successful for students of natural and technical sciences than students of social and humanistic sciences.

The results of both statistical tests were at odds with hypothesis three and suggested that the personality characteristics of the students of natural and technical sciences were better determinants of good teacher evaluations than those of social and humanistic sciences students.

4 General Discussion

The evaluation of teacher personality as the most important factor in educational work is the subject of many studies. This paper presents the results of a study in which students of natural and technical sciences and social and humanistic sciences were asked to evaluate the characteristics of a good teacher, based on the features included in the five-factor model of personality. We sought to answer two questions: (1) is it possible to predict the dimensions and personality traits attributed to a good teacher on the basis of personality characteristics of the appraisers; and (2) in which group of students is such prediction more successful?

The results for the sample as a whole suggested that the personality traits of the student assessors were significant predictors of their evaluations of a good teacher. All statistical analyses showed that a high percentage of the criterion variance (evaluations of a good teacher at the level of BFI dimensions or individual questionnaire items) could be explained on the basis of predictors (students' self-assessments for BFI dimensions or items of the questionnaire). The percentage for dimensions as predictors based on either the total sample or the student groups (determined by linear correlation and regression analysis) was about 20% (with the canonical correlation it was 28.5%). In both cases the intensity of the different dimension contributions in explaining the criteria followed the same order. The most significant contribution was from openness to experience, followed by (in descending order): agreeableness, neuroticism, extroversion and conscientiousness. The students' evaluation of the teacher was primarily formed based on perceptions of the teacher's openness to experience, agreeableness and emotional stability. The intensity of the domain influence ranged from strong (24.5%) to medium (9.8%) when determined by linear correlation. However, when the impact of overlapping between the predictor variables (which is the result of interactions between domains and the disruptive effects of the sample characteristics) was removed, the predictive power of the domains was significantly reduced. Only openness to experience remained a significant predictor. Analysis at the level of individual questionnaire items revealed the percentage of common variance to be 15%, which is a finding worth mentioning.

In addition, the self-assessments of the natural and technical sciences students contributed significantly more to the aforementioned regularities than those of students in social and humanistic disciplines. The percentage of explained variance of criteria based on dimensions was 42%, with each dimension making a significant contribution. The self-assessments of social and humanistic sciences students were not at odds with these findings, but their impact was less pronounced and less convincing.

The current study found that the evaluations of a good teacher were largely determined by properties of the assessor. When applied to educational practice, this essentially means that such evaluations may vary depending on the personality characteristics of pupils or students with whom the teacher works. This inevitably raises the question of what the desired personality of the staff in the pedagogical profession should be. This question is often approached in a somewhat naive form: is there a personality profile that would be ideal for educational work? If we define the ideal teacher as a teacher who is, according to their personality characteristics, optimally suitable to all students or groups of students with whom some form of pedagogical intervention is required, the answer is negative. Attempts to determine such a personality profile are doomed to failure because the goal is too general, and thus unrealizable. The reason is simple: the needs of members of all those groups with whom teachers work, especially in grammar schools, vary as much as general human needs. A desirable structure of personality in educational work can be determined only in relation to the structure of personality of pupils/students. Greater heterogeneity of educational groups reduces the likelihood that the educator's personality would optimally suit all in such a group. It seems that the essential psychological characteristics of teachers likely to be perceived as good, at least with respect to the properties covered by the BFI questionnaire, are properties of the openness to experience dimension, and also include the aspects of agreeableness and neuroticism that are necessary for good interpersonal relationships. Moderate extraversion and conscientiousness are additional characteristics of a good teacher.

Conclusions

This paper investigates to what extent students' personality traits can predict their attribution of the personality characteristics which a good teacher should have, and whether this prediction is more successful with students of natural and technical sciences, or with students of social sciences and humanities. A significant positive correlation between self-assessed personality traits and assessed traits of a good teacher was predicted. It was also assumed that in both groups the variance in ratings of a good teacher would be explained by personality traits of students to the same extent.

The analysis of the results revealed the following regularities:

1) Evaluations of personality qualities of a good teacher are determined by the personality characteristics of the assessors. Various correlational analyses of the

relationship between self-assessment and the evaluation of a good teacher for the BFI domains for the entire sample indicated about 20% of the common variance, i.e. a strong effect of the predictors on the criterion.

2) Self-assessments for the dimensions of openness to experience, agreeableness and neuroticism were better predictors of a good teacher than self-assessments in the domain of extraversion and conscientiousness. Students with a high degree of openness to experience and agreeableness, and low neuroticism (i.e. expressed emotional stability) required these domains to be even more pronounced in a good teacher, and conservative assessors preferred even greater conservatism from teachers.

3) Standard regression analysis showed that the dimension of openness had the highest predictive power, followed by (in descending order): agreeableness, neuroticism, extroversion and conscientiousness. However, when the potential impact of disruptive variables was removed, only openness to experience maintained an effect of medium intensity in estimating the variance of a good teacher.

4) Analysis at the level of individual items revealed that 15% of the variability in the assessment of a good teacher could be explained by the self-assessment. The evaluation of a good teacher was primarily based on items from the domains of openness to experience and agreeableness, and on the grounds of items relating to (good) mood (the latter of which comes under the domain of neuroticism but relates to pleasant (agreeable) and open (unconventional) behaviour). It seems that the general mood of a teacher was the most important characteristic in evaluation of "a good teacher".

5) The personality characteristics of natural and technical sciences students were more successful predictors of evaluations of a good teacher than those of students in social and humanistic sciences.

The results have confirmed the first hypothesis. However, they are at odds with the second hypothesis.

Acknowledgement

The paper was written within the projects No 179010 and No 47020, supported by the Ministry of Education and Science of the Republic of Serbia. The authors are grateful to Louise Venables for her valuable comments and editing.

References

[1] K. Feldman: The Perceived Instructional Effectiveness of College Teachers as Related to Their Personality and Attitudinal Characteristics: A Review and Synthesis, Research in Higher Education, 24, 2, 1986, pp. 139-213

[2] S. Suplicz: What Makes a Teacher Bad? Trait and Learnt Factors of Teachers' Competencies, Acta Polytechnica Hungarica, 6, 3, 2009, pp. 125-138

[3] T. A. Lamke: Personality and Teaching Success. The Journal of Experimental Education, 20, 2, 1951, pp. 217-259

[4] R. B. Cattell, M. B. Marshall, S. Georgiades: Personality and Motivation: Structure and Measurement, Journal of Personality Disorders, 19, 1, 1957, pp. 53-67

[5] M. K. Mount, M. R. Barrick, G. L. Stewart: Five-Factor Model of Personality and Performance in Jobs Involving Interpersonal Interactions, Human Performance, 11, 1998, pp. 145-165

[6] A. Aidla, M. Vida: Personality Traits Attributed to Estonian School Teachers, Review of International Comparative Management, 11, 4, 2010, pp. 591-602

[7] J. M. Digman: Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology, 41, 1990, pp. 417-440

[8] J. P. Rushton, J. Morgan, M. Richard: Teacher's Myers-Briggs Personality Profiles: Identifying Effective Teacher Personality Traits, Teaching and Teacher Education, 23, 4, 2007, pp. 432-441

[9] C. G. Jung: Psychological Types. Collected Works of C.G. Jung, Volume 6, Princeton University Press, Princeton, New Jersey, 1971

[10] G. J. Boyle: Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations, Australian Psychologist, 30, 1995, pp. 71-74

[11] H. Q. Li, Z. Y. Wu: Comparative Study on the Personality Patterns of “Good Teacher” and “Bad Teacher” as Perceived by College Students, Advanced Materials Research, 271-273, 2011, pp. 760-763

[12] H. J. Eysenck: Dimensions of Personality: 16, 5 or 3?—CRITERIA for a Taxonomic Paradigm, Personality and Individual Differences, 12, 8, 1991, pp. 773-790

[13] C.-S. Liao, C.-W. Lee: An Empirical Study of Employee Job Involvement and Personality Traits: The case of Taiwan, International Journal of Economics and Management, 3, 1, 2009, pp. 22-36

[14] M. R. Barrick, M. K. Mount: The Big Five Personality Dimensions and Job Performance: A Meta-Analysis, Personnel Psychology, 44, 1, 1991, pp. 1-26

[15] T. A. Judge, D. Heller, M. K. Mount: Five-Factor Model of Personality and Job Satisfaction: A Meta-Analysis, Journal of Applied Psychology, 87, 3, 2002, pp. 530-541

[16] M. M. Sanchez, R. Martinez-Pecino, Y. T. Rodriguez, P. T. Menero: Student Perspectives on the University Professor Role, Social Behavior and Personality: An International Journal, 39, 4, 2011, pp. 491-496

[17] L. Genc, J. Pekić, A. Genc: Struktura ličnosti dobrog nastavnika u modelu Velikih pet (Personality Structure of a Good Teacher from Students Perspective according to Big-Five Model), [Manuscript submitted for publication]

[18] S. A. Basow: Student Evaluations of College Professors: When Gender Matters, Journal of Educational Psychology, 87, 4, 1995, pp. 656-665

[19] O. P. John, S. Srivastava: The Big-Five Trait Taxonomy: History, Mesurement, and Theoretical Perspectives. In L. A. Pervin, O. P. John (eds.): Handbook of personality: Theory and research, Guilford Press, New York, 1999, pp. 102-139

[20] P. Čolović: Tipološka prespektiva u psihologiji ličnosti: tradicionalni i taksometrijski pristup (Typological Perspective in Personality Psychology: Traditional and Taxometric Approaches), Unpublished doctoral dissertation, Novi Sad, University of Novi Sad, Faculty of Philosophy, 2012

[21] S. Rózsa: A BFQ Tesztkönyve. Oktatási segédanyag, Budapest, ELTE Személyiség- és Egészségpszichológiai Tanszék (The BFQ Testbook: Educational Supplement, Budapest, ELTE, Personality and Health Psychology Department), 2000

[22] O. P. John, E. M. Donahue, R. L. Kentle: The Big Five Inventory: Versions 4a and 54, Institute of Personality and Social Research, University of California, Berkley, 1991

[23] J. Cohen: Statistical Power Analyses for the Behavioral Sciences (2nd ed.), Erlbaum, Hillside, New Jersey, 1988

[24] L. Tenjović, S. Smederevac: Mala reforma u statističkoj analizi podataka: malo p nije dovoljno, potrebna je i veličina efekta (A Small Reform in the Data Analisys in Psychology: a Amll p is not Eough, Effect Size is Needed too), Primenjena psihologija, 4, 2011, pp. 317-333

[25] P. Tóth: Learning Strategies and Styles in Vocational Education, Acta Polytechnica Hungarica, 9, 3, 2012, pp. 195-216

Mathematical Basis of Sliding Mode Control of an Uninterruptible Power Supply

Károly Széll and Péter Korondi

Budapest University of Technology and Economics, Hungary
Bertalan Lajos u. 4-6, H-1111 Budapest, Hungary
szell@mogi.bme.hu; korondi@mogi.bme.hu

The sliding mode control of Variable Structure Systems has a unique role in control theories. First, the exact mathematical treatment represents numerous interesting challenges for the mathematicians. Secondly, in many cases it can be relatively easy to apply without a deeper understanding of its strong mathematical background and is therefore widely used in the field of power electronics. This paper is intended to constitute a bridge between the exact mathematical description and the engineering applications. The paper presents a practical application of the theory of differential equation with discontinuous right hand side proposed by Filippov for an uninterruptible power supply. Theoretical solutions, system equations, and experimental results are presented.

Keywords: sliding mode control; variable structure system; uninterruptable power supply

1 Introduction

Recently most of the controlled systems are driven by electricity as it is one of the cleanest and easiest (with smallest time constant) to change (controllable) energy source. The conversion of electrical energy is solved by power electronics [1]. One of the most characteristic common features of the power electronic devices is the switching mode. We can switch on and off the semiconductor elements of the power electronic devices in order to reduce losses because if the voltage or current of the switching element is nearly zero, then the loss is also near to zero. Thus, the power electronic devices belong typically to the group of variable structure systems (VSS). The variable structure systems have some interesting characteristics in control theory. A VSS might also be asymptotically stable if all the elements of the VSS are unstable itself. Another important feature that a VSS - with appropriate controller - may get in a state in which the dynamics of the system can be described by a differential equation with lower degree of freedom than the original one. In this state the system is theoretically completely independent of changing certain parameters and of the effects of certain external disturbances (e.g. non-linear load). This state is called sliding mode and the control based on this is called sliding mode control which has a very important role in the control of power electronic devices.

The theory of variable structure system and sliding mode has been developed decades ago in the Soviet Union. The theory was mainly developed by Vadim I. Utkin [2] and David K. Young [3]. According to the theory sliding mode control should be robust, but experiments show that it has serious limitations. The main problem by applying the sliding mode is the high frequency oscillation around the sliding surface, the so-called chattering, which strongly reduces the control performance. Only few could implement in practice the robust behavior predicted by the theory. Many have concluded that the presence of chattering makes sliding mode control a good theory game, which is not applicable in practice. In the next period the researchers invested most of their energy in chattering free applications, developing numerous solutions [4-8].

According to [9]: Uninterruptible power supplies (UPS's) are being broadly adopted for the protection of sensitive loads, like PCs, air traffic control system, and life care medical equipment, etc., against line failures or other ac mains' perturbations. Ideally, an UPS should be able to deliver: 1) a sinusoidal output voltage with low total harmonic distortion during normal operation, even when feeding nonlinear loads (particularly rectifier loads). 2) The voltage dip and the recovery time due to step load change must be kept as small as possible, that is, fast dynamic response. 3) The steady-state error between the sinusoidal reference and the load regulation must be zero. To achieve these, the Proportional Integral (PI) controller is usually used [10]. However, when the system using PI controller under the case of a variable load rather than the nominal ones, cannot obtain fast and stable output voltage response. In the literature there can be found some solutions to overcome this problem [11-15].

The structure of this paper is as follows. After the introduction, the second section summarizes the mathematical foundations of sliding mode control based on the theory of the differential equations with discontinuous right-hand sides, explaining how it might be applied for a control relay. The third section shows how to apply the mathematical foundations on a practical example. The fourth section presents experimental results of an uninterruptable power supply (UPS).

2 Mathematical Foundations of Sliding Mode Control

2.1 Introductory Example

The first example introduces a problem that can often be found in the engineering practice. Assume that there is a serial L-C circuit with ideal elements, which can be shorted, or can be connected to the battery voltage by a transistor switch (see Figure 1, where the details of the transistor switch are not shown). Assume that our reference signal has a significantly lower frequency than the switching frequency of the controller. Thus we can take the reference signal as constant.

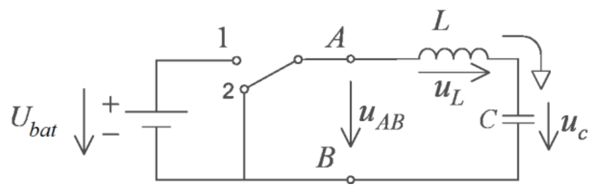


Figure 1
L-C circuit

Assume that we start from an energy free state, and our goal is to load the capacitor to the half of the battery voltage by switching the transistor. The differential equations for the circuit elements are:

$$C\frac{d}{dt}u_c = i_c \text{ and } L\frac{d}{dt}i_L = u_L \tag{1}$$

Due to the serial connection $i_c = i_L$, thus the differential equation describing the system is:

$$u_{AB} = u_c + LC\frac{d^2}{dt^2}u_c \tag{2}$$

Introduce relative units such way, that $LC = 1$ and $U_{bat} = 1$. Introduce the error signal voltage $u_e = U_r - u_c$, where $U_r = 1/2$ is the reference voltage of the capacitor. Thus, the differential equation of the error signal has the form:

$$u = u_e + \frac{d^2}{dt^2}u_e \tag{3}$$

,where

$$u = \begin{cases} \frac{1}{2}, & \text{if the switch is in state 1} \\ -\frac{1}{2}, & \text{if the switch is in state 2} \end{cases}$$

It is easy to see that the state belonging to the solution of equation (3) moves always clockwise along a circle on the phase plane $u_e, \frac{d}{dt}u_e$ (see Figure 2).

The center of the circle depends on the state of the transistor. The state-trajectory is continuous, so the radius of the circle depends on in what state the system is at the moment of the last switching. Assume that we start from the state

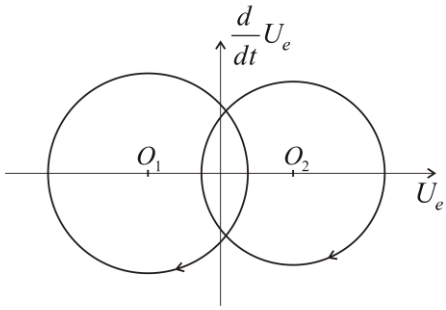


Figure 2

Possible state-trajectories

$u_e = 1/2, \frac{d}{dt}u_e = 0$ and our goal is to reach by appropriate switching the state $u_e = 0, \frac{d}{dt}u_e = 0$. Introduce the following switching strategy:

$$u = \frac{1}{2}\,\mathrm{sign}(s) \tag{4}$$

,where

$$s = u_e + \lambda \frac{d}{dt} u_e$$

This means that if the state-trajectory is over the $s = 0$ line, then we have to switch the circle centered at O_1, if it is below the line, then we have to switch the circle centered at O_2. Examine how we can remove the error. Consider Figure 3, according to (4) at first we start over the $s = 0$ line on a circle centered at O_1.

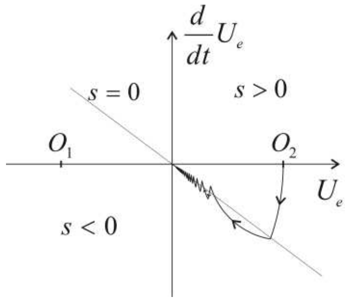


Figure 3
Removing the error

Reaching the line we switch to the circle centered at O_2 so that the state-trajectory remains continuous. After the second switching we experience an interesting phenomenon. As the state trajectory starts along the circle centered at O_1, it returns immediately into the area, where the circle centered at the O_2 has to be switched, but the state-trajectory cannot stay on this circle either, new switching is needed. For the sake of representation, the state-trajectory on Figure 3 reaches significantly over to the areas on both sides of the $s = 0$ line. In ideal case the state trajectory follows the $s = 0$ line on a curve broken in each points consisting of infinitely short sections switched by infinitely high frequency. In other words, the trajectory of the error signal slides along the $s = 0$ line and therefore is called sliding mode.

Based on the engineering and geometric approach we feel that after the second switching, the behavior of the error signal can be described by the following differential equation instead of the second order (3):

$$0 = u_e + \lambda \frac{d}{dt} u_e \tag{5}$$

This is particularly interesting because (5) does not include any parameter of the original system, but the λ we have given. Thus we got a robust control that by certain conditions is insensitive to certain disturbances and parameter changes. Without attempting to be comprehensive investigate the possible effects of changing some attributes and parameters of the system. If we substitute the ideal lossless elements by real lossy elements, then the state-trajectory instead of a circle moves along a spiral with decreasing radius. If the battery voltage fluctuates, the center of the circle wanders. Both changes affect the section before the sliding mode and modifies the sustainability conditions of the sliding mode, but in both cases, the sliding mode may persist (the state trajectory cannot leave the switch line), and if it persists, then these changes will not affect the behaviour of the sliding mode of the system.

The next section will discuss how we can prove our conjecture mathematically.

2.2 Solution of Differential Equations with Discontinuous Right-Hand Sides

Consider the following autonomous differential equation system:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)) \text{ and } \mathbf{x}(t = T_0) = \mathbf{X}_0 \tag{6}$$

,where

$\mathbf{x} \in \Re^n$

$\mathbf{f}(\mathbf{x}) : \Re^n \to \Re^n$

If **f(x)** is continuous then we can write (6) as the integral equation:

$$\mathbf{x}(t) = \mathbf{X}_0 + \int_{T_0}^{t} \mathbf{f}(\mathbf{x}(\tau))d\tau \tag{7}$$

The approach of (6) according to (7) is called Carathéodory solution, which under certain conditions may also exist when **f(x)** is discontinuous [16]. Recently, several articles and PhD theses dealt with it how to ease the preconditions which guarantee the existence of (7) concerning **f(x)**, but for the introduced example none of the cases might be applied, we need a completely different solution.

Filippov recommends a solution, which is perhaps closer to the engineering approach described in the previous section [17], [18]. Filippov is searching the solution of (6) at a given point based on how the derivative behaves in the neighborhood of the given point, allowing even that the behavior of the derivative may completely differ from its neighborhood on a zero set, and regarding the solution ignores the derivative on this set. Filippov's original definition concerns non-autonomous differential equations, but this article deals only with autonomous differential equations.

Consider (6) and assume that **f(x)** is defined almost everywhere on an open subset of $\Re^n$. Assume also that **f(x)** is measurable, locally bounded and discontinuous. Define the set *K(***x***)* for **x** $\subset$ **f(x)** by:

$$K(\mathbf{x}) := \bigcap_{\delta>0} \bigcap_{\mu N=0} \operatorname{conv} \mathbf{f}(Q(\mathbf{x},\delta) - N) \tag{8}$$

,where $Q(\mathbf{x},\delta)$ denotes the open hull with center **x** and the radius δ, μ is the Lebesgue measure, *N* is the Lebesgue null set and the word "conv" denotes the convex closure of the given set.

Filippov introduced the following definition to solve the discontinuous differential equation systems:

An absolutely continuous vector-valued function $\mathbf{x}(t) : [T_0, T_2] \to G \in \Re^n$ is the solution of (6) if

$$\frac{d}{dt}\mathbf{x} \subset K(\mathbf{x}(t)) \tag{9}$$

for almost every $t \subset [T_0, T_2]$. Note that if **f(x)** is continuous, then set *K(***x***)* has a single element for every **x**, namely **f(x)**, thus the definition of Filippov is consistent with the usual differential equations (with continuous right-hand side). However, if **f(x)** is not continuous, then this definition allows us searching a solution for (6) in such a domain of **x**, where **f(x)** is not defined.

2.3 Control Relays

Apply the definition of Filippov as a generalization of the introductory example in the case of such a controller with state feedback, where in the feedback loop only a relay can be found (see Figure 4). Assume that the state of the system can be described by the differential equation (6) where the vectorfunction **f(x)** is sbtanding on the right-hand side rapidly varies depending on the state of the relay. The control (switching) strategy should be the following. In the domain of the space $G \in \Re^n$ defined by the fedback state variables define an *n-1* dimensional smooth regular hypersurface *S* (which can also be called as switching surface) using continuous $s(\mathbf{x}): \Re^n \to \Re$ scalar-vector function in the following way:

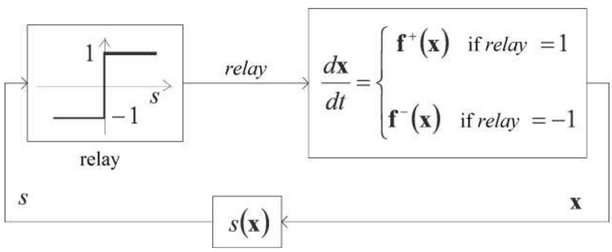


Figure 4
Controller with relay

$$S := \{\mathbf{x} : s(\mathbf{x}) = 0\} \tag{10}$$

The goal of the controller is to force the state-trajectory to this surface. Mark the points of the surface *S* with $\mathbf{x}_s$. With the help of this surface we can divide the domain *G* into two parts:

$$G^+ := \{\mathbf{x} : s(\mathbf{x}) > 0\} \tag{11}$$

$$G^- := \{\mathbf{x} : s(\mathbf{x}) < 0\}$$

Let the differential equation for **x** on domain *G* and our switching strategy have the following form:

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}) = \begin{cases} \mathbf{f}^{+}(\mathbf{x}), & \text{if} \quad \mathbf{x} \in G^{+} \\ \mathbf{f}^{-}(\mathbf{x}), & \text{if} \quad \mathbf{x} \in G^{-} \end{cases} \tag{12}$$

,where $\mathbf{f}^{+}(\mathbf{x})$ and $\mathbf{f}^{-}(\mathbf{x})$ are uniformly continuous vector-vector functions.

Note that $\mathbf{f}(\mathbf{x})$ is not defined on the surface *S*, and we did not specify that $\mathbf{f}^{+}(\mathbf{x})$ and $\mathbf{f}^{-}(\mathbf{x})$ must be equal on both sides of the surface *S*.

Outside the surface *S* we have to deal with an ordinary differential equation. Solution of (12) might be a problem in the $\mathbf{x_s(t)}$ points of the surface *S*. According to definition (9), K is the smallest closed convex set, which you get in the following way: let's take an arbitrary $\delta > 0$ hull of all $\mathbf{x_s(t)}$ points of the $\mathbf{x_s}$ belonging to the surface *S*, exclude $\mathbf{f(x_s)}$, where $\mathbf{f(x)}$ is not defined (remark: it is a null set *N* domain), and we complete the set of $\mathbf{f(x)}$ vectors belonging to the resulting set to a closed convex set. Obviously, the smaller the value of $\delta > 0$ is, the smaller the resulting closed convex set is. Finally, we need to take the intersection of the closed convex sets in the hull of all $\delta > 0$ and *N*. Since $\mathbf{f(x)}$ is absolutely continuous, the following limits exist at any point of the surface *S*:

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{x_s}} \mathbf{f}(\mathbf{x}) &= \mathbf{f}^{+}(\mathbf{x}_S) \quad \text{if} \quad \mathbf{x} \in G^{+} \\ \lim_{\mathbf{x}\to\mathbf{x_s}} \mathbf{f}(\mathbf{x}) &= \mathbf{f}^{-}(\mathbf{x}_S) \quad \text{if} \quad \mathbf{x} \in G^{-} \end{aligned} \tag{13}$$

It means that the $\bigcap\limits_{\delta>0}\bigcap\limits_{\mu N=0}\mathbf{f}(Q(\mathbf{x},\delta)-N)$ set belonging to any point $\mathbf{x_s(t)}$ of the surface *S* has only two elements, $\mathbf{f}^{+}(\mathbf{x_s})$ and $\mathbf{f}^{-}(\mathbf{x_s})$. We have to take the convex closure of these two vectors, which will be the smallest subset belonging to all $\delta > 0$ values. In summary, the differential equation (6) with a (12) form discontinuity in the $\mathbf{x_s(t)}$ points of *S* surface according to definition (9) can be described in the following form:

$$\frac{d}{dt}\mathbf{x}_S = \zeta\mathbf{f}^{+}(\mathbf{x}_S) + (1-\zeta)\mathbf{f}^{-}(\mathbf{x}_S) \tag{14}$$

To illustrate (14) see Figure 5, where we drew $\mathbf{f}^{+}(\mathbf{x_{SP}})$ and $\mathbf{f}^{-}(\mathbf{x_{SP}})$ vectors belonging to the point *P* of surface *S*. We marked the normal vector belonging to the point *P* of the surface with n_p. The change of the state trajectory in point $\mathbf{x_{SP}}$ is given by the equivalent vector $\mathbf{f_{eq}(x_{SP})}$, which is the convex sum of vectors $\mathbf{f}^{+}(\mathbf{x_{SP}})$ and $\mathbf{f}^{-}(\mathbf{x_{SP}})$.

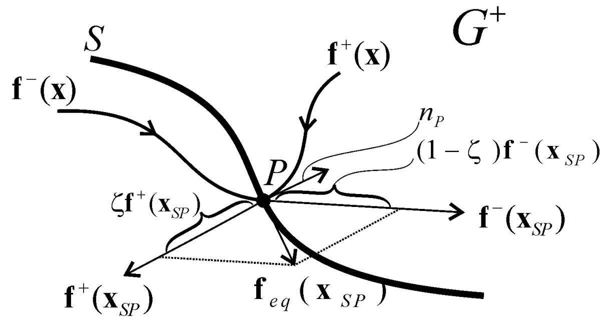


Figure 5

The state trajectory sliding along the surface *S*

Denote by $L_f s(\mathbf{x})$ the directional derivative of the scalar function $s(\mathbf{x})$ concerning the vector space $\mathbf{f(x)}$:

$$L_f s(\mathbf{x}) = (\mathbf{f}(\mathbf{x}) \bullet \text{grad}(s(\mathbf{x}))) \tag{15}$$

,where ($\mathbf{a} \bullet \mathbf{b}$) denotes the scalar product of vectors **a** and **b**. Since s(**x**) is uniformly continuous, the following limits exist at any point of the surface *S*:

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{x}_s} L_f s(\mathbf{x}) &= L_{f^+} s(\mathbf{x}_S) \quad \text{if} \quad \mathbf{x} \in G^+ \\ \lim_{\mathbf{x}\to\mathbf{x}_s} L_f s(\mathbf{x}) &= L_{f^-} s(\mathbf{x}_S) \quad \text{if} \quad \mathbf{x} \in G^- \end{aligned} \tag{16}$$

The value of ζ should be defined such that $\frac{d}{dt}\mathbf{x}_S$ and $\mathbf{f_{eq}(x_S)}$ are orthogonal to the normal vector of the surface *S* (see Filippov's 3. Lemma [18]):

$$L_f s(\mathbf{x}_S) = (\mathbf{f}_{eq}(\mathbf{x}_S) \bullet \text{grad}(s(\mathbf{x}_S))) = 0 \tag{17}$$

The equation (17) can be interpreted in the following way: in sliding mode, in the $\mathbf{x_s}$ points of the sliding surface the change of the state trajectory can be described by an equivalent $\mathbf{f_{eq}(x_S)}$ vector function that satisfies condition (17). Based on (14) and (17), we obtain

$$\zeta L_{f^+} s(\mathbf{x}_S) + (1-\zeta) L_{f^-} s(\mathbf{x}_S) = 0 \tag{18}$$

ζ can be expressed from (18):

$$\zeta = \frac{L_{f^-} s(\mathbf{x}_S)}{L_{f^-} s(\mathbf{x}_S) - L_{f^+} s(\mathbf{x}_S)} \tag{19}$$

If $L_{f^+} s(\mathbf{x}_S) < 0$ and $L_{f^-} s(\mathbf{x}_S) > 0$ then on both sides of the surface *S* the vector space $\mathbf{f(x)}$ points towards the surface *S* (see Figure 6). Therefore, if the state trajectory once reaches the surface *S*, it can not leave it. The state trajectory slides along the surface and therefore this state is called sliding mode.

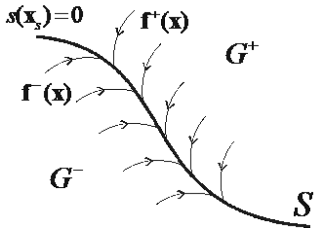


Figure 6
The f(x) vector space pointing towards the surface *S*

Note that the two conditions separately defined on both sides of the surface *S*.

$$L_{f^+} s(\mathbf{x}) < 0, \quad \text{if} \quad \mathbf{x} \in Q(\mathbf{x}_S, \delta) \cap G^+ \tag{20}$$

$$L_{f^-} s(\mathbf{x}) > 0, \quad \text{if} \quad \mathbf{x} \in Q(\mathbf{x}_S, \delta) \cap G^-$$

can be substituted by one inequality:

$$\frac{d}{dt} s(\mathbf{x})^2 < 0, \quad \text{if} \quad \mathbf{x} \in Q(\mathbf{x}_S, \delta) - S \tag{21}$$

Which can be interpreted as Lyapunov's stability criterion concerning whether the system remains on the surface *S*.

3 The Solution of the Differential Equation of the Introductory Example

Steps presented in Section II are applied for a circuit seen in Figure 1. As later on we can see our experimental setup for UPS (shown in Fig. 7) can be modeled with a simple L-C circuit.

There are two energy storage elements (L and C) in the circuit of the introductory example, therefore the behavior of the circuit can be described by two state variables. The goal is to remove the voltage error, so it is practical to choose the error signal and the first time derivative of it as the state variables.

$$\mathbf{x} = \begin{bmatrix} u_e \\ \frac{d}{dt} u_e \end{bmatrix} = \begin{bmatrix} U_r - u_c \\ -\frac{d}{dt} u_c \end{bmatrix} \tag{22}$$

The state equation of the error signal, assuming that the reference signal U_r is constant:

$$\frac{d}{dt}\begin{bmatrix} u_e \\ \frac{d}{dt}u_e \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{LC} & a_{22} \end{bmatrix}\begin{bmatrix} u_e \\ \frac{d}{dt}u_e \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{LC} \end{bmatrix}(U_r - u_{AB}) \tag{23}$$

,where $a_{22} = 0,$ if we neglect the losses assuming ideal *L-C* elements, while $a_{22} = -R/L$, if we model the losses of the circuit with serial resistance.

The design of a sliding mode controller (SMC) consists of three main steps. First, the design of the sliding surface the second step is the design of the control law which holds the system trajectory on the sliding surface and the third and key step is the chattering-free implementation [19].

Based on (4), let the scalar function defining the sliding surface be (first step in the design of the SMC):

$$s(\mathbf{x}) = \begin{bmatrix} 1 & \lambda \end{bmatrix}\begin{bmatrix} u_e \\ \frac{d}{dt}u_e \end{bmatrix} \tag{24}$$

Rewriting the matrix equation (23) to the form (12), we obtain (second step in the design of the SMC):

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}) = \begin{cases} \mathbf{f}^{+}(\mathbf{x}) = \begin{bmatrix} f_1 \\ f_2 - u_+ \end{bmatrix}, & \text{if the switch is in state 1} \\ \mathbf{f}^{-}(\mathbf{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, & \text{if the switch is in state 2} \end{cases} \tag{25}$$

,where

$$f_1 = \frac{d}{dt}u_e$$

$$f_2 = -\frac{1}{LC}u_e - a_{22}\frac{d}{dt}u_e + \frac{1}{LC}U_r \tag{26}$$

$$u_+ = \frac{1}{LC}U_{bat}$$

The directional derivative of the scalar function s(x) concerning the vector space f(x) on both sides of the surface S is:

$$L_{f^+}s(\mathbf{x}_s) = \frac{d}{dt}u_{es} + \lambda(-\frac{1}{LC}u_{es} - a_{22}\frac{d}{dt}u_{es} - u_+) = f_{1s} + \lambda(f_{2s} - u_+) \tag{27}$$

$$L_{f^-} s(\mathbf{x}_s) = \frac{d}{dt}u_{es} + \lambda(-\frac{1}{LC}u_{es} - a_{22}\frac{d}{dt}u_{es}) = f_{1s} + \lambda f_{2s}$$

Note that in our case **f** $^+$**(x)** and **f** $^-$**(x)** can be defined on the surface *S*, therefore we do not need to calculate the limits in (16), the points $u_{es}, \frac{d}{dt}u_{es}$ belonging to the surface *S* can be directly substituted. At the same time, (21) is met only in the following domain:

$$0 < \frac{d}{dt}u_{es} + \lambda(-\frac{1}{LC}u_{es} - a_{22}\frac{d}{dt}u_{es}) < u_+ \tag{28}$$

It means that, by the given control relay, only on a limited part of the surface *S* can be in sliding mode. By completing the relay control laws with additional members, we can reach that the condition of sliding mode is fulfilled on the whole *S* surface [20]. In case of the control relay, based on (19) and (27), we get:

$$\zeta = \frac{f_{1s} + \lambda f_{2s}}{\lambda u_+} \tag{29}$$

Based on (14), (25), (27), and (29) the differential equation describing the system in sliding mode will be:

$$\frac{d}{dt}\begin{bmatrix} u_{es} \\ \frac{d}{dt}u_{es} \end{bmatrix} = \frac{f_{1s} + \lambda f_{2s}}{\lambda u_+}\begin{bmatrix} f_{1s} \\ f_{2s} - u_+ \end{bmatrix} + \frac{\lambda u_+ - f_{1s} - \lambda f_{2s}}{\lambda u_+}\begin{bmatrix} f_{1s} \\ f_{2s} \end{bmatrix} = \begin{bmatrix} f_{1s} \\ -\frac{1}{\lambda}f_{1s} \end{bmatrix} \tag{30}$$

The differential equation (30) is basically the same as the equation of the sliding line, and thus we can describe the original system as a first order differential equation instead of a second order one:

$$s(\mathbf{x}_s) = 0 = u_{es} + \lambda\frac{d}{dt}u_{es} \tag{31}$$

This way we proved, that the state belonging to the smooth regular sliding line *S* can be accurately followed by a state trajectory broken in each points consisting of infinitely short sections switched by infinitely high frequency. The solution of (30) is:

$$u_{es} = U_{es0}e^{-\frac{1}{\lambda}t} \tag{32}$$

,where U_{es0} is the u_{es} error signal at the moment when the state trajectory reaches the surface *S*. Based on (32) we can see that λ is the characteristic time constant of the sliding line. Note that equation (32) does not include any parameter of the original system. This means that in the above-described ideal sliding mode the relay control law leads to a robust controller, insensitive to certain parameters and disturbances. The above derivation is only concerned with how the system behaves on the sliding surface, but we did not deal with the practically very important question of how to ensure that the state trajectory always reaches the sliding surface and stays on it.

Of course, in reality such an ideal sliding mode does not exist. From engineering point of view the challenge is the realization of a so-called chattering-free approximation of it.

4 Uninterruptible Power Supply

The measurement setup is an uninterruptable power supply. A simplified diagram of the inverter and the filter is shown in Figure 7. This circuit can be modelled as a simple L-C circuit seen in Figure 1.

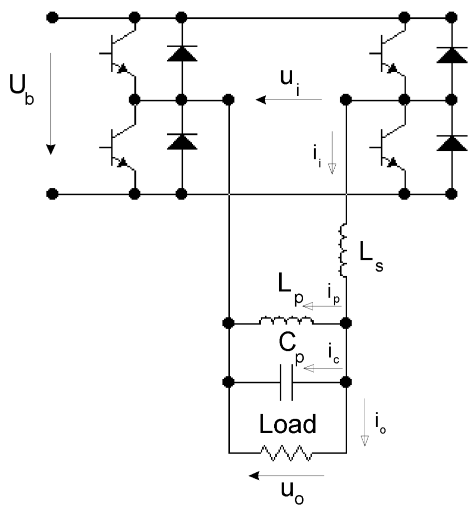


Figure 7

Simplified diagram of the inverter

Ideally the sliding mode needs an infinitely high switching frequency. The frequency is limited by a hysteresis relay. The minimum interval between two switches is determined by the hysteresis. Because of the hysteresis the error signal trajectory is chattering around the origin and harmonics will appear in the output voltage. As the third step in the design of a sliding mode controller to realize a chattering-free implementation, a proper filter is used with respect to the hysteresis, thus harmonic distortion remains at an acceptable level. Positive and negative half periods can be separated by dead-band (see Figure 8).

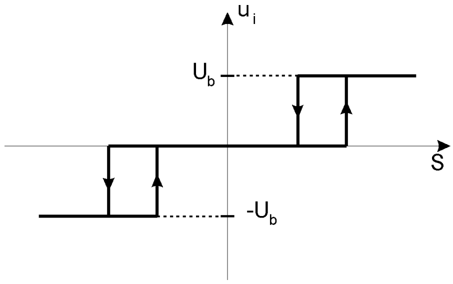


Figure 8
Dead-band hysteresis relay

As a first step the equations of the UPS shown in Figure 7 are rearranged to the form described in Section III. The load is connected to the inverter through a transformer. The L_s in Fig. 7 is the leakage inductance of the transformer, which has a special structure to increase and set the value of L_s. The main field inductance Lp cannot be ignored from the point of view of the resonant circuit. Because of the main field inductance instead of (2) the filter circuit can be described by the following differential equation:

$$Bu_i = BL_s C_p \frac{d^2 u_o}{dt^2} + u_o \tag{33}$$

,where

$$B = \frac{L_p}{L_s + L_p} \tag{34}$$

Let us use per unit system:

$$t_{rel} = \frac{1}{\sqrt{BL_s C_p}} t \tag{35}$$

From now on we calculate only in per unit system and do not denote it. The differential equation (33) in per unit system:

$$Bu_i = \ddot{u}_o + u_o \tag{36}$$

, where $\bullet$ is the operator: d/dt. (36) corresponds to (3). Thus we can use the results described in Section III. The difference in this case is that we have more switching states and the reference signal is a sine wave. Take the influence of this sine wave into consideration.

If u_{ref} is the reference signal, we obtain the following equations:

$$u_{ref} = U_{ref} \sin(\omega t + \varphi) \tag{37}$$

$$u_e = u_{ref} - u_o \tag{38}$$

$$\ddot{u}_e = -\omega^2 u_{ref} - \ddot{u}_o \tag{39}$$

$$\ddot{u}_e + u_e = (1-\omega^2) u_{ref} - Bu_i \tag{40}$$

Solution of the differential equation:

$$u_e = A_1 \cos t + A_2 \sin t + u_{ref} - Bu_i \tag{41}$$

Values of constants A_1 and A_2 can be determined using the initial conditions:

$$A_1 = U_{e0} - U_{ref} \sin\varphi + Bu_i \tag{42}$$

$$A_2 = \dot{U}_{e0} - \omega U_{ref} \cos\varphi \tag{43}$$

,where $u_e(t=0)=U_{e0}$ and $\dot{u}_e(t=0) = \dot{U}_{e0}$. Using (41), the error signal trajectory can be described as follows:

$$(u_e - u_{ref} + Bu_i)^2 + (\dot{u}_e - \dot{u}_{ref})^2 = A_1{}^2 + A_2{}^2 \tag{44}$$

The curve defined by (44) can be plotted as a $(A_1^2+A_2^2)^{1/2}$ radius circle in the phase plane $u_e - \dot{u}_e$, the center of which runs along an ellipse given by the parameter equation system below:

$$u_e = U_{ref} \sin(\omega t + \varphi) - Bu_i \tag{45}$$

$$\dot{u}_e = \omega U_{ref} \cos(\omega t + \varphi) \tag{46}$$

The ellipse has three possible positions in the phase plane, depending on the value of u_i (see Figure 9).

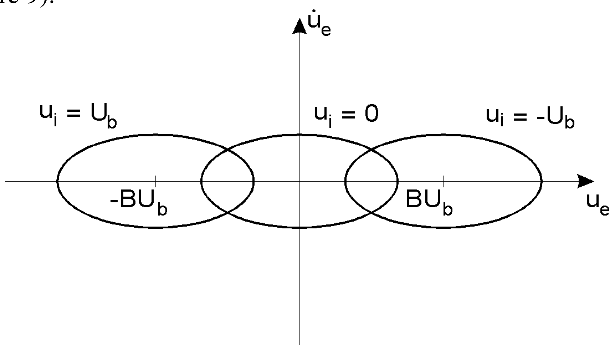


Figure 9
Ellipses describing the movement of the center of the circle

4.1 Robustness Analysis

Let us check the condition of the sliding mode based on (21). Deriving (24) and eliminating the second order term using (40), we obtain:

$$\dot{s}(\mathbf{x}) = \dot{u}_e + \lambda[U_{ref}(1-\omega^2)\sin(\omega t+\varphi) - Bu_i - u_e] \tag{47}$$

(47) corresponds to (5). According to Lyapunov's stability, if $s(\mathbf{x})>0$ (similar to (28)):

$$\dot{u}_e - \lambda u_e < -\lambda U_{ref}(1-\omega^2)\sin(\omega t+\varphi) + \lambda Bu_i \tag{48}$$

If $s(\mathbf{x})<0$:

$$\dot{u}_e - \lambda u_e > -\lambda U_{ref}(1-\omega^2)\sin(\omega t+\varphi) + \lambda Bu_i \tag{49}$$

Both inequalities define a half plane, and the boundaries are perpendicular to the sliding surface changing their position according to a sine wave. The phase plane is divided into four sections (see Figure 10).

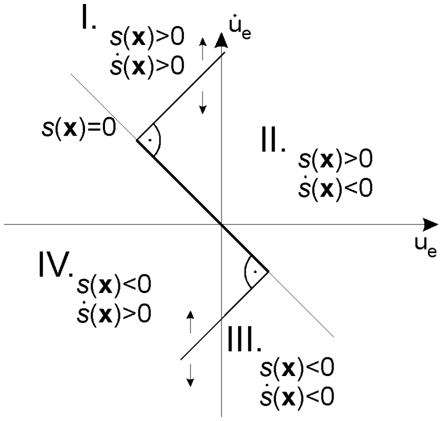


Figure 10
Phase plane of the error signal

Introduce the following switching strategy:

$$u_i = \begin{cases} U_b, & \text{if } s(\mathbf{x}) > 0 \\ -U_b, & \text{if } s(\mathbf{x}) < 0 \end{cases} \tag{50}$$

Usually $U_{ref} < U_b$ and $\omega < 1$. Thus the half planes defined by (48) and (49) always have an intersection which goes through the origin where (21) is valid on both sides of the switching surface. The sliding mode occurs on the common boundary of sections II. and IV.

4.2 Experimental Verification

Using the sliding mode control an experimental measurement is carried out for the UPS. The parameters of the experimental setup:

S_n= 10 kVA	U_n= 230 V	ω= 314 rad/sec	λ= 1/24
B= 0,9	L_s= 0,35 mH	C_p= 3200 μF	L_p= 3,2 mH

Based on (32) and the per unit system (35) the characteristic time constant of the sliding line:

$$T = \sqrt{BL_sC_p}\lambda = 42\mu\sec \tag{51}$$

During the measurement the current and the voltage of the load are measured. Operational amplifiers are used to provide the value of *S(t)*, which signal of the controller is measured with an oscilloscope. Using the notation of Figure 7 the measurements with the oscilloscope are shown in Figure 11 and 13. The harmonic distortion is less than 1.5%. The error trajectory cannot be measured directly, thus they are modeled by computer simulation. Figure 12 shows the steady-state trajectory and switching lines of the dead-band hysteresis relay. Switching delays can also been seen.

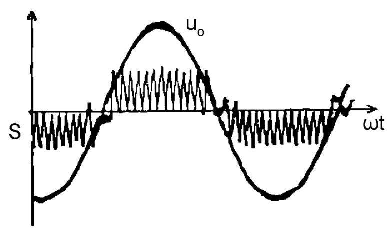


Figure 11
Experimental results for steady-state

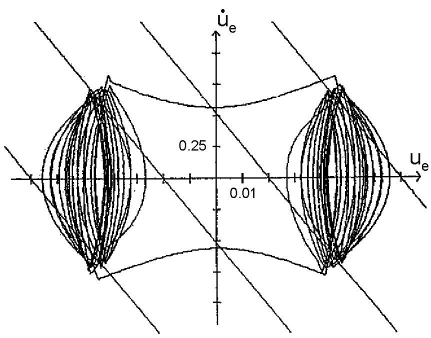


Figure 12
Steady-state error-trajectory

Figure 13 and Figure 14 show the system's response for a 100 % ohmic step change in the load. As we can see first the system has to reach the sliding surface and then it can follow the switching strategy of the case above. It takes only some switching, about 10% of the time period until the system settles.

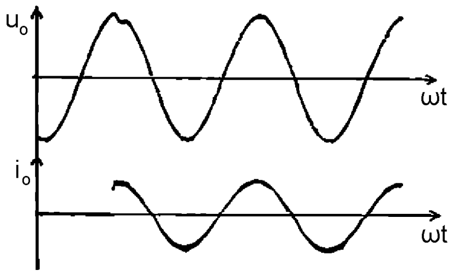


Figure 13
Experimental results for step-change

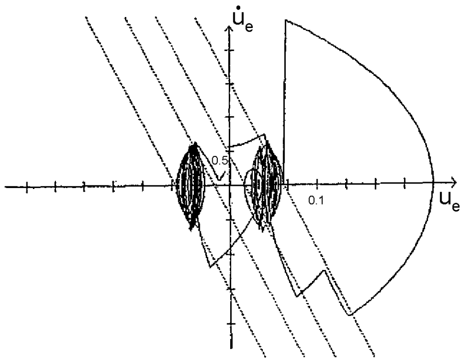


Figure 14
Step-change error-trajectory

Acknowledgements

The authors wish to thank the support to the Hungarian Automotive Technicians Education Foundation, to the Hungarian Research Fund (OTKA K100951), and the Control Research Group of HAS. The results discussed above are supported by the grant TÁMOP-4.2.2.B-10/1--2010-0009.

Conclusions

When a VSS is in sliding mode its trajectory lies on the switching line. The theoretical sliding mode is an idealization. In practice, when switching delays are present the steady state trajectory chatters around the origin. In spite of the chattering the harmonic distortion is less than 2%. The response time for step change of the load is very short due to the instantaneous nature of the sliding mode control. Apart from the short transient process the uninterruptable power supply is insensitive to the load variation. Though sliding mode controller is very simple, it provides a good performance. By the application of the definition of Filippov to the UPS, the paper presented a practical application of the theory of differential equation with discontinuous right hand side proposed by Filippov

References

[1] P. Bauer, R. Schoevaars: Bidirectional Switch for a Solid State Tap Changer, 34th Annual IEEE Power Electronics Specialists Conference. Vol. 1-4, 2003, pp. 466-471

[2] V. I. Utkin: Variable Structure Control Optimization, Springer-Verlag, 1992

[3] K. D. Young: Controller Design for Manipulator using Theory of Variable Structure Systems, IEEE Trans. on System, Man and Cybernetics, Vol. SMC-8, 1978, pp. 101-109

[4] H. Yu, U. Ozguner: Adaptive Seeking Sliding Mode Control, IEEE American Control Conference, Minneapolis, USA, 2006

[5] En-Chih Chang, Fu-Juay Chang, Tsorng-Juu Liang, Jiann-Fuh Chen: DSP-Based Implementation of Terminal Sliding Mode Control with Grey Prediction for UPS Inverters, the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA) , 2010, pp. 1046-1051

[6] R. A. Hooshmand, M. Ataei, M. H. Rezaei: Improving the Dynamic Performance of Distribution Electronic Power Transformers Using Sliding Mode Control, Journal of Power Electronics, Vol. 12, 2012, pp. 145-156

[7] F. F. A. van der Pijl, M. Castilla, P. Bauer: Adaptive Sliding-Mode Control for a Multiple-User Inductive Power Transfer System Without Need for Communication, IEEE Transactions on Industrial Electronics, Vol. 60, 2013, pp. 271-279

[8] Yung-Tien Liu, Tien-Tsai Kung, Kuo-Ming Chang, Sheng-Yuan Chen: Observer-based Adaptive Sliding Mode Control for Pneumatic Servo System, Precision Engineering 37, 2013, pp. 522-530

[9] Tsang-Li Tai, Jian-Shiang Chen: UPS Inverter Design Using Discrete-Time Sliding-Mode Control Scheme, IEEE Trans. Ind. Electron., Vol. 49, No. 1, Feb. 2002

[10] R. Caceres, R. Rojas, O. Camacho: Robust PID Control of a Buck-Boost UPS Converter, Proc. IEEE INTELEC'02, 2000, pp. 180-185

[11] Khalifa Al-Hosani, Andrey Malinin, Vadim I. Utkin: Sliding Mode PID Control of Buck Converters, The European Control Conference, 2009

[12] Hongmei Li, Xiao Ye: Sliding-Mode PID Control of DC-DC Converter, the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA) 2010, pp. 730-734

[13] S. Saraswathy, K. Punitha, D. Devaraj: Implementation of Current Control Techniques for Uninterruptable Power Supply, International Conference on Circuits, Power and Computing Technologies, 2013, pp. 589-595

[14] F. Tahri, A. Tahri, A. Allali, S. Flazi: The Digital Self-Tuning Control of Step a Down DC-DC Converter, Acta Polytechnica Hungarica, Vol. 9, No. 6, 2012, pp. 49-64

[15] A. Tahri, H. M. Boulouiha, A. Allali, T. Fatima: A Multi-Variable LQG Controller-based Robust Control Strategy Applied to an Advanced Static VAR Compensator, Acta Polytechnica Hungarica, Vol. 10, No. 4, 2013, pp 229-247

[16] D. C. Biles, P. A. Binding: On Caratheodory's Conditions for the Initial Value Problem, Proceedings of the American Mathematical Society, Volume 125, Number 5, May 1997, pp. 1371-1376

[17] A. G. Filippov: Application of the Theory of Differential Equations with Discontinuous Right-hand Sides to Non-linear Problems in Automatic Control, 1st IFAC Congr. Moscow, 1960, pp. 923-925

[18] A. G. Filippov: Differential Equations with Discontinuous Right-hand Side, Ann. Math Soc. Transl., Vol. 42, 1964, pp. 199-231

[19] B. Takarics, G. Sziebig, B. Solvang, P. Korondi: Multimedia Educational Material and Remote Laboratory for Sliding Mode Control Measurements, Journal of Power Electronics, Vol. 10, 2010, pp. 635-642

[20] F. Harashima, T. Ueshiba, H. Hashimoto: Sliding Mode Control for Robotic Manipulators, 2nd Eur. Conf. on Power Electronics, Grenoble Proc., 1987, pp. 251-256

A Fuzzy Logic Approach to Model and Predict HV Cable Insulation Behaviour under Thermal Aging

Lakhdar Bessissa[1], Larbi Boukezzi[2], Djillali Mahi[1]

[1] Laboratory of Studies and Development of Semiconductor and Dielectric Materials, LeDMaScD, University Amar Telidji of Laghouat, BP 37G route of Ghardaïa, Laghouat 03000, Algeria
bsisa.lakhdar@email.univ-djelfa.dz, d.mahi@mail.lagh-univ.dz

[2] Materials Science and Informatics Laboratory, MSIL, University Ziane Achour of Djelfa, BP 3117 route of Moudjbara, Djelfa 17000, Algeria
lboukezzi@email.univ-djelfa.dz

Abstract: Given the non-linearity of changes in the dielectric and mechanical properties of insulation, it is difficult to find neither a theoretical nor an experimental model for the aging mechanism. Hence, the need to model the evolution of mechanical properties of the XLPE material under thermal aging using new methods is necessary. The use of fuzzy logic is one of these methods and presents an original idea in high voltage prediction problems. Several factors were considered, namely the aging time and the level of applied temperature. The study of the impact of these parameters on the quality of the fuzzy algorithm is carried out. The characteristics of the modelling function (trapezoidal or triangular) and the number of membership functions partitions was also taken into consideration. The obtained results were encouraging. It has been proved that fuzzy logic is a powerful tool that can be used in predicting the properties of insulation. Furthermore, the analytical findings were in good agreement with the experimental results with an acceptable error margin.

Keywords: Thermal aging; Crosslinked polyethylene (XLPE); Fuzzy logic; Fuzzy Inference System (FIS); Number of partitions

1 Introduction

Electrical power systems include a large number of expensive and important high voltage cables of different ages, manufactured and mounted during decades. Repair and replacement of important cable system are expensive and correct scheduling for those, would give large saving of costs. Insulation system of high voltage power cable and their accessories are subjected to different kind of stress during their life, and thus to degradation and deterioration. These can lead to a reduction of life and so to a lower reliability of electrical power systems [1, 2]. Now that such insulating materials have been in service for several decades the opportunity exists to ascertain their current condition so that decisions can be taken regarding replacement.

The great importance of XLPE as insulation encourages researchers in laboratories through the world to investigate a lot of experimental techniques in order to get more information and characterize well the degradation mechanisms of the material under service conditions. Recent investigations aimed with the effect of thermal and electrical stresses on the endurance capability of the insulation materials [3], degradation mechanisms of insulation cable materials under thermal and radiation aging [4] and the effect of water tree aging on the properties of cables insulation [5]. These investigations are costly and time consuming. They need sometimes few years to get sufficient database to solve economical problems of energy and making maintenance in a simplest way. This obstacle forces researchers to find powerful modelling methods to solve this problem. Recently, several intelligent systems have been developed which help scientist and engineer to use in efficient way the database and get with more precision future states of the insulation system. Among these intelligent systems, artificial neural networks (ANN) and fuzzy logic (FL) present a powerful tools to predict and diagnosis high voltage insulation materials. In recent studies, a novel extension neural network algorithm was used in recognition of partial discharge pattern in high voltage power apparatus [6]. Other application aimed with classification and diagnostic of transformer oil [7]. However a small amount of papers aimed with applying of FL in high voltage field can be found [8].

The objective of this work is to develop a fuzzy logic algorithm using the non-linear properties variations of the material versus aging time and for different thermal constraint values, to predict XLPE insulation behaviour when the learning time is clearly less than the laboratory test time. We shall note that the use of FL in this paper presents an original idea in high voltage prediction problems. This study treats also the impact of the modelling function characteristics (trapezoidal or triangular) and the number of membership functions partitions on the quality of the fuzzy algorithm. Parts of the obtained results are recently presented in several international conferences [9, 10].

2 Experimental Setup

The used material is XLPE UNION CARBIDE 4201 which is used as an insulator in medium and high voltage cables. Plates of 2 mm thickness were moulded using a heat press machine. According to the International Electrotechnic Committee (IEC) 540 publication [11], dumbbell shape samples of 7.5 cm length were cut and will be performed to mechanical tests. The thermal aging experiments were carried out in the forced air ventilating oven which can maintain the average temperature of sample within ±2°C. Three temperatures have been considered namely 80°C, 100°C and 120°C. The full aging time is 5000 hours for 80°C and 100°C and 2500 hours for 120°C. After each 500h aging time, 10 samples were taken, and then subjected to tensile stress. The experiments consist in breaking the sample, at ambient temperature, using a dynamometer which moves with speed of 50 mm/min. This test allows us to measure the tensile strength of the material. The variation of tensile strength according to aging time is presented in Figure 1[12].

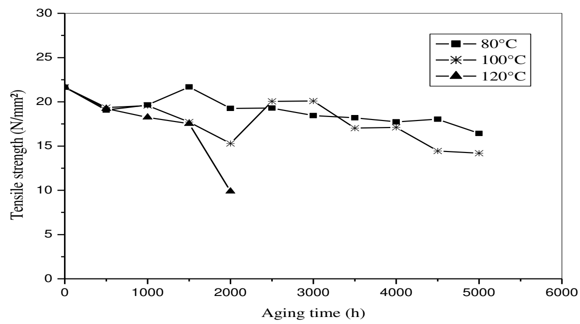


Figure 1
Variation of the tensile strength in function of the aging time

This figure shows that tensile strength decreases as a function of aging time. This decrease is even more pronounced as the temperature is higher. The mechanisms which are responsible for this depletion are generally the thermo-oxidative degradation, the chains' break, the decrease of the crosslinking degree and the variation in the crystallinity degree. For aging temperatures 80°C and 100°C, this property is stored and remains substantially higher than the lower limit value (12.5 N/mm²) required by IEC 502 [13]. However, for other temperatures, the degradation was reached after relatively short time.

3 Modelling and Prediction by Fuzzy Logic

3.1 Fuzzy Logic

The concept of ''fuzzy set" was preliminarily introduced by Zadeh [14, 15], who pioneered the development of fuzzy logic (FL) replacing Aristotelian logic which has two possibilities only. Zadeh has motivated his work on FL with the observation that the key elements in human thinking are not numbers but levels of fuzzy sets [15, 16]. Fuzzy approach performs numerical computation by using linguistic labels stimulated by membership functions. In the learning process, membership functions characterize the fuzziness in a fuzzy set, whether the elements in the set are discrete or continuous in a graphical form for eventual use in the mathematical formalism of fuzzy set theory [17, 18]. Although FL was brought forward by Zadeh [14, 15] in 1965, fuzzy concepts and systems attracted attention after a real control application in 1975 conducted by Mamdani and Assilian [15, 19, 20]. For control applications, fuzzy logic operations that include comparison of two or more membership functions are needed [21]. Figure 2 shows a typical membership function for small, medium and large class sizes in a universe U.

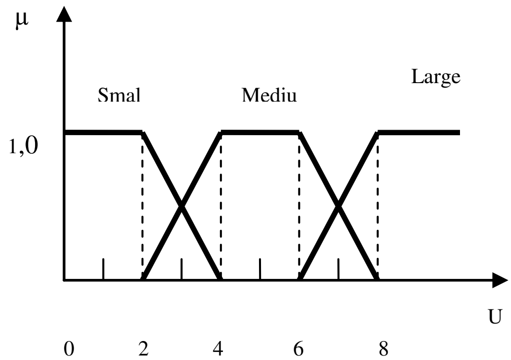


Figure 2
Fuzzy subsets

3.2 Fuzzy Inference System

Basically, a fuzzy inference system is composed of five functional blocks as follows [22, 23]:

(i) A rule base containing a number of fuzzy IF-THEN rules. The general form of a fuzzy IF-THEN rule is as follows; Rule: IF Z is A THEN F is B.

(ii) A database, which defines the membership functions of the fuzzy sets used in the fuzzy rules.

(iii) A decision-making unit, which performs the inference operations on the rules.

(iv) A fuzzification inference, which transforms the crisp inputs into degree of match with linguistic values.

(v) A defuzzification inference, which transforms the fuzzy results of the inference into a crisp output. These functional blocks are shown in Figure 3.

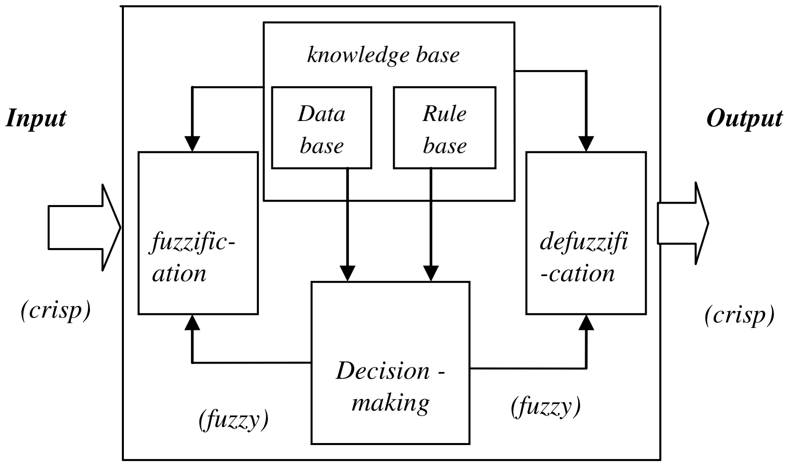


Figure 3
Block diagram for a fuzzy inference system

3.3 Modelling and Prediction

In our case, the input fuzzy variables are: the level of temperature T which takes 80 °C, 100 °C and 120 °C as values, and the aging time t. The output variable is the tensile strength.

Three fuzzy labels were attributed to the variations of these variables and various combinations were distinctively created:

Case 1: 3 fuzzy sets were created, S, M and L, which means that the variations are respectively Small, Medium or Large.

Case 2: 6 fuzzy sets were created: SS, SL, MS, ML, LS and LL, which means that the changes are Small-Small, Small-Large, Medium-Small, Medium-Large, Large-Small or Large-Large.

Case 3: 9 fuzzy sets were created: SS, SM, SL, MS, MM, ML, LS, LM and LL, which means that variations are Small-Small, Small-Medium, Small-Large, Medium-Small, Medium-Medium, Medium-Large, Large-Small, Large-Medium or Large-Large.

Membership functions can take any form. In our case, for reasons of simplification of the calculations, we have used two kinds of functions: trapezoidal and trapezoidal-triangular.

The fuzzification is to evaluate the used membership functions. Each fuzzy set is characterized by a degree of membership μ which represents the degree of truth of the membership function as shown in Figure 4.

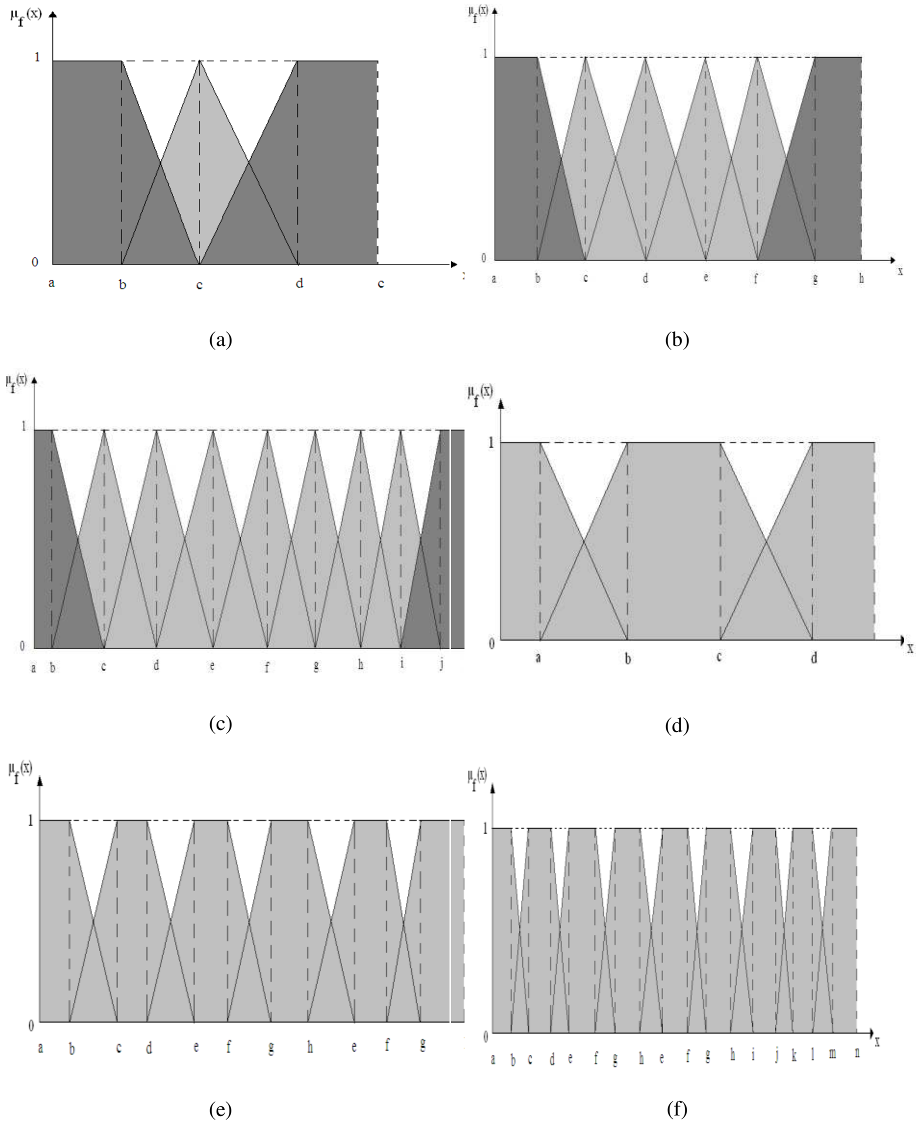


Figure 4
The membership functions of input / output system variables:
(a) mixed membership function with 3 partitions, (b) mixed membership function with 6 partitions, (c) mixed membership function with 9 partitions, (d) trapezoidal membership function with 3 partitions, (e) trapezoidal membership function with 6 partitions, (f) trapezoidal membership function with 9 partitions.

The fuzzy system used in this work is a Mamdani type we have used the temperature and the aging time as input variables and tensile strength as output variable. Tables 1, 2 and 3 present the fuzzy inference system and gives the fuzzy rules presented in following type: IF "Condition 1" AND "Condition 2" Then "Condition 3"

Table 1

Matrix inference functions for three partitions

Input \ Output		Tensile strength
t	S	L
	M	M
	L	S

Table 2

Matrix inference functions for six partitions

Input \ Output		Tensile strength
t	SS	LL
	SL	LS
	MS	ML
	ML	MS
	LS	SL
	LL	SS

Table 3

Matrix inference functions for nine partitions

Input \ Output		Tensile strength
t	SS	LL
	SM	LM
	SL	LS
	MS	ML
	MM	MM
	ML	MS
	LS	SL
	LM	SM
	LL	SS

The conditions being dependant between them by a logical operator AND, the degree of membership function, which is the conclusion, will be then equal for least two degrees of membership conditions. The membership's functions of output are built by aggregations of membership functions obtained by the rules in whole. For example, the degree of membership function of output (tensile strength) will be calculated three, six and nine times since there are exactly three, six and nine rules that arrive at this conclusion; one thus calculates the maximum between these membership functions. To calculate the numerical value of membership functions of output from the linguistic variables, we have used the gravity center method.

4 Results and Discussion

To model the behaviour of high voltage insulation, we need a long time to have a complete test database to give a good prediction. The use of fuzzy logic in modelling of insulators properties under thermal aging is economic and reduces the aging time. The work carried out consists of modelling and prediction of the XLPE properties under thermal aging. The experimental results of tensile strength under thermal aging [11] were used to build up the fuzzy rules and also for introducing a fuzzy inference system (FIS) in the prediction. Modelling of the chosen property is done for the entire interval of experiments which gives us a possibility to validate the developed fuzzy model. For 80°C and 100°C the experiments time is going up until 5000 h and for 120°C it reaches 2000 h. However, in the prediction phase the model gives the values of the property for aging times not reached by experiments.

Figure 5 presents the results of the tensile strength modelling in the case of 80°C as temperature of aging. In Figure 5 (a) we have represented the modelling results using the trapezoidal function with different partitions compared with the experimental results. The results of the use of mixed function (trapezoidal and rectangular) are presented in Figure 5 (b). Similar presentation as 80°C has been presented in the case of other aging temperatures (Figure 6 for 100°C and Figure 7 for 120°C).

From the obtained results, it is clear that fuzzy logic has great power to solve modelling problems in high voltage cables field. The chosen trapezoidal and mixed shape of the membership functions give good representations of linguistic variables. In Figures 5 and 6 where the variations of the tensile strength are not monotone and present some peaks at the beginning of aging times, the experimental curve is not reproduced closely by fuzzy model. The non ability of fuzzy logic to reproduce non monotone function is one of the most limitations of this modelling method. However when the XLPE property presents a monotone variation, as in the case of 120°C (Figure 7), the curves generated by the fuzzy model reproduce in a manner that is very close to the experimental results.

In order to test the developed FL model, we have divided the used experimental results into training and testing parts. Herein, 60% of experimental data were used for training, whereas 40% ones were employed for testing. The testing step is done in two different ways. First, we choose the training and the testing points in alternating way. Second, we take the first 60% of points to train the model and the last 40% of data for testing.

In Table 4, we have presented the medium and maximum relative errors for training and the first case of testing (alternating way). In this Table, we have also taken into consideration all proposed membership. If we choose the medium relative error as criterion of comparison, the trapezoidal function gives the best results. Moreover, in the almost cases (in the case of aging temperatures 100°C and 120°C), the model which use six partitions gives the best results and presents a minimum value of medium relative error. For 80°C, the medium relative errors in training and testing phases were found equal to 4.35% and 1.69%, respectively. For 100°C, also in the training and testing phases, the medium relative errors were determined as 4.52% and 10.01%, respectively. In the case of 120°C, the training is done with 0.5% as medium relative error, and the testing is achieved with 0.16% as medium relative error.

In Table 5, we present the similar analysis results as Table 4 but for the second case of testing. In this case, the best result of prediction at 80°C is obtained with mixed membership function and six partitions. However, for 100°C the best result is obtained with always mixed membership function but with three partitions. In the case of 120°C, the best result of prediction is obtained with trapezoidal membership function and six partitions. In terms of medium relative error, it reaches 3.77% in the training phase and 2.26% in testing phase. These results concern the aging temperature 80°C. At 100°C, the medium relative errors are 8.97% and 6.6% for training and testing respectively. In the case of 120°C, medium relative errors reach 0.53% and 0.2% for training and testing respectively.

5 Thermal Endurence of XLPE

It is crucial to the operator to know the changes in material properties resulting from prolonged exposure to high operating temperatures in order to ensure that the material exhibits throughout his life, the minimum properties required in its use [24]. In this section, fuzzy logic was applied to predict the behaviour of the cable insulation under thermal operating conditions for longer period of aging more than the experimental interval. The used fuzzy model gives the best results in the modelling step discussed in the preceding section. This model uses trapezoidal form of function and six partitions. XLPE tensile strength is predicted for an aging time of 9500 hours for temperatures of 80°C and 100°C, and an aging time of 3500 hours for temperature of 120°C. The obtained results are summarized in

Figure 8. This figure is used to determine the XLPE lifetime for a decrease of 50% from the initial value of the characteristic in question. This limit is considered to be the criterion of material end life [12].

Figure 9 is used to calculate the temperature values for each level of aging time. By extrapolation of the thermal endurance graph to life length of 20000 hours [11, 24], the index temperature is deduced to be 78.82°C (Table 6). The obtained value is slightly higher than that found in literature [25] via conventional method.

Table 4
Comparison of experimental results with training and testing results of Fuzzy logic model (first case)

	Applied temperature	80°C			100°C			120°C		
	Type of membership function	Mixed								
	Number of partitions	3	6	9	3	6	9	3	6	9
Testing	Medium relative error (%)	2,38	1,96	1,69	11,11	10,98	10,38	9,73	5,92	3,7
	Maximum relative error (%)	2,99	2,47	2,52	17,58	19,22	20	9,73	5,92	3,7
Training	Medium relative error (%)	5	4,77	4,4	6,43	5,14	4,76	17,52	11,50	7,65
	Maximum relative error (%)	12,63	12,13	12,49	12,42	11,32	10,62	31,00	20,77	13,78
	Type of membership function	Trapezoidal								
	Number of partitions	3	6	9	3	6	9	3	6	9
Testing	Medium relative error (%)	3,33	1,96	1,69	11,28	10,93	10,01	9,9	0,16	13,1
	Maximum relative error (%)	5,19	2,36	2,60	19,41	19,93	20	9,9	0,16	13,1
Training	Medium relative error (%)	5,42	4,56	4,35	7,09	4,3	4,52	18,79	0,5	10,67
	Maximum relative error (%)	14,75	12,45	12,68	12,27	10,02	10,67	34,45	1,20	26,94

Table 5
Comparison of experimental results with training and testing results of Fuzzy logic model (second case)

	Applied temperature	80°C			100°C			120°C		
	Type of membership function	Mixed								
	Number of partitions	3	6	9	3	6	9	3	6	9
Testing	Medium relative error (%)	2,39	2,26	2,45	6,6	7,53	8,07	31	20,77	13,78
	Maximum relative error (%)	5,66	6,21	6,76	11,63	14,46	15,5	31	20,77	13,78
Training	Medium relative error (%)	4,23	3,77	3,9	8,97	8,85	8,66	9,50	4,75	3,16
	Maximum relative	12,63	12,13	12,49	16,86	19,80	20,2	19,12	12,5	8,33

	error (%)									
	Type of membership function	Trapezoidal								
	Number of partitions	3	6	9	3	6	9	3	6	9
Testing	Medium relative error (%)	3,08	2,4	2,49	9,37	8,1	7,96	34,45	0,2	3,34
	Maximum relative error (%)	6,1	6,7	6,94	16,40	14,26	15,36	34,45	0,2	3,34
Training	Medium relative error (%)	5,47	3,67	3,89	9,22	8,44	8,74	8,31	0,53	13,1
	Maximum relative error (%)	14,75	12,45	12,68	18,43	19,80	20,39	19,52	1,20	26,94

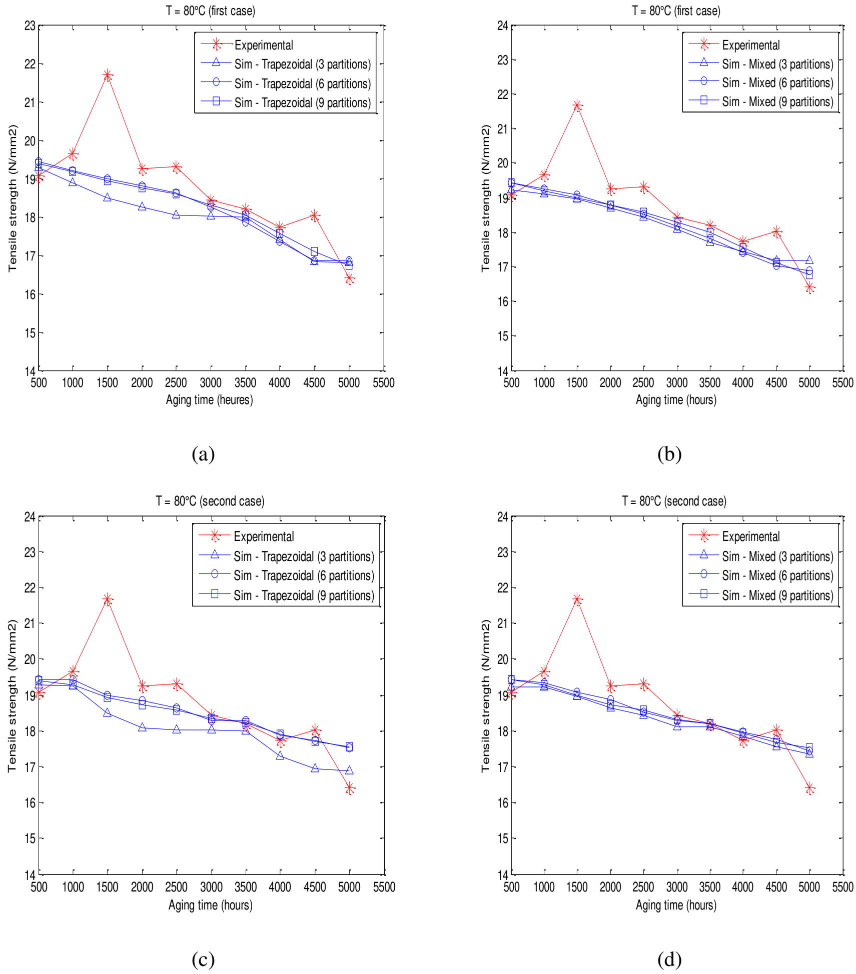
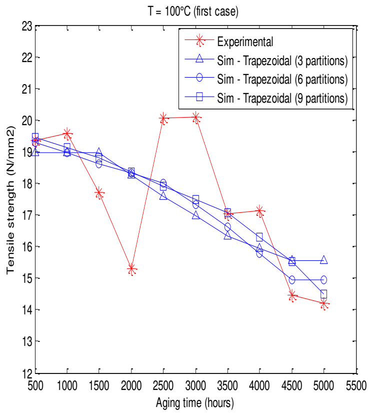
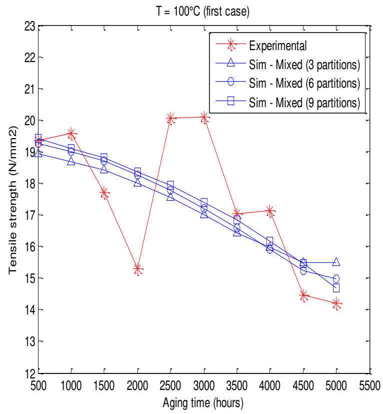


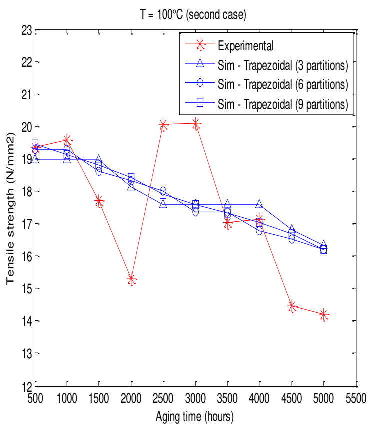
Figure 5
Modelling of tensile strength at T = 80°C:
(a): Trapezoidal function (first case), (b): Mixed function (first case),
(c): Trapezoidal function (second case), (d): Mixed function (second case)



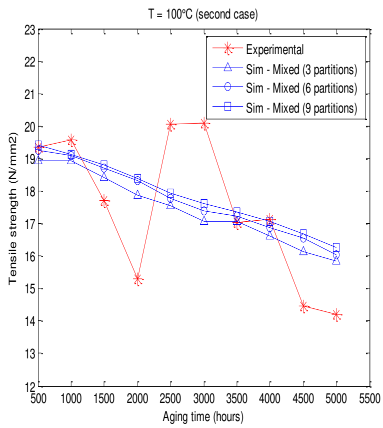
(a)



(b)



(c)



(d)

Figure 6

Modelling of tensile strength at T = 100°C:

(a): Trapezoidal function (first case), (b): Mixed function (first case),

(c): Trapezoidal function (second case), (d): Mixed function (second case).

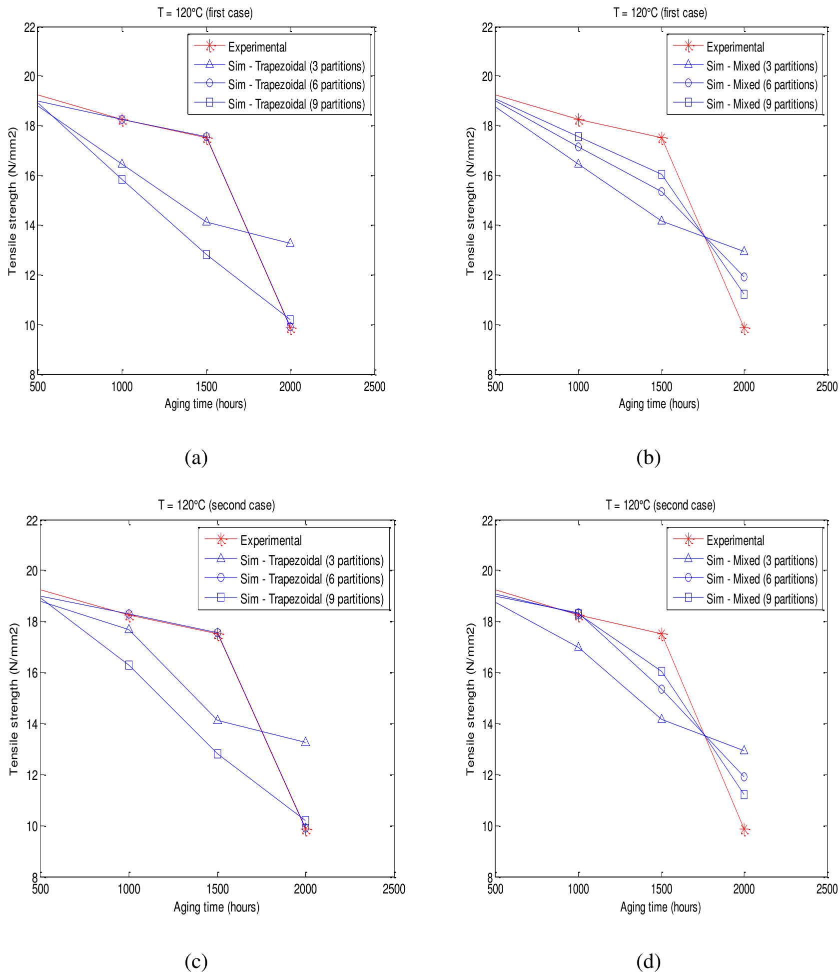


Figure 7
Modelling of tensile strength at T = 120°C:
(a): Trapezoidal function (first case), (b): Mixed function (first case),
(c): Trapezoidal function (second case), (d): Mixed function (second case).

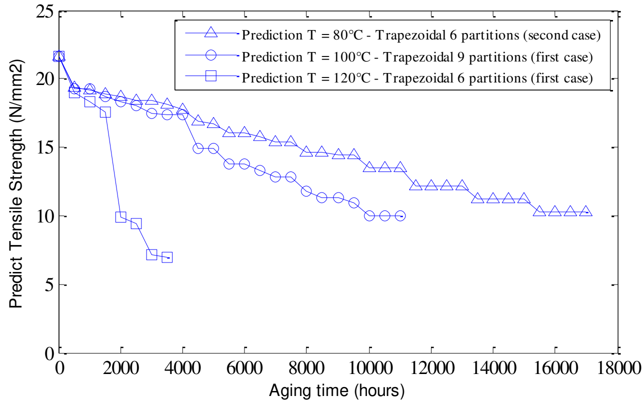


Figure 8

Prediction of tensile strength for different aging temperatures

In Figure 9, we have presented the lifetime endurance of the insulation, where we have reported the variation of aging time (in logarithmic) as a function of the reciprocal of absolute temperature [24].

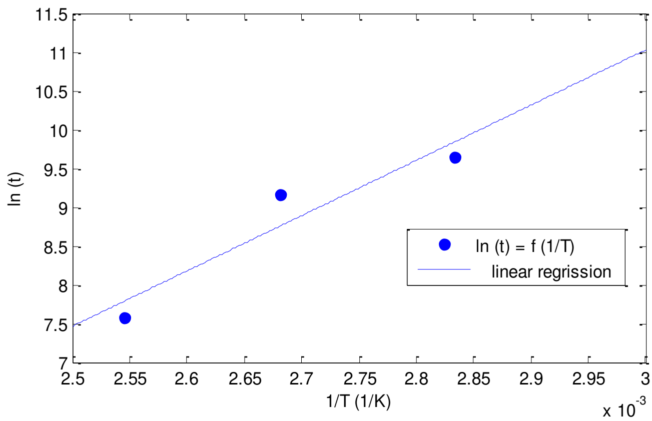


Figure 9

Variation of aging time logarithm vs reciprocal of absolute temperature

Table 6

Different values of aging times and temperatures for 50% end life criterion

Aging time (hours)	Logarithm of time ln (t)	1/T (K^{-1})	T (°C)
1939	7,5699	0,002545	120
9552	9,1645	0,002681	100
15447	9,6452	0,002833	80
20000	9,9035	0,002842	78,82

Conclusions

The process of aging of the insulation material can be monitored by several techniques based on numerous experimental studies or modelling aspect. To predict the mechanical behaviour of the insulation, a long time is needed to have a complete test database enough to give a good prediction.

In this paper, a new approach based on an expert system of fuzzy logic modelling was proposed to predict the behaviour of XLPE used as insulation in high voltage cables. From the obtained results, it can be concluded that fuzzy logic by its flexibility in establishment helped us to model and predict the XLPE properties in a simple way. This intelligent modelling technique can be considered as an alternative solution for conventional methods since it avoids mathematical modelling especially for complex systems. The use of this method proves a great effectiveness.

Several Fuzzy model parameters have been also studied in this paper. It seems that the trapezoidal function form presents linguistic variables in a best way and gives very close results to experimental results with quite small error rates and in relatively short time. Finally, we can say that it is economic to apply fuzzy logic in investigations on the HV insulation since it allows us to reduce both tests time and costs in laboratory experiments.

References

[1] P. Hyvönen, B. Oyegoke, M. Aro: Diagnostics and Testing of High Voltage Cable Systems, Helsinki University of Technology, High Voltage Institute, Finland 2003

[2] A. R. Yazdandoust, F. Haghjoo, S. M. Shahrtash: Insulation Status Assessment in High Voltage Cables Based on Decision Tree Algorithm, IEEE Electrical Power & Energy Conference, 2008

[3] A. Tzimas, S. Rowland, L. A. Dissado, M. Fu, U. H. Nilsson: Effect of Long-Time Electrical and Thermal Stresses upon the Endurance Capability of Cable Insulation Material, IEEE Trans. Dielect. Electri. Insul., Vol. 16, No. 5, pp. 1436-1443, October 2009

[4] T. Seguchia, K. Tamura, T. Ohshima, A. Shimada, H. Kudoh: Degradation Mechanisms of Cable Insulation Materials During Radiation Thermal Ageing in Radiation Environment, Radiation Physics and Chemistry, Vol. 80, pp. 268-273, 2011

[5] J. Li, X. Zhao, G. Yin, S. Li, J. Zhao, B. Ouyang: The Effect of Accelerated Water Tree Ageing on the Properties of XLPE Cable Insulation, IEEE Trans. Dielect. Electri. Insul., Vol. 18, No. 5, pp. 1562-1569, October 2011

[6] H. C. Chen, F. C. Gu, M. H. Wang: A Novel Extension Neural Network-based Partial Discharge Pattern Recognition Method for High-Voltage

Power Apparatus, Expert Systems with Applications, Vol. 39, pp. 3423-3431, 2012

[7] L. Mokhnache, A. Boubakeur: Comparison of Different Back-Propagation Algorithms Used in the Diagnosis of Transformer Oil, IEEE, Ann. Rep. Conf. Electrical Insulation and Dielectric Phenomena (CEIDP), pp. 224-247, 2002

[8] Y. Bourek, L. Mokhnache, N. Nait Said, R. Kattan: Determination of Ionization Conditions Characterizing the Breakdown Threshold of a Point-Plane Air Interval Using Fuzzy Logic, Electric Power Systems Research, Vol. 81, pp. 2038-2047, 2011

[9] L. Bessissa, L. Boukezzi, D. Mahi, A. Boubakeur: Modelling and Prediction of HV Insulation Cables Behaviour under Thermal Aging: Use of Fuzzy logic Approach, Third International Meeting on Dielectric Materials IMDM’3, Monastir - Tunisia, November 13-18, 2011

[10] L. Bessissa, L. Boukezzi, Z. Turki, D. Mahi, A. Boubakeur: Influence of the Variation of Fuzzy Algorithms Parameters on the Quality of HV Cable Insulation Properties Modeling Subjected to Thermal Aging, 5th International Workshop on Impedance Spectroscopy IWIS’5, Chemnitz-Germany, November 26-28, 2012

[11] I.E.C. 540: Test Methods for Insulation and Sheaths of Rigid and Flexible Electrical Cables (elastomeric and thermoplastic blends) 1983 (in French)

[12] L. Boukezzi: Influence of Thermal Aging on the Properties of Cross-linked Polyethylene Insulation used in High Voltage Cables Ph.D. Thesis ENP of Algiers Algeria, June 2007

[13] I.E.C. 502: Power Cables Insulated with Solid Dielectrics for RATED Voltages from 1kV to 30kV, Edition 1997 (in French)

[14] L. Zadeh: Fuzzy Sets, Inform Control, 8: pp. 338-53, 1967

[15] M. Sarıdemir, I. B. Topçu, F. Özcan, M.H. Severcan: Prediction of Long-Term Effects of GGBFS on Compressive Strength of Concrete by Artificial Neural Networks and Fuzzy Logic, Construction and Building Materials, 2008

[16] Z. Sen: Fuzzy Algorithm for Estimation of Solar Irradiation from Sunshine Duration, Sol Energy, Vol. 63, No. 1, pp. 39-49, 1998

[17] W. Pedrycz: Fuzzy Control and Fuzzy Systems Research Studies, John Wiley and Sons, N.Y. 1989

[18] R. Gayakwad: Optimized Fuzzy Logic for Motion Control, Acta Polytechnica Hungarica, Vol. 7, No. 5, pp. 161-168, 2010

[19] K. M. Passino, S. Yurkovich: Fuzzy control, Addison-Wesley, 1998

[20] DWC. Ho, P.A. Zhang, J. Xu: Fuzzy Wavelet Networks for Function Learning, IEEE Trans Fuzzy Syst, Vol. 9, No. 1, pp. 200-11, 2001

[21] K. Nagy, S. Divéki, P. Odry, M. Sokola: A Stochastic Approach to Fuzzy Control, Acta Polytechnica Hungarica, Vol. 9, No. 6, pp. 29-48, 2012

[22] R. Bakhtyar, A. Y. Bakhtiary, A. Ghaheri: Application of Neuro-Fuzzy Approach in Prediction of Runup in Swash Zone, Applied Ocean Research, Vol. 30, pp. 17-27, 2008

[23] J. SR. Jang. ANFIS: Adaptive-Network-based Fuzzy Inference Systems, IEEE Trans Syst Man Cybern, Vol. 23, No. 3, pp. 665-85, 1993

[24] B. Fallou: New Rules for the Determination of Materials Thermal Endurance, R.G.E., Vol. 83, No. 7/8, pp. 445-52, July/August 1974

[25] L. Boukezzi, M. Nedjar, L. Mokhnache, M. Lallouani, A. Boubakeur: Thermal Aging of Cross-linked Polyethylene, Annales de Chimie, Science des Matériaux, Vol. 31, No. 5, pp. 561-569, 2006

The Microstructure Evolution and Its Effect on Corrosion Properties of 18Cr-12Ni-2,5Mo Steel Annealed at 500-900°C

Mária Dománková, Edina Kocsisová, Ivan Slatkovský, Peter Pinke

Institute of Materials Science, Faculty of Material Science and Technology in Trnava, Slovak University of Technology in Bratislava, Bottova 25, 917 24 Trnava, Slovakia, e-mail: maria.domankova@stuba.sk, edina.kocsisova@stuba.sk, ivan.slatkovsky@stuba.sk, peter.pinke@stuba.sk

Abstract: The process of the secondary phases precipitation controls the mechanical and physical properties of the stainless steels. The process of precipitation and its influence on corrosion resistance was analysed in the experimental steel AISI 316. The corrosion test ASTM A 262 practice A was used for evaluation of the experimental steel to intergranular corrosion. Results showed that the solution-annealed samples (1100°C/1h – followed by water quenching) did not present susceptibility to intergranular corrosion. The heat treatment in the range from 500 to 900°C resulted in susceptibility to intergranular attack. Precipitation of carbide $M_{23}C_6$ accelerates sensitisation of the experimental steel. Increasing of holding time cased precipitation σ-phase and carbide M_6C, which were detected at grain boundaries. Transmission electron microscopy was used for identification of particles at the grain boundaries, measurement of their size and studied of their morphology.

Keywords: austenitic stainless steel; precipitation; sensitisation; intergranular corrosion

1 Introduction

Intergranular corrosion (IGC) and stress corrosion cracking (SCC) of austenitic stainless steels (ASSs) are the most important corrosion process that affects the service behaviour of these materials. Austenitic stainless steels are the most favoured construction materials of various components required in chemical, petrochemical and nuclear industries. The selection of these is made basically due to a good combination of mechanical, fabrication and corrosion resistance properties. But austenitic stainless steels which have undergone treatment in the temperature range between 500-900°C or have been cooled slowly from annealing temperatures (1000-1200°C) become sensitised. The phenomenon of sensitisation in stainless steels refers to their susceptibility to intergranular corrosion resulting from microstructural changes. Grain boundary chemistry is known to be a critical factors impacting structural reliability of ASSs [1-3, 11-14]. If the chromium content near the grain boundaries drops under the passivity limit 12 wt.%, the steel becomes to be sensitised. In the sensitised condition, the steels are quite susceptible to the intergranular corrosion (IGC) and intergranular stress corrosion cracking (IGSCC) that can result in premature failures of the fabricated components. The sensitisation temperature range is often encountered during isothermal heat treatment, slow cooling from the solution annealing temperature, the improper heat treatment in the heat affected zone of the welds or welding joints or hot working of the material [4-6].

Degree of the sensitisation (DOS) is influenced by factors such as the steel chemical composition, grain size, degree of strain or temperature and time of isothermal annealing [5-7]. The sensitisation involves both nucleation and growth of intermetallic phases at the grain boundaries. Sensitisation resulting from isothermal exposure is normally represented by time-temperature-sensitisation (TTS) diagrams, which are plots of aging time versus temperature necessary for sensitisation. These are C - shaped curves, which separate sensitised and non-sensitised regions. The TTS diagrams show the time required for isothermal sensitisation at various temperatures and can be used to solve problems such as the selection of conditions of annealing which will not result in sensitisation. The nose of this curve specifies the critical temperature at which the minimum time (t_{min}) is required for sensitisation [13, 14]. The ageing reactions, nature of precipitates, and precipitation kinetics during high-temperature exposure were studied in the past. The weight fraction of carbides was often assess by dint of the usual electrochemical dissolution methods, applied to austenitic stainless steels with high carbon contents. The use of analytical electron microscopy has allowed the chromium distribution at grain boundaries to be measured and it was possible to compare these results with value predicted by different models [8].

However, in austenitic stainless steels except chromium also other alloying elements may effect and change the precipitate consequention or kind of precipitates. For instance Hull [9] investigated a great number of standard carbon and nitrogen free iron-chromium-nickel alloys with regard to the alloying addition influence on the precipitation of σ-phase and χ-phase at the temperature 816°C and precipitation times up to 1000 hours. From the data obtained, he calculate the chromium equivalent for the material embrittlement connected to σ-phase formation:

$$\begin{aligned} Cr_{eqv} &= \%Cr + 0{,}31\%Mn + 1{,}7\%Mo + 1{,}58\%Si + \\ &+ 2{,}44\%Ti + 1{,}22\%Ta + 0{,}97\%W - 0{,}266\%Ni - 0{,}177\%Co \end{aligned} \quad (1)$$

In this research article, we report of the effects temperature and aging time on sensitisation and precipitation of the secondary phases in AISI 316 stainless steel.

2 Material and Experimental Procedures

Wrought austenitic stainless steel AISI 316 containing 1% of δ-ferrite was used during the course of this investigation and its chemical composition is given in Table 1. All samples were solution heat-treated at 1100°C/1 h. followed by water quenching. The samples were heat treated at various temperatures from 500 to 900°C for duration range from 10 min. to 1000 h in evacuated tubes. The samples were water quenched after the heat treatments.

Table 1
Chemical composition (in wt.%) of austenitic stainless steel

Steel	C	N	Si	Mn	P	S	Cr	Ni	Mo	Fe
AISI 316	0,05	0,032	0,47	0,86	0,003	0,001	17,55	11,56	2,10	Bal.

The specimens for light optical microscopy (LOM) examination were polished up to fine diamond (~1 μm) finish. The specimens were etched chemically for 60 sec. using solution: 10 ml H_2SO_4 + 10 ml HNO_3 + 20 ml HF + 50 ml distilled H_2O. Then the screening of microstructures was done using a light microscope NEOPHOT 32 equipped with the CCD camera.

The oxalic acid etch test was used for the determination of steel sensitivity to intergranular corrosion. The specimens were electrolytically etched in 10% oxalic acid for 90 sec. at a current density of 1 A/cm^2. The etched microstructures were then examined at 250x and were classified into the following types [10, 12, 14-16]:

- step structure: steps only between grains, no ditches at grain boundaries,
- dual structure: some ditches at grain boundaries, but no single grain completely surrounded by ditches,
- ditch structure: one or more grains completely surrounded by ditches and the steel is classified as sensitised in this case. The steel is susceptible to the intergranular corrosion.

For the individual secondary phase identification transmission electron microscopy (TEM) of the extraction carbon replicas was utilised. The carbon extraction replicas were obtained from mechanically polished and etched surfaces. The replicas were stripped from the specimens in the solution of CH_3COOH : $HClO_4$ = 4 :1 at 20°C, 20 V [15].

Thin foils suitable for transmission electron microscopy (TEM) observation were prepared from each of the samples. Small discs of 3 mm in diameter and thick about 0.1 mm were jet-electropolished in electrolyte HNO_3 : CH_3OH = 3 : 7, at 0°C and 15V to obtain transparent areas near the central hole. The jet-electropolishing was done by TENUPOL 5.

TEM observations were performed using JEOL 200 CX operated at 200 kV and Philips CM 300 operating at 300 kV equipped with energy-dispersive X-ray spectrometer (EDX), which was used for the microchemical analyses. The analysis was supplemented by selected area electron diffraction for the phases identification.

3 Results

3.1 LOM Observations

The results of the light microscopy (LOM) examination are summarised in Fig. 1. The microstructure of AISI 316 after solution annealing consists of polyhedral austenitic grains with twinning typical for fcc microstructure. The average austenitic grain size in this state is about 35±5 μm (Fig. 1a). A small amount of δ-ferrite was also recorded. No precipitates were observed at the grain boundaries (GB) of solution annealed steels. The microstructures of aged states are shown in Fig. 1b-1d.

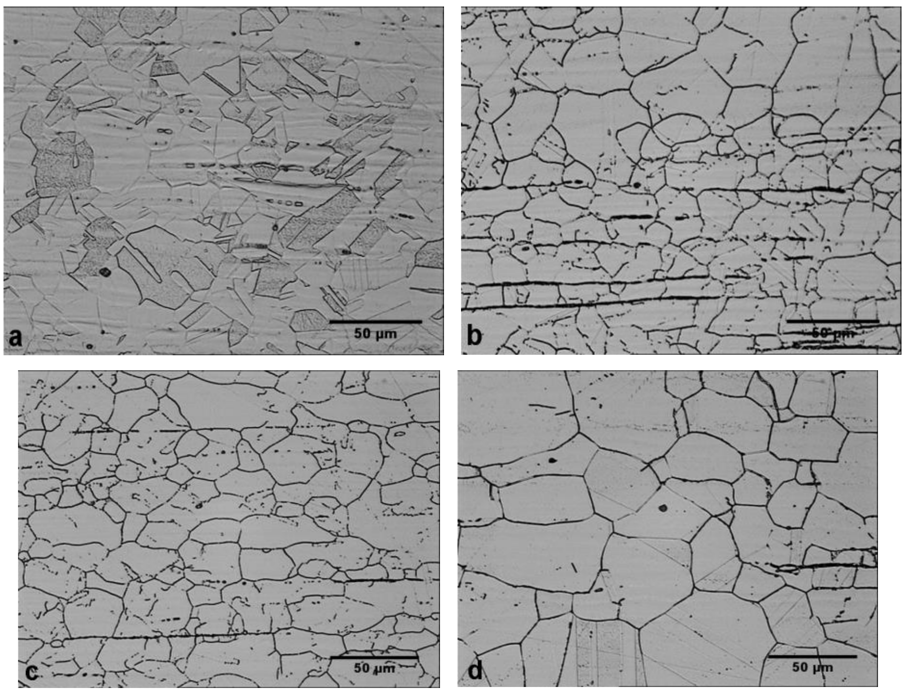


Figure 1
Microstructure of AISI 316 a) after solution annealing, b) after aging 750°C/1 h, c) after aging 750°C/300 h, d) after aging 750°C/1000 h

Significant etching of grain boundaries was observed by LOM with increasing of holding time at the temperature. This effect is caused by precipitation of secondary phases at the grain boundaries. Fig. 1b shows the evolution of the precipitation at the GB in the isothermally aged sample (750°C/1h). Fig. 1c shows microstructure in the isothermally aged sample (750°C/300 h) and Fig. 1d shows microstructure of the sample after the heat treatment 750°C/1000 h. The growth of grains was observed in the case of this sample. The average austenitic grain size is about 60±8 μm (Fig. 1d).

The rapid oxalic acid etch test was used for analysis of the grain boundary sensitisation development. Examples of microstructure obtained by etch test for AISI 316 are given in Fig. 2. TTS diagram was obtained by plotting sensitisation test results on a temperature versus log soaking time axes and drawing a line which demarcates the sensitised and non-sensitised regions (Fig. 3). The nose of this curve for the experimental steel AISI 316 was at the temperature 800°C and t_{min}=20 min.

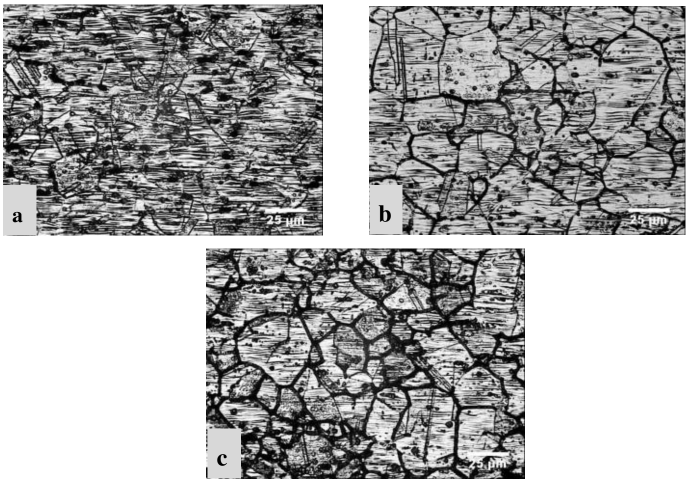


Figure 2
Examples of AISI 316 microstructure obtained after oxalic acid etch test a) after solution annealing (step), b) after aging 800°C/10 min. (dual), c) after aging 800°C/10 h (ditch)

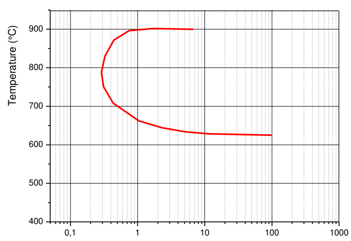


Figure 3
TTS diagram for AISI type 316 austenitic stainless steel established according to the test ASTM A262 practice A

3.2 TEM Observations

The TEM microstructure analysis was focused especially on the identification of secondary phases, which precipitated at the grain boundaries during the annealing. The changes of chemical composition and the relative size of particles were evaluated. Fig. 4 shows the microstructure of AISI 316 in the solution annealing state observed by TEM using carbon replica. The pure grain boundaries are documented by the figures. Fig. 5 shows the typical GBs formed in an early stage of the aging treatment (750°C/10 min.), only a few particles of secondary phase were observed at the grain boundaries using TEM. As first $M_{23}C_6$ carbide at the austenitic grain boundaries was detected (Fig. 6). Besides $M_{23}C_6$ carbide, σ-phase and M_6C carbide were detected at the grain boundaries. The density and size of precipitates at the grain boundaries growth with increasing of the aging time (Fig. 7 and Fig. 8).

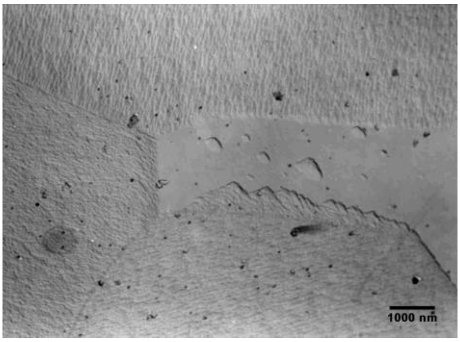


Figure 4

Microstructure of AISI 316 in the solution annealing state

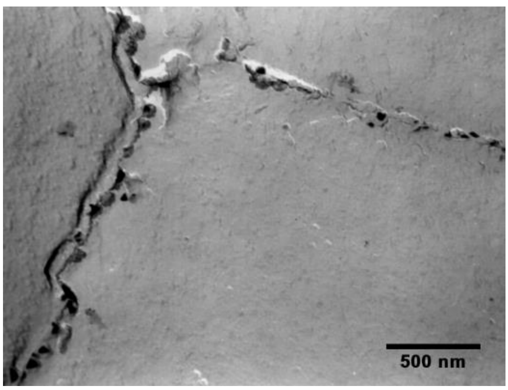


Figure 5

Precipitation of $M_{23}C_6$ carbide at the GB – after aging 750°C/10 min.

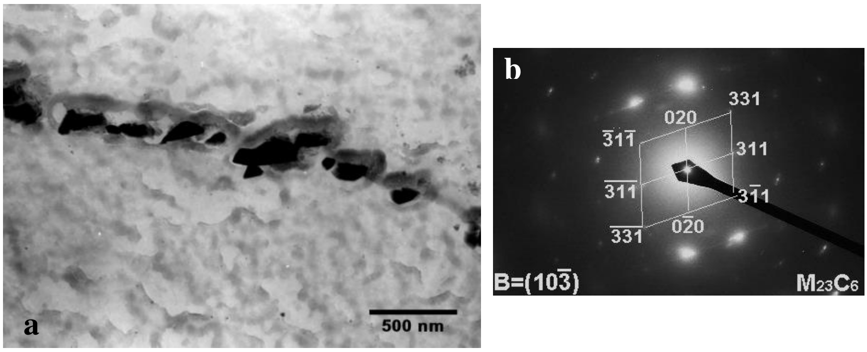


Figure 6

a) Detail of the grain boundary with particles of irregular shape after aging 750°C/5 h, b) particles were identified by SAED as carbide $M_{23}C_6$

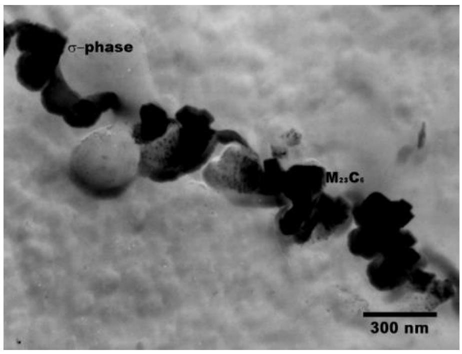


Figure 7

Precipitation of secondary phases at the grain boundary after aging 750°C/300 h

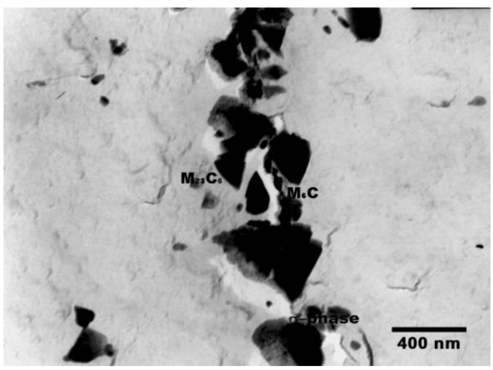


Figure 8

Precipitation of secondary phases at the grain boundary after aging 750°C/1000 h

The change of the particles size during aging at the temperature 750°C is shown in Fig. 9. The medium size of particles increased from 40 nm to 120 nm. The maximum of the particles size was found at the temperature at 750°C with holding time 30h. Subsequently the medium particles size decreased with increasing of holding time. In our opinion decreasing of particles size was caused by the change of phase composition in the experimental steel. Carbide $M_{23}C_6$ content decreased and content of carbide M_6C increased.

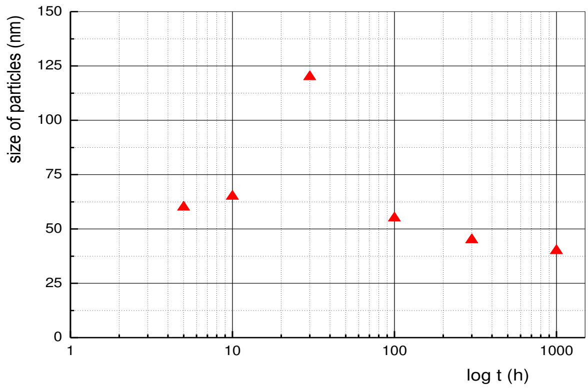


Figure 9
Growth of the particles in the experimental steel AISI 316 after aging at the temperature 750°C

Fig. 10 summarises the average metal composition of carbide $M_{23}C_6$ after the aging treatment at the temperature 750°C measured by EDX analysis. The measurement of chemical composition shows the slight growth of the chromium contain with increasing of the aging time.

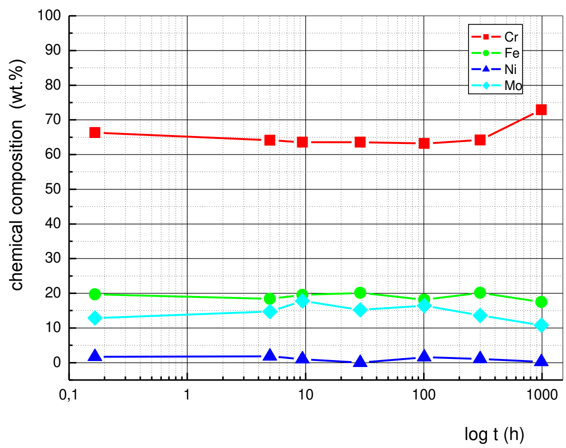


Figure 10
Changes of the chemical composition of carbide $M_{23}C_6$ during annealing at the temperature 750°C

Fig. 11 shows the changes of chemical composition of σ-phase during the annealing at the temperature 750°C. The main element of this phase is iron, and a slight increase in Fe content can be observed as a function of increasing aging time. The changes of other elements were not significant. The changes of chemical composition of the carbide M_6C is shown by Fig. 12. The main element of this phase is molybdenum. The significant changes of the chemical composition of the carbide M_6C during the annealing were not observed by EDX analysis.

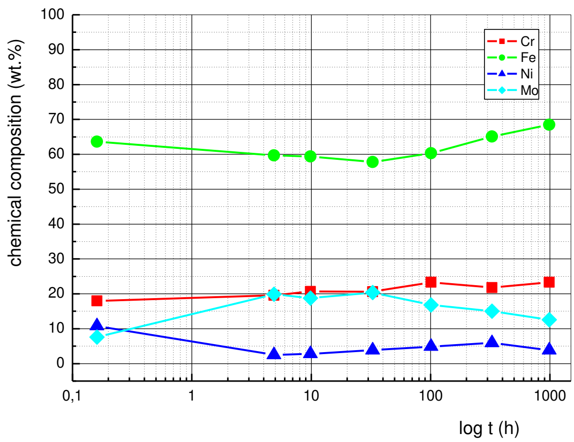


Figure 11

Changes of the chemical composition of σ-phase during annealing at the temperature 750°C

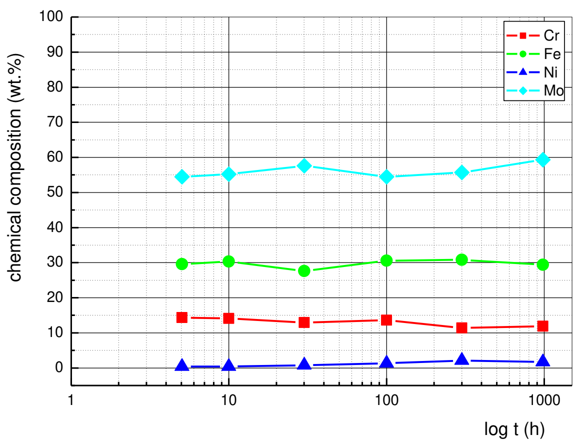


Figure 12

Changes of the chemical composition of carbide M_6C during annealing at the temperature 750°C

The changing fractions of identified secondary phases in steel AISI 316 during aging at temperature 750°C are demonstrated in Fig. 13.

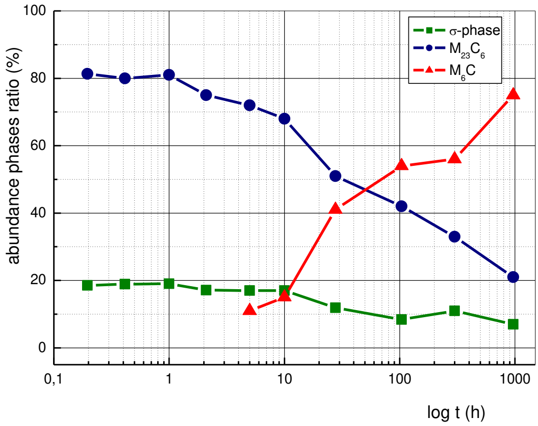


Figure 13
Fractions of the secondary phases in AISI 316 after aging at the temperature 750°C

Fig. 14 shows substructure of the experimental steel after solution annealing observed using thin foils. The grain boundaries were pure without precipitates. Fig. 15 shows details of the substructure after annealing 800°C/10 min – early stage of the precipitation process. The small precipitates of irregular shape were observed at some grain boundaries. EDX line analysis across the boundary (Fig. 16) did not show the change of the chemical composition between the matrix and the grain boundary in the case of solution annealed state. Small decreasing of chromium contain across the boundary was found by EDX line analysis across the boundary (Fig. 16), but chromium content was sufficient to maintain of corrosion resistance of experimental steel. Precipitation during heat treatment 800°C/10 min did not cause sensitisation of the steel AISI 316. This results is in good agreement with results of the rapid oxalic acid etch test and construction of TTS diagram (Fig. 3), respectively.

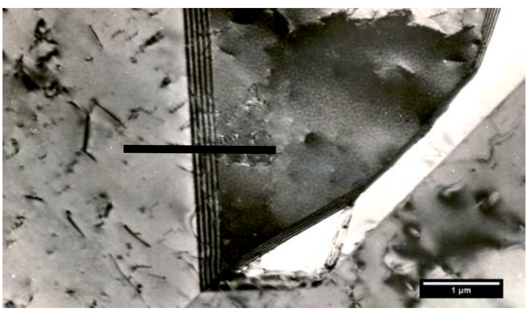


Figure 14
Details of the grain boundaries in the experimental steel after solution annealing (line marked area of EDX analysis)

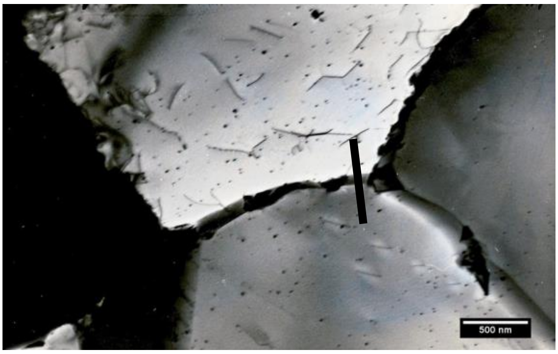


Figure 15

Details of the grain boundaries in the experimental steel after annealing 800°C/10 min (line marked area of EDX analysis)

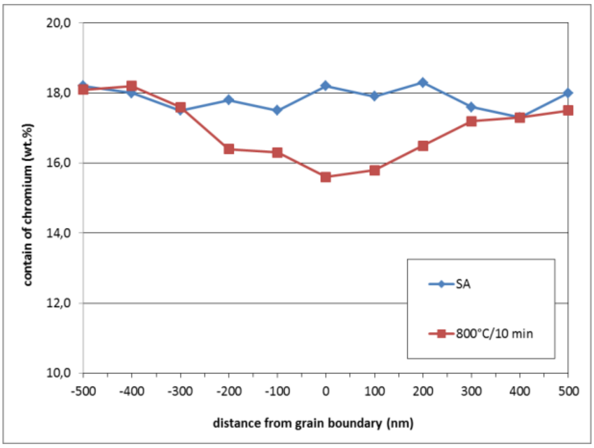


Figure 16

Changes of chromium contain in the area of grain boundary in the experimental steel after solution annealing (SA) and heat treatment 800°C/10 min measured by EDX analysis

Conclusions

The precipitation behaviour of AISI 316 austenitic stainless steel has been investigated in the temperatures range from 500 to 900°C for duration range from 10 min to 1000 h. The following conclusions were drawn:

1. TTS diagrams of experimental steel after oxalic acid etch test ASTM A262 practice A were constructed. The nose of this curve was at the temperature 800°C and t_{min}=20 min.

2. Sensitisation of the experimental steel accelerated the precipitation of the carbide $M_{23}C_6$, besides carbide $M_{23}C_6$, σ-phase and carbide M_6C were detected at the austenitic grain boundaries (Fig. 12).

3. The medium size of particles increased from 40 nm to 110 nm with increasing of holding time during the heat treatment.
4. The measurements of the chemical composition of the secondary phases showed slight increasing the main elements (chromium and iron) in the case of the carbide $M_{23}C_6$ and σ-phase with increasing of the time annealing.
5. The small decreasing of chromium contain near of the grain boundary after heat treatment 800°C/10 min was found by EDX line measurement, but the chromium content was sufficient to maintain of corrosion resistance of experimental steel.

Acknowledgement

The authors wish to thank the financial support of Scientific Grant Agency of Slovak Republic (VEGA) under grant No. 1/0402/13

References

[1] Lo, K. K.; Shek, C. H.; Lai, J. K. L.: Recent Developments in Stainless Steels. In: Materials Science and Engineering:R:Reports, 65(2009), pp. 39-104

[2] Tavares, S. S. M.; Moura, V.; da Costa, V. C.; Fereira, M. L. R.; Pardal, J. M.: Microstructural Changes and Corrosion Resistance of 310S Steel Exposed to 600-800°C. In: Material Characterization, 60 (2009), pp. 573-578

[3] Nikulin, I.; Kipelova, A.; Kaibyshev, R.: Effect of High-Temperature Exposure on the Mechanical Properties of 18Cr-8Ni-W-Nb-V-N Stainless Steel. In: Materials Science and Engineering A 554 (2012), pp. 61-66

[4] Kanek, K.; Fukunaga, T.; Yamada, K.; Nakada, N.; Kikuchi, M.; Saghi, Z.; Barnard, J. S.; Midgley, P. A.: Formation of $M_{23}C_6$ – Type Precipitates and Chromium-depleted Zones in Austenitic Stainless Steel. In: Scripta Materialia 65 (2012), pp. 509-512

[5] Ahmadabadi, P. M.; Kain, V.; Dang, B. K., Samajdar, I.: Role of Grain Boundary Nature and Residual Strain in Controlling Sensitisation of Type 304 Stainless Steel. In: Corrosion Science 66(2013), pp. 242-255

[6] Bitay, E.; Kovacs, T.: The Effect of Laser Surface Treatment on the Wear Resistance. In: Materials Science Forum 649 (2010), pp. 107-112

[7] Tsay, L. W.; Lin, Y. J.; Chen, C.: The Effects of Rolling Temperature and Sensitisation Treatment on the Sulphide Stress Corrosion Cracking of 304 L Stainless Steel. In: Corrosion Science 63 (2012), pp. 267-274

[8] Sanzay, M.; Barard, K.; Karlen, W.: TEM Observation and Finite Element Modelling of Channel Deformation in Pre-irradiated Austenitic Stainless Steels – Interactions with the Surfaces and Grain Boundaries. In: Journal of Nuclear Materials 406 (2010), pp. 152-165

[9] Hull, F. C.: Effects of Composition on Embrittlement of Austenitic Stainless Steels. In: Welding Journal 52 (1973), pp. 104-113

[10] ASTM Standard Practice in A 262 for Detecting Susceptibility to Intergranular Corrosion in Austenitic Stainless Steels

[11] Leiva-García, R.; Munoz-Portero, M. J.; García-Antón, J.: Evaluation of Alloy 146, 279, 900 and 926 Sensitization to Intergranular Corrosion by Means of Electrochemical Methods and Image Analysis. In: Corrosion Science 51 (2009), pp. 2080-2091

[12] Záhumenský, P.; Tuleja, S.; Országová, J.; Janovec, J., Siládiová, V.: Corrosion Resistance of 18Cr-12Ni-2,5-Mo Steel Annealed at 500-1050°C. In: Corrosion Science 41 (1999), pp. 1305-1322

[13] Terada, M.; Saiki, M.; Costa, I.; Padilha, A. F.: Microstructure and Intergranular Corrosion of Austenitic Stainless Steel 1.4970. In: Journal of Nuclear Materials 358 (2006), pp. 40-46

[14] Dománková, M.; Marek, P.; Magula, V.: Corrosion Resistance of 18Cr-12Ni-2,5Mo Steel Annealed at 500-900°C. In: Acta Metallurgica Slovaca, Special Issue Vol. 13 (2007), pp. 487-492

[15] Bakajová, J.; Dománková, M.; Čička, R.; Eglsäer, S.; Janovec, J.: Influence of Annealing Conditions on Microstructure and Phase Occurrence in High-Alloy CrMnN Steels. In: Materials Characterization 61 (2010), pp. 969-974

[16] Matula, M.; Hyspecká, L.; Svoboda, M.; Vodarek, V.; Dagbert, C.; Galland, J.; Stonawska, Z.; Tuma, L.: Intergranular Corrosion of AISI 316L Steel. In: Materials Characterization 46 (2001), pp. 203-210

Analysis of Heat Pump Condenser's Performance using the Mathematical Model and a Numerical Method

Jozsef Nyers

Óbuda University, Budapest, Bécsi út 96, 1034 Budapest, Hungary
Subotica Tech, Marko Oreskovic 16, 24000 Subotica, Serbia
e-mail: jnyers@vts.su.ac.rs

Laszlo Garbai

Budapest University of Technology and Economics, 1111 Budapest, Műegyetem rkp. 3-9, Hungary; e-mail: garbai@epgep.bme.hu

Arpad Nyers

Subotica Tech, Marko Oreskovic 16, 24000 Subotica, Serbia
Tera Term co., Vig Istvan 1, 24000 Subotica, Serbia
e-mail: nyarp@vts.su.ac.rs

Abstract: this article aims to investigate the performance of the heat pump's condenser and functionality, suitability of the steady-state mode mathematical model and the used numerical method. The heat transferred through the condenser is the performance of the condenser. The condenser is plate exchanger. The refrigerant is R 134a and the heated medium is water. This wide-range study investigates the condenser's performance depending on four independent external variables. The independent external variables are: inlet temperature of the refrigerant, refrigerant mass flow rate, the water inlet temperature and the water mass flow rate.The investigation is carried out by using a steady state mode mathematical model with lumped parameter. The model consists of two parts. Part one describes mathematically the cooling of the superheated vapor, while the second part describes the condensation itself. The physical model is simple, only the plate condenser and the input, output parameters are included. This article does not take into consideration the compressor and circulating pump model. That remains the subject of future investigation. The results can be represented in three-dimensional graphics. For example, it shows the changes of heat transfer in the condenser depending on the refrigerant mass flow rate and the water mass flow rate.

Keywords: condenser; heat pump; circulator pump; performance; mathematical model; mass flow rate; efficiency

Nomenclature

$\dot{m}$	mass flow rate $[kg/s]$
α	convective heat transfer coefficient $[W/m^2/°C]$
λ	conductive heat transfer coefficient $[W/m/°C]$
k	overall heat transfer coefficient $[W/m^2/°C]$
C_p	specific heat, p=const $[J/kg/°C]$
t	temperature [°C]
Δt	temperature difference [°C]
F	surface $[m^2]$
A	cross section area of flow $[m^2]$
di	latent heat $[J/kg]$
q	heat flux, performance $[W]$
Re	Reynolds number $[-]$
Pr	Prandtl number $[-]$
C	coefficient $[s/kg^2/m]$
ρ	density $[kg/m^3]$
δ	thickness $[m]$
d	diameter $[m]$
ξ	coefficient of flow resistent $[-]$

Subscripts and superscripts

v	water
f	refigerant (freon)
i	input
o	output
m	middle
1	vapor cooling section
2	condensation section
p	pressure

1 Introduction

The condenser plays a decisive role to the efficiency of the heat pump. The heat transfer between the refrigerant and the heated water is formed in the condenser. The efficiency of the heat transfer process is very important. On the efficiency of the heat transfer process have strong effects the construction of the condenser, the refrigerant and the heated water mass flow rate.

The answer to the extensive investigation of the interactions among the operating condenser's parameters can be obtained by measurements or by the solution of the mathematical model. The mathematical model have been developed to correspond to the above-defined tasks. In the center of the mathematical model is the condenser as well as the refrigerant and water to transport the heat.This study does not take into consideration the compressor's and circulating pump's physical and mathematical model. It is the topics of the future investigation.

The investigated condenser is plate exchanger. Within the condenser during the processing are formed two parts. The first part is the superheated vapor cooling section, the second is the condensation section itself. The aim of research is to investigate the interaction of condenser and its environment in terms of energy efficiency. More precisely, the determination and investigation of condenser's coefficient of performance-COP depending on the compressor and circulating pumps. In this paper the influence of the compressor and circulation pump on the perfomances of the condenser is taken into account through the mass flow rate and inlet temperature of the refrigerant and hot water.The investigation of the COP can only be justified in steady state operation mode. The transitional operation mode occurs only in the case of switching the heat pump on/off and possibly in the case of changing the system parameter in the time.

On the base the above specified aim the mathematical model is steady state, with lumped parameter and coupled algebraic equations. In the mathematical model the equations are divided into the governing and auxiliary equations. The governing are the energy balance equations for the heated water and the refrigerant. The auxiliary equation are the equations of the convective heat transfer coefficients and the latent heat of the refrigerant. In the studied system four independent variables are present, the refrigerant mass flow rate, the refrigerant inlet temperature, the water mass flow rate and the water inlet temperature. Changing the independent variables the condenser's performance and the coefficient of performance-COP or rather energy efficiency can be changed, as well.

In principle, each of the four independent variables can be changed simultaneously but in this case separatly the effect of the variables to the condenser performance will not be possible to detect. It is preferred to change two independent variables at the same time during the investigation of the condenser performance. Four possible combinations exist when two system parameters is changing. But the most physically relevant investigation is the change of the heat transfer in condenser depending on the refrigerant mass flow rate and the heated water mass flow rate.

Furthermore, the simulation of the remaining three possible combinations is also carried out. This was the change of heat transfer in condeser depending on the refrigerant mass flow rate and inlet temperature, the water mass flow rate and inlet temperature as well as the refrigerant and water inlet temperature.

To solve the system of equations in the created mathematical model in algebraic form is almost not possible. Some numerical method is applicable only. The Newton-Taylor numerical method was used to solve the system of equation. The obtained numerical results can be displayed in a very suitable three dimensioned form. The great advantage of graphics is that the behavior of the condenser's performance can be seen over a wide range of the parameters.

In the scientific press has many articles researched the heat pump's behavior in stationary operating mode used various mathematical models with lumped parameters. Many of the mathematical models contain all four components including the condenser, however, the goal of research is focused on the behavior of the entire system rather than individual components. A smaller number of articles that dealt with the study of the condenser 's behavior as a component of a heat pump.

Hatef Madani and al. [1]. Have dealt with the investigation of control of the heating system using heat pump. In order to investigate the control of the system are developed a mathematical model with lumped parameters and for stationary regime. As part of the complete mathematical model developed the mathematical model of the condenser, too. The description of the condenser are used well known equations based on conservation energy. As a system parameter used enthalpy. Elias Kinab and al. [2]. Their aim was investigating the optimal seasonal performance of reversible heat pump system. Mathematical model of the condenser is created in analytical form based on the data of the existing physical condenser. Hongtao Qiao and al. [3]. Set a new mathematical model of the convective heat transfer coefficient to the refrigerant and water within the plate condenser. Róbert Santa [4]. Compared the results of condensing heat transfer coefficients of several authors. According to him paper the best characteristics showed Shah model among the investigated models. Yi-Yie Yan and al. [5]. Created a new mathematical model of condensing heat transfer coefficient of the plate condenser and carried out the model verification at the factory realized condenser.

The above articles mainly investigated the processes within the condenser in terms of the heat exchange and the results are presented in two-dimensional graphics. In the enclosed article, research has focused on investigate the efficiency of heat exchange in the condenser but as a function of external influences.

Results obtained by numerical simulation are presented in the three-dimensional graphics. Three-dimensional graphics are very suitable for the analysis because it provides a wide range of possibilities of displaying heat efficiency of the condenser, depending on the two external influences the same time.

2 Physical Model

The physical model includes the condenser, the refrigerant and the water for transport the heat. The created model does not include the compressor and circulating pump. The compressor and circulating pump’s effect on the condenser's performance through the refrigerant and the water mass flow rate and inlet temperature. The investigated condenser is plate heat exchanger. Inside the condenser the refrigerant and the water may stream in the co-current or counter-current direction.

The direction of the stream has a negligibly small effect on the condenser's performance because the temperature of the condensation can be considered as constant. The near constant temperature is the result of the low pressure drop across the condensor. In the superheated vapor cooling section the counter flow is more preferable. The four independent variables are the refrigerant and the water mass flow rate and the inlet temperature. These four variables can be changed independently of each other.

Incorporating the compressor and circulation pump in a mathematical model situation something is changing. Independent variable mass flow rate of water will be dependent on the performance of the circulating pump, and refrigerant's mass flow rate, of the compressor's performance.

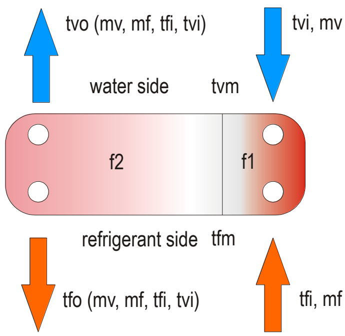


Figure 1
Heat Pump's Condenser with flow of refrigerant and hot water

3 Mathematical Model

3.1 General Description

The condenser’s mathematical model was developed with lumped parameters and on steady-state operation mode. The governing equations are formed based on the energy balance equations separately on the vapor section, and separately on the condensate section.

Within each section three balance equations can be formed. The balance equation of refrigerant, water, and the balance equation of the heat transfer between them. By analyzing the energy efficiency only the steady-state operation mode is interesting. The dynamic-state may take up to 1% of the operation time.

3.2 Approximations

The next approximations have been applied during the modeling.

- lumped parameters
- steady-state operation mode
- pressure drop in plate exchanger is small therefore are neglected.
- the condenser's wall thermal resistance is too small so it has been neglected.
- specific heat of water is taken as constant because weakly depends on the temperature.
- specific heat of refrigerant vapor is taken as constant because weakly depends on the temperature.
- the condenser's heat loss to the environment is neglected.

3.3 The Governing Equations

3.3.1 Vapor Cooling Section

Heat delivered by the refrigerant vapor.

$$q_1 = C_{pf} \cdot \dot{m}_f \cdot \Delta t_{f1} \tag{1}$$

Heat transferred through the condenser in the vapor cooling section.

$$q_1 = k_{f1} \cdot F_1 \cdot \Delta t_{ln1} \tag{2}$$

Heat received by the water.

$$q_1 = C_{pv} \cdot \dot{m}_v \cdot \Delta t_{v1} \tag{3}$$

3.3.2 Condensing Section

Heat delivered by the condensation of the refrigerant.

$$q_2 = \dot{m}_f \cdot \Delta i_f \tag{4}$$

Heat transferred through the condenser in the condensation section.

$$q_2 = k_{f2} \cdot F_2 \cdot \Delta t_{ln2} \tag{5}$$

Heat received by the water.

$$\mathrm{q_2 = C_{pv} \cdot \dot{m}_v \cdot \Delta t_{v2}} \tag{6}$$

3.4 Auxiliary Equations

Logarithmic mean temperature difference

$$\Delta t_{ln} = \frac{\Delta t_{max} - \Delta t_{min}}{ln \frac{\Delta t_{max}}{\Delta t_{min}}} \tag{7}$$

Overall heat transfer coefficient

$$k = \frac{1}{\frac{1}{\alpha_v} + \sum\frac{\delta}{\lambda} + \frac{1}{\alpha_f}} = \frac{\alpha_v \cdot \alpha_f}{\alpha_v + \alpha_f} \tag{8}$$

The water's convective heat transfer coefficient. [4]

$$\alpha_v = 0.2121 \cdot \frac{\lambda_v}{d_{ekv}} \cdot Re_v^{0.78}(\dot{m}_v) \cdot Pr_v^{1/3} \tag{9}$$

The refrigerant vapor's convective heat transfer coefficient when x=1. [4]

$$\alpha_{fs} = 4.118 \cdot \frac{\lambda_{fl}}{d_{ekv}} \cdot Re_f^{0.4}(\dot{m}_f, x) \cdot Pr_{fl}^{1/3} \tag{10}$$

The condensing heat transfer coefficient is the arithmetic mean value by increment dx = 0.1 for the interval x = 0-1. Using equation (10) [4]

$$\alpha_f = \frac{\sum_{i=0}^{i=n} \alpha_{fs}(\dot{m}_f, x_i)}{n+1} \tag{11}$$

$$x_{i+1} = x_i + dx$$

$$x_0 = 0, \qquad x_n = 1, \qquad dx = 0.1$$

The latent heat as the function of the condensation temperature.

$$\mathrm{di} = \mathrm{a_o} + \mathrm{a_1} \cdot \mathrm{t_{fo}^1} + \mathrm{a_2} \cdot \mathrm{t_{fo}^2} \tag{12}$$

Where the constants of the refrigerant R 134a are:

$$\mathrm{a_o} = 200.5965715$$

$$\mathrm{a_1} = -0.709168$$

$$\mathrm{a_2} = -0.00596796$$

The condenser's total surface.

$$F = F_1 + F_2 \tag{13}$$

The total amount of heat transferred through the surface of the condenser.

$$q = q_1 + q_2 \tag{14}$$

3.5 The Condition of Linking the Two Parts of Model

Change of refrigerant phase are the cause of the forming two parts of the model. The condition of linking the two parts of the model is the condensing temperature and the temperature of the cooled refrigerant vapor on the end of first section must to be equal

$$t_{fm} = t_{fo} \tag{15}$$

4 Mathematical Method

A mathematical model of the heat pumps condenser based on lumped parameters and in steady-state. The model is stationary so the parameters do not depend on the time or on space because the parameters are lumped. Thanks to these two assumptions, equations in the model are in algebraic format, without exception. Regardless, analytically to solve the mentioned mathematical model due to the complexity is not possible.

The model's complexity and size has almost no restrictions imposed on the numerical method. The mathematical equations in the model can be coupled, nonlinear and implicit. Applied numerical method is iterative therefore it must be assumed the appropriate initial solutions.

The non-linear system of equations has multiple solutions so the initial values have to be determined properly. The solution of the non-linear system of equations converges to the initial value but maybe it is not the right solution. In such a case, the initial values have to be changed.

In this study, a well-known numerical method, such as the Newton linearization and Gauss elimination method has been applied. The governing equations are connected and must be arranged in implicit form. Before the starts of the first iteration each variable should get the initial value. As it was noted earlier, the system of equations is non-linear therefore it is very important to determine the initial value of the solutions.

The simulation is running in a wide range and in three dimensions. Simulations made 9OO calculated points. The calculation of the variables in every pont is done iteratively, while the calculation of the next point can be considered a recursive step, since the initial values of solution was equalized with the final results from the previous step. This solution offeres a convergence and that is good if the initial values of solution are close to the values of the solution.

4.1 Implicit Mathematical Model of the Condenser

The condition of using the Newton-Taylor's linearisation method the equations must be written in an implicit form.

$$f_1 = -q_1 + C_{pf} \cdot \dot{m}_f \cdot \Delta t_{f1} = 0 \tag{16}$$

$$f_2 = -q_1 + k_{f1} \cdot F_1 \cdot \Delta t_{ln1} = 0 \tag{17}$$

$$f_3 = -q_1 + C_{pv} \cdot \dot{m}_v \cdot \Delta t_{v1} = 0 \tag{18}$$

$$f_4 = -q_2 + \dot{m}_f \cdot \Delta i_f = 0 \tag{19}$$

$$f_5 = -q_2 + k_{f2} \cdot F_2 \cdot \Delta t_{ln2} = 0 \tag{20}$$

$$f_6 = -q_2 + C_{pv} \cdot \dot{m}_v \cdot \Delta t_{v2} = 0 \tag{21}$$

$$f_7 = -di + a_o + a_1 \cdot t_{fo}^1 + a_2 \cdot t_{fo}^2 = 0 \tag{22}$$

4.2 Newton-Taylor's Linearisation Method of Implicit Non-Linear Equations

The (16) - (22) implicit equations have to develope in the Taylor series. Implicit discretized equation after developed in the Taylor series in 1 + i iteration.

$$f_{i+1} = f_i + \frac{\partial f_i}{\partial x} \cdot dx + \cdots \tag{23}$$

From the above relation the increment of variable between i and i +1 iterations can be expressed.

$$dx = \frac{f_{i+1} - f_i}{\frac{\partial f_i}{\partial x}} \tag{24}$$

In the i +1 iteration, the new initial values of the variables

$$x_{i+1} = \mathrm{dx} + x_i \tag{25}$$

5 Simulations

Several independent variables were investigated to explain the behaviour of the heat pump's condenser. The investigations carried out by the mathematical model and the numerical method.

The independent variables are:

• the compressor output parameters, the refrigerant mass flow and the superheated vapor temperature.

• the hot water flow rate circulated circulation pump and inlet temperature.

The condensator performance was simulated in three dimensions.

The simulation includes the following investigations:

a. The condenser performance as a function of the heated water mass flow rate and refrigerant mass flow rate.

$$q = f(\dot{m}_f, \dot{m}_v)$$

b. The condenser performance as a function of the heated water and the refrigerant inlet temperature.

$$q = f(t_{fi}, t_{vi})$$

c. The condenser performance as a function of the refrigerant mass flow of superheated vapour temperature.

$$q = f(\dot{m}_f, t_{fi})$$

d. The condenser performance as a function of the heated water mass flow rate, the inlet temperature

$q = f(\dot{m}_v, t_{vi})$

6 Simulation Results

On the basis of the created mathematical model, the Newton-Taylor linearisation and Gauss iteration the obtained results are presented in the Figures 2, 3, 4.

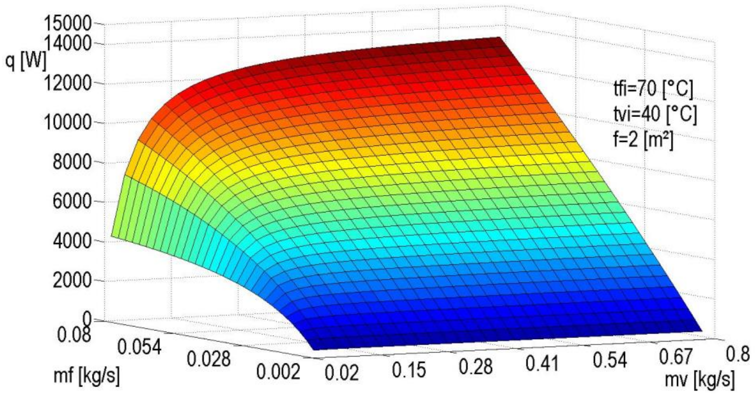


Figure 2

Condenser's performance as a function of the hot water and the refrigerant mass flow rate

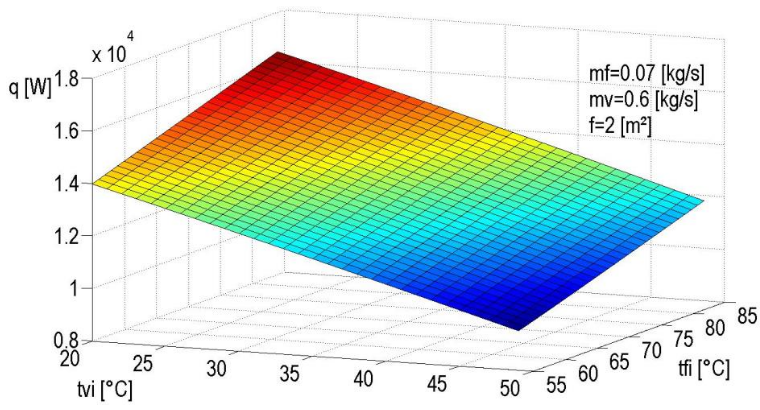


Figure 3

Condenser's performance as a function of the hot water and the refrigerant mass flow rate

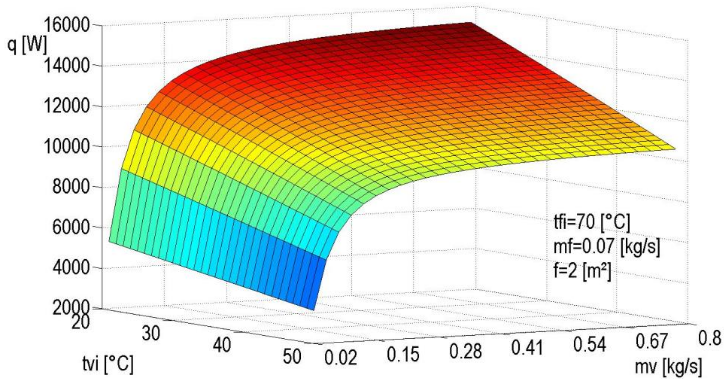


Figure 4

Condenser's performance as a function of the hot water mass flow rate and the inlet temperature

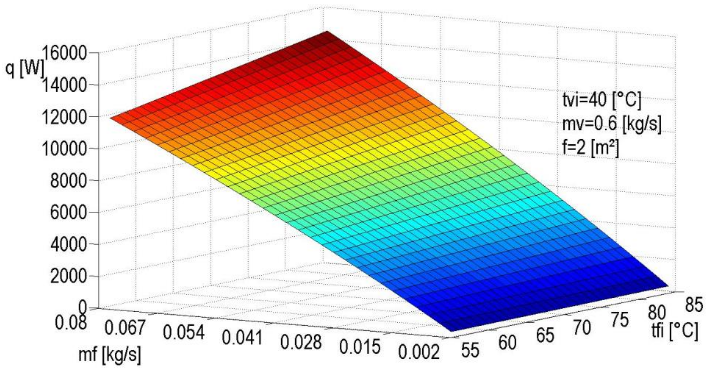


Figure 5

Condenser's performance as a function of the refrigerant mass flow rate and the inlet temperature

Conclusions

a. Comments on the mathematical model

- the mathematical model is stationary with lumped parameters
- difference of temperatures is logarithmic
- the convective heat transfer coefficients depends on the mass flow rate
- latent heat depends on the condensation temperature
- the thermal resistance of the thin wall of condenser is very small, therefore it has been neglected.

b. Comments on the numerical methods

- the numerical method is iterative with Newton-Taylor linearization and Gaussian elimination method.

- the values of the finally numerical solutions largely depend on the initial solutions.
- in the case of improper initial solutions, the obtained solutions are not physically appropriate.
- the investigation of the obtained numerical results is possible in graphical form
- the applied graphics are three-dimensional

c. Simulation results

- the created mathematical model and the applied numerical method provide a wide range of opportunities to investigate the heat pump's plate condenser in the steady-state operating mode
- the research extended to investigate the plate condenser performance as a function of independent input variables.
- the independent input variables are the refrigerant and water mass flow rates and temperatures.

Based on the simulation results, the conclusions are as follows:

a. The condenser's performance increases nonlinearly as a function of hot water and refrigerant mass flow rate.
b. In the case of zero mass flow rates the condenser performance is zero. For the range of low quantity of the mass flow rates, the performance improvement is extremely intense.
c. The further linear increase in the mass flow rates resulted an asymptotic increasing of performance. It is not worth increasing the mass flow rates after certain quantity because the increase of performance is negligibly little. The investment in flow energy may not be recoverable.
d. The condenser's performance mildly linearly increases as a function of the hot water and the refrigerant temperature difference.
e. The condenser's performance nonlinearly increases depending on the increases of the hot water mass flow rate.
f. If the water inlet temperature increases, the performance decreases proportionally.
g. The condenser's performance weakly non-linear increases if the refrigerant mass flow rate in growth. The increases of the refrigerant's inlet temperature resulted in a clear linear increasing of performance.

New scientific results

- The new formulation of the investigation of heat pump condenser's performance.

- The condenser is divided into cooling and condensation sections.

- The superheated vapor cooling in the condenser, was taken into account

- The convective heat transfer coefficients depend on the mass flow rates and temperatures.

- The latent heat depends on the condensing temperature

• For the purpose a suitable complexity mathematical model created

- The model consists of two coupled parts, cooling and condensing section

- The two models are coupled with an internal moving boundary

- Determined the cooling and condensing surface area

• The presentation of the obtained numerical results by three-dimensional graphics

- The new numerical package enables the 3D graphical presentation of results therefore thus provides the comprehensive investigation of condenser's performance and evaluation of new results.

References

[1] Hatef Madani, Joachim Claesson, Per Lundqvist: „Capacity Control in Ground Source Heat Pump Systems Part I: Modeling and Simulation" I J Refrigeration, Vol 34 (2011), pp 1338-1347

[2] Elias Kinab, Dominique Marchio, Philippe Riviere, Assaad Zoughaib: „Reversible Heat Pump Model for Seasonal Performance Optimization" I J Energy and Buildings, Vol 42 (2010), pp 2269-2280

[3] Hongtao Qiao, Vikrant Aute, Hoseong Lee, Khaled Saleh, Reinhard Radermacher: "A New Model for Plate Heat Exchangers with Generalized Flow Configurations and Phase Change" I J Refrigeration Vol. 36 (2013) pp. 622-632

[4] Róbert Sánta: „The Analysis of Two-Phase Condensation Heat Transfer Models Based on the Comparison of the Boundary Condition" Acta Polytechnica Hungarica Vol. 9, No. 6, 2012, pp 167-180

[5] Yi-Yie Yan, Hsiang-Chao Lio, Tsing-Fa Lin: "Condensation Heat transfer and Pressure Drop of Refrigerant R-134a in a Plate Heat Exchanger" International Journal of Heat and Mass Transfer Vol. 42 (1999) pp 993-1006

[6] Imrich Bartal, Hc László Bánhidi, László Garbai: "Analysis of the Static Thermal Comfort Equation" Energy and Buildings Vol. 49 (2012) pp 188-191

[7] Garbai L., Jasper A.: A matematikai rendszerelmélet feldolgozása és alkalmazása épületgépészeti optimalizációs feladatok megoldására; Magyar Épületgépészet LX. évfolyam, 2011/3. szám, pp. 3-6

[8] Garbai L., Méhes Sz.: Energy Analysis of Geothermal Heat Pump with U-tube Installations. IEEE International Symposium CFP 1188N-PRT "EXPRES 2011." Proceedings, pp. 107-112. Subotica, Serbia. 11-12 03. 2011

[9] M. M. Awad, H. M. Mostafa, G. I. Sultan, A. Elbooz, A. M. K. El-ghonemy: Performance Enhancement of Air-cooled Condensers, Acta Polytechnica Hungarica, Volume 4, No. 2, pp. 125-142, 2007

[10] Nyers J. Stoyan G.:" A Dynamical Model Adequate for Controlling the Evaporator of Heat Pump", Internationale Journal of Refrigeration, Vol. 17, Issue 2, pp. 101-108, 1994

[11] Nyers J., Nyers A.: "COP of Heating-Cooling System with Heat Pump "IEEE International Symposium CFP 1188N-PRT "EXPRES 2011" Proceedings, pp. 17-21, Subotica, Serbia, 11-12 03 2011

Impact of Perceived Brand Name Origin on Fashion Brand's Perceived Luxury

Zoran Krupka, Durdana Ozretic-Dosen, Jozo Previsic

Faculty of Economics and Business, University of Zagreb
J. F. Kennedy Square 6, 10000 Zagreb, Croatia
E-mail: zkrupka@efzg.hr, dozretic@efzg.hr, jprevisic@efzg.hr

Abstract: The aim of this paper is to investigate the impact of perceived brand name origin and perceived country of origin (COO) on fashion brand's perceived luxury. Results of exploratory research indicate that perceived brand name origin has a strong impact on perceived luxury, and subsequent willingness to pay a price premium, regardless whether the brand is fictional and only sounds as if originated from a certain country (e.g. "Italian sounding brand") or is a real brand. Results also indicate that by naming fashion brands to have a desirable (e.g. Italian) tone, potential negative impacts of real COO and even a country of manufacture (COM) could be mitigated.

Keywords: Brand Name Origin; Luxury Fashion Brands; Country of Origin; Fashion Marketing

1 Introduction

Luxury brand management has been a topic of interest to many scientists and practitioners during the last decade. That is not surprising because luxury sector has grown from a value of $ 20 billion in 1985 [27] to its current over € 200 (estimated $ 260) billion worth and it is expected to grow 2-4 percents annually for the next two years [3]. Factors like growing globalization, development of digital and communication media, opportunities for easy money earning, countries like China and Russia opening for foreign brands, and cultural homogenization [8] have led to global success of numerous luxury brands, especially luxury fashion brands.

But not all countries have a pedigree for developing luxury brands. That is even more obvious in fashion where over 80 percent of all luxury brands originate from only a few countries [27]. Because of that, it is very important from which country the brand is perceived to originate. That leads us to another, very important factor in developing luxury fashion brands - Country-of-Origin construct (COO).

Although, COO has been present in literature for over five decades, it is still topic of many current studies (e.g. [35], [5], [28]).

COO effect is one of several extrinsic attributes that potentially influence consumer attitudes towards a product [9]. Approximately one-quater of consumers make purchase decisions on the basis of COO information [23]. Another extrinsic attribute that influence consumers perception of product is definitely brand name. It has been suggested that the brand name cue evokes not only beliefs about the brand itself but also triggers recall of the country associated with it as its COO [19] and that both have broad and specific effects on consumer behavior [31].

Topics like country of origin and luxury brand management, lately even fashion marketing, are well studied, but there is still lack of research in the area of impact of brand name's perceived origin on perceived brand luxury. Although, there were some research that were pointed in that direction [2] [23], need for further understanding of this important issue still exists. In addition, most research was done in developed countries [6], thus omitting the understanding of developing countries which are opening up for luxury brands and thus becoming increasingly of interest to luxury brand marketers.

The question that this research aims to answer is whether it is possible, by selecting a brand name that is associated with desirable COO, to affect the perceived brand's luxuriousness and to mitigate potentially negative impact of true COO and even a Country-of-Manufacture (COM).

2 Literature Review

2.1 Luxury Fashion Brands

Generally speaking, brand is a promise of performance to be delivered in return for the trust placed in a brand [37]. Thus brands embody the symbol, an implicit value proposition, and personality.

Luxury brands from an economics perspective can be defined as those whose ratio of functional utility to price is low while the ratio of intangible and situational utility to price is high [26]. From the product perspective, luxury brands are frequently defined in terms of their excellent quality, high transaction value and distinctiveness [16]. In addition to possible functional attributes, luxury brands are even more bought for their symbolic values arising from exclusivity, premium prices, image, status [18], premium quality across product lines, heritage of craftsmanship, recognizable style and design [7].

Level of brand's luxuriousness is "in the eye of the beholder", i.e. presents a perceptual category, influenced by cultural specificities [14]. While luxury products are often purchased simply because they cost more, without providing any additional direct utility over their cheaper counterparts [13], luxury brands may be used to enhance one's social status [22]. This is also implicit in the ideas of Christodoulides, Michaelidou and Li [11] who reported that the primary value of luxury brands is psychological, and that their consumption is dependent on a distinctive mix of social and individual cues.

As stated earlier, the luxury goods market is significant, not only in terms of its market value, but also in terms of its rate of growth [3] [16]. According to Seringhaus [34] France and Italy are the most important luxury brand source countries in the world. Those two countries command more than one-half of the global consumer luxury brands - Italy with 30 percent and France with 25 percent of the market. Taking into account fact that Gucci, Chanel, Dolce & Gabbana, Cavalli, Armani, Yves Saint Laurent, and Lacroix are (just some of) Italian or French luxury brands, the share of luxury brands in fashion industry is expected to be even higher. The success around the world of items "Made in Italy" is mostly due to the Italian brand's ability to transfer a certain sense of product quality in concert with values and experiences of beauty, elegance, tradition, luxury, and life quality [36].

Brand's association with country of origin tends to have a strong impact on brand's reputation, especially in a specific industry such as fashion [26]. Piron [32] stated that a product's country of origin has a stronger effect when considering luxury products (i.e. higher monetary risk) than necessities. Thus, for luxury brands, especially for luxury fashion brands, COO presents a make-or-break of brand's market success.

2.2 Country of Origin and Luxury Brands

Product's COO presents an important cue that consumers use to make their judgments and their subsequent purchase behaviors [29]. Importance of country of origin effect has mostly been discussed in the consumer behavior and international marketing literature (e.g., [25], [15]), while being relatively neglected in the brand management literature. Moreover, in consumer behavior literature, COO has been primarily analyzed as the real country of origin, while in brand management literature, it makes more sense to focus on the perceived COO and its impact on brand success [5]. Brand management literature has generally analyzed the impact of the actual COO on brand image (e.g. [1], [10]), but recent research on country of origin has shown that consumers often do not know the real origin of many (even of the well-known) brands [5].

Country of origin can be defined as an extrinsic product attribute indicating the country where a product was made, was assembled, or both [24]. This definition, although widely accepted, implies some operationalization difficulties. Thus, in this research we analyze the COO as the country in which corporate headquarters of the company marketing the brand is located, regardless of the place in which the brand in question is produced [4].

Consumers tend to recall the stored information about the brand and the country and then they relate the brand name with the COO to form a brand image and infer the product evaluation [21]. When a product's COO is known to consumers prior to exposure to the product's attributes, then the image of the country as a country of origin affects the brand's image. Because of that, it is very important to explore what is the impact of COO on perceptions of brand luxury so that companies know what strategy to apply (whether to emphasize country of origin or to depart from it).

When we consider fashion as industry, it is clear that some countries (e.g. Italy and France) tend to be perceived as countries that have an innate fashion capability [34]. Such perceptions provide companies from those countries a certain competitive advantage and privileges for building luxury brands in the fashion industry. Numerous aspects provide grounds for emphasizing country-brand associations, and one of the most prominent is surely brand name which reflects country's cues [33].

2.3 Brand Name Origin And Luxury Perceptions

Thakor and Kohli [37] define brand-origin as the place, region, or country where brand is perceived to belong by its target consumers. Past studies suggested that brand name origin associations play potentially powerful role in the formation of brand attitudes [38] [30] [39] and that the brand name cue evokes not only beliefs about the brand itself but also triggers recall of the country associated with it as its country of origin [19].

Products with foreign brand names are frequently associated with the perceived COO of the brand [20]. Even in case of completely unfamiliar brands, consumers would develop some expectations and perceptions of that brand based on image of a country that seems to be associated to the brand name. Morover, literature recognizes [24] that countries are perceived to be experts in creating and producing certain product category (e.g. Japan and electronic products; USA and jeans; Germany and cars etc.). Because of that, COO image that those countries have for that exact product category is positive. So, many manufacturers and retailers use brand names that suggest origins different from the brands' real COO or COM and associate them with countries that have positive country of origin image for that product category. For example, Alba, a British manufacturer of electronic equipment, has introduced the Hinari brand, a name that does not convey the real origins of the product [4].

Literature has analyzed the role of real country of origin on consumer perceptions and brand name on consumer perceptions (e.g. [12]). However, this research aims at understanding how brand name itself can be a cue for perceived country of origin and thus generate positive brand associations. In that sense, we expect that the mere perceived origin of a brand name can transfer country-related beliefs to the brand.

Han [17] suggest that brand name may be even a more powerful summary construct than the COO. In that sense, brand name in itself implies existence of a powerful "story" linked to that brand, thus making country of origin effect less important in case of well-known brands. Moreover, if brand name is closely linked to (i.e. sounds as if it originates from) a certain country, we expect that the perceived COO resulting from that association will present a much stronger cue than possible cues arising from either real COO and/or COM associations [31].

Pecotich and Ward [31] suggested that country of origin or brand-origin operates in one of two ways. First, it may serve as a hallo effect to infer beliefs about attributes that make up the attitude towards a product or service, i.e. consumer evaluations of products and services are based on their perception of the country. And second, it may be used as a means of abstracting previous beliefs about attributes of products and services from a particular country into information called the summary construct.

Country of origin concept is still the topic of many current research projects (e.g. [5]) as it implies that brands would succeed by "nature" (i.e. by their COO) and not "nurture" (i.e. management that overturns potential negative COO effects). Although, topics like country of origin and fashion marketing, lately even luxury brand management, are well studied, there is lack of research in the area of impact of brand name's perceived origin on perceived brand luxury [33]. Even there was some research that has pointed in that direction (e.g. [2]), further research into nurturing brands to diminish possible negative COO effects is still needed.

3 Research

3.1 Objective and Hypotheses

The main purpose of this paper is to fill the gap in existing literature and research by investigating connection between perceptions of the country of origin given a brand name with perceived brand luxury in fashion marketing.

Based on the literature review, we propose two hypothesis:

H1: Perceived brand's COO has an impact on brand's perceived luxury.

H2: If a brand name is associated with a desirable COO, the COM does not have an impact on perceived luxury.

3.2 Methodology

This research presents a first step in our research project aiming to understand the complex relationship between brand's perceived and true country of origin, on one side, as well as the impact this relationship may have on brand's performance, leading to potential strong brand management implications (e.g. to brand luxury).

Data was collected on the sample of 276 younger consumers in Croatia. Each respondent filled out a questionnaire for 8 different brands, some of which were fictional (4 in total) and some real (4 in total) thus resulting in the total of 1808 data points after clearing out the dataset (deleting incomplete responses; 50 or 18.11% of returned questionnaires were incomplete).

Although having sample limited to young consumers presents a certain limitation, it is consistent with other research on related topics (e.g. [30]) and is only the first step in the overall research project. Moreover, Bain and Forsythe [8] reported that younger consumers represent current and important future market for luxury brands, they have knowledge about them and they are establishing life-long buying patterns and loyalties.

Given that real fashion luxury brand names are often personal names, fictional brand names were developed by randomly picking 20 different first-last names in France, Germany, Italy and USA from phonebooks. Further, a pretest was done on the sample of 8 experts in the fashion industry to rate which of the names seem the most probable brand names for luxurious fashion products. Questionnaire was pretested on a sample of 25 respondents.

3.3 Results

In terms of the demographic profiles of all 226 respondents, 171 or 75.66% were female and 55 respondents or 24.34% were male. Of all respondants, 89 of them (39.38%) were between age of 18-21, 102 respondents (45.13%) were between age of 22-25 and 35 respondents (15.49%) were between 26-29 years old.

Dependent variable in this research was brand's perceived luxury which, for robustness reasons, was measured in three different ways: using an existing reflective scale, using a formative scale, and using a single-item measure of whether the brand is luxurious or not. Results were tested using the three different measures of the dependent variable and there were no significant differences between the models (see Table 1). Thus for simplicity reasons we present our results only with a reflective scale of perceived brand luxury. Moreover, as this is the first step in our research project, in the regressions we did not focus on control variables but on the main effects.

Table 1
COO Impact on Perceived Brand Luxury (both fictional and real brands) – use of different scales for dependent variable

	Brand luxury Formative scale	Brand Luxury Reflective scale	Brand Luxury Single Item
COO_AUSTRIA	-,057*	-,055*	-,065*
COO_BELGIUM	,022	,024	,026
COO_FRANCE	,236***	,244***	,190***
COO_GERMANY	-,052	-,054	-,080
COO_ITALY	,246***	,239***	,226***
COO_ISRAEL	-,053*	-,047*	-,061**
COO_CANADA	-,060*	-,057*	-,042
COO_NEITHERLANDS	-,007	-,005	-,013
COO_SPAIN	-,012	-,013	-,025
COO_SWISS	-,046*	-,048*	-,049*
COO_UK	,062	,069	,051
COO_USA	,037	,038	,010
R^2	,104	,105	,100
F	17,394***	17,389***	16,643***

** p<,05; ** p<,01; *** p<,001*

First we tested how familiar respondents were with the fictional brands. Results indicated that 12,9% believed they have heard of fictional brands while 87,1% did not. Also, average familiarity with fictional brands was 1,38 (on a 5 point scale), i.e. significantly below the theoretical mean of 3 (t=-77,457; sig=,000). Thus, selected brands were unfamiliar to customers. This implies that brand names were well chosen in the sense that they did not trick all the participants (e.g. what might have happened if we only altered the spelling of existing brands). However, they were good enough to "trick" 12,9% of participants to believe that they have heard of the brand before.

When fictional and real brands are jointly considered, results indicate that perceived COO has an impact on Perceived Brand Luxury (R^2=,109) with brands originating from Italy (β=,239; p<,001) and France (β=,244; p<,001) having the strongest positive impact (see Table 2). When we consider only fictional brands the effect of perceived COO of provided brands is still significant. When only real brands are considered, the impact of perceived COO is even stronger.

These results indicate that H1 can be accepted and that perceptions of country of origin do influence the brand's perceived luxury, especially if perceived COO is Italy or France. Thus, even if luxury fashion brands are originating from countries other than Italy and France, they could benefit if their brand names are set to mimic Italian or French origin.

Table 2
COO and COM Impact on Perceived Brand Luxury

	COO			COM		
	All brands	Fictional	Real	All brands	Fictional	Real
AUSTRIA	-,055*	-,046		-,037	-,034	
BELGIUM	,024	,056*		,037	,042	
CHINA				-,003	-,074	-,076
EU				-,017	-,027	-,006
FRANCE	,244***	,111*	,341*	,255*	,107	,289*
GERMANY	-,054	-,077	,070	-,016	-,070*	,071
INDIA				-,021	-,038	,042
INDONESIA				-,021	-,053	-,038
ITALY	,239***	,147*	,342*	,256*	,100*	,281*
MEXICO				-,052*	-,064*	-,086
NEITHERLANDS	-,004	,009		,001	,014	
ROMANIA				,022	-,010	,029
SPAIN	-,013	-,014	,050	,002	-,002	,044
TAIWAN				-,020	-,067*	-,036
TURKEY				,010	-,064*	,074
UK	,069	-,006	-,092	,046	-,019	-,058
USA	,038	-,038	-,038	,060*	-,018	-,020
VIETNAM				,021	,013	-,066
R^2	,109	,067	,182	,119	,067	,189
F	17,389***	7,836***	17,203***	10,391***	4,095***	5,517***

** p<,05; ** p<,01; *** p<,001*

Results regarding the COM indicate that brands perceived as produced in France (β=,255; p<,05), Italy (β=,256; p<,05) and USA (β=,060; p<,05) are perceived as more luxurious. When focusing on the real brands, the same holds true only for France and Italy with no negative impacts on perceived luxury of where the well-known luxury brands are actually produced.

When we focus on the fictional brands, there is a positive impact only of Italy as COM, while fictional brands perceived as produced in other countries exhibit either negative (Germany, Mexico, Taiwan, Turkey) or non significant impact (see Table 2). This could be since these countries are generally not linked to high fashion industry.

In this respect, we were expecting that China as the country of manufacture would diminish positive country of origin effects on brand's perceived luxury. However, contrary to what is often believed, China as COM has no significant impact on the perceived brand luxury. Such results might imply that consumers have gotten used to the idea that fashion products tend to be produced in China and thus believe that Chinese production has achieved a level of quality that is acceptable and/or they accept that fashion products tend to be produced in countries like China.

When we analyze the impact of perceived COM, depending on the fictional brand's perceived COO, results indicate that there is no effect of country of manufacture only for brands that are believed to originate from Italy (see Table 3).

One possible explanation is that Italian luxury brands stress their design while brands originating from countries like USA, France and Germany communicate their quality. In that sense, if perceived country of origin is Italy, then consumers still recognize the luxury stemming from the design, while they might have a problem accepting that high quality German brand is produced abroad.

Thus, H2 is partially accepted. If perceived COO is so strongly linked to luxury (like in the case of Italy), it does not matter where the product is actually originating from or manufactured. However, if perceived country of origin is not so strongly linked to luxury, COM exhibits significance. Further research is planned in order to understand the mechanism between different facets of luxury (e.g. luxury grounded on design, quality, etc).

Table 3

COM Impact on Brand's Perceived Luxury if a fictional brand is perceived to originate from Italy, France, Germany, USA

	Fictional Italian	Fictional French	Fictional German	Fictional USA
COM_AUSTRIA			,001	
COM_BULGARIA		-,034	,156*	
COM_CAMBODIA	-,020	-,049	,006	-,201*
COM_CHINA	-,041	-,179	-,270*	-,179
COM_EU		-,065	-,033	
COM_GERMANY			-,178	
COM_FRANCE		-,037		
COM_INDIA	-,028	-,152*	-,061	-,041
COM_INDONESIA	-,008	,021	-,145*	-,100
COM_ITALY	,004	,136*		
COM_MEXICO				-,121
COM_POLAND		-,002	-,130*	
COM_ROMANIA	,014	,019	-,089	
COM_TAIWAN	,012	-,156*	-,090	-,150*
COM_THAILAND	-,006	-,163*	-,074	,091
COM_TURKEY	-,094	-,095	-,159*	
COM_UK				-,008
COM_USA		-,058	-,017	-,050
COM_VIETNAM	,067	-,026	,022	-,068
R^2	,017	,114	,103	,097
F	,521	2,087*	1,774*	2,200*

* *p<,05; ** p<,01; *** p<,001*

When results for real luxury fashion brands are considered (Table 4), controlling for the perceived COO, there is no significant negative impact of COM on the perceived luxury. This is logical since consumers, if they know the brand, they evaluate the brand image and do not need other cues to make expectations about brand characteristics.

Table 4
COM Impact on Brand's Perceived Luxury if a real brand is perceived to originate from Italy, France, Germany, USA

	Real Italian	Real French	Real German	Real USA
COM_AUSTRIA				
COM_BULGARIA		,044		
COM_CAMBODIA				
COM_CHINA	,298	,085	,200	-,075
COM_EU				,009
COM_GERMANY			,491	
COM_FRANCE		,335		
COM_INDIA		,097		
COM_INDONESIA	,168			-,137
COM_ITALY	,585*			
COM_MEXICO				-,163
COM_POLAND		,097	-,109	
COM_ROMANIA	,059			
COM_TAIWAN	,117	,036	,256	-,168
COM_THAILAND	,112	,045		,152
COM_TURKEY	,359	-,017		,012
COM_UK				
COM_USA		,097		,058
COM_VIETNAM		-,104		,036
R^2	,077	,101	,167	,115
F	1,522	1,324	1,305	1,181

* *p*<,05; ** *p*<,01; *** *p*<,001

In addition, we have offered respondents 24 pairs of brands (real and/or fictional) used in a questionnaire and asked them for each pair to chose what brand would they buy. In each pair countries of origin of the brands were different.

When looking at the results for real brands paired together, in most cases Italian brands were chosen as ones that respondents would buy, followed by French, USA and German brands. When looking at pairs of fictional brands results are similar. Most respondent chosen Italian sounding brands, then French, USA and German sounding brands.

Results for pairs consist of real and fictional brands show that in most cases (89.38%) respondents would chose real brands over fictional ones. However, in 11.62% of cases respondents have chosen fictional over real brands. Their choices were mostly made in favor of Italian sounding fictional brand. That also imply that COO or at least perceived COO influence fashion brand's perceived luxury and consumers willingness to buy brands from certain country.

3.4 Limitation

The main limitation of this research is related to the sample. For this reasearch we have used a convenient sample od younger respondents, and therefore findings cannot be generalizable. However, a lot of research in luxury branding field (e.g. [8], [30]) have been made on convenient sample of younger consumers as there is no reason to expect systematic differences, other than disposable income role.

In addition, since the focus of our paper is on understanding the perceived brand name origin on fashion brand's perceived luxury, having younger consumers with no (or with limited) experience with such brands is actually beneficial. Such respondents will not be biased by performance of those brands but will base their perceptions only on expectations.

Conclusions

Brand management is of constant interest of both academics and practitioners. Fashion marketing and especially luxury brand management have become a hot topic lately because of sales volume achieved and growth potential even in the time of crises.

Brand naming is of great importance for luxury fashion brand management. Based on a brand name, consumers create brand expectations and perceptions. In the lack of other cues, brand's country of origin presents an important cue for making expectations on brand's luxury. Regarding this initial research into impact of perceived brand name origin and COO on fashion brand's perceived luxury, the mere name association to Italy or France as a country of origin makes the consumer perceive the brand as more luxurious.

In addition, we have also analyzed the impact of perceived country of origin on price premium consumers are willing to pay for the brand. And again the results indicate that Italian and French sounding brands can automatically expect consumers to be willing to pay a price premium.

Also, interesting conclusion is that potential negative impacts of country of manufacture could be mitigated by naming fashion brands to sound as if they are originating from a desirable and favorable country.

Although this research fills some of the gaps in the literature, there is still need for further research in this area. First of all, it would be beneficial to use representative sample so that data and conclusions could be generalizable. One of possible directions for the future research would be to understand the mechanism between different facets of luxury (e.g. luxury grounded on design, quality, etc). As it is noticed in the business, many fashion industry companies that originate from emerging countries that do not have positive COO (like China) begun to manufacture in countries that do have positive country of origin for luxury fashion (like Italy) simply to ensure that they could use the "Manufacture in (e.g.) Italy" label to enhance favorable associations toward the brands and mitigate unfavorable COO associations [23]. So, in the future research more attention should be given to the impact of country of manufacture on perceived luxury.

References

[1] Ahmed, S. A., d'Astous, A. (1996). Country of Origin and Brand Effects: a Multi-Dimensional and Multi-Attribute Study. *Journal of International Consumer Marketing*, *9 (2),* 93-115

[2] Aiello, G., Donovito, R., Godey, B., Pederzoli, D., Wiedmann, K-P., Hennings, N., Siebels, A., Chan, P., Tsuchiya, J., Rabino, S., Ivanova, S. I., Weitz, B., Oh, H., Singh, R. (2009). An International Perspective on Luxury Brand and Country-of-Origin Effect. *Brand Management*, *16 (5/6),* 323-337

[3] Bain and Company, www.bain.com (November 2012)

[4] Balabanis, G., Diamantopoulos, A. (2008). Brand Origin Identification by Consumers: A Classification Perspective. *Journal of International Marketing, 16 (1),* 39-71

[5] Balabanis, G., Diamantopoulos, A. (2011). Gains and Losses from the Misperception of Brand Origin: The Role of Brand Strength and Country-of-Origin Image. *Journal of International Marketing, 19 (2)*, 95-116

[6] Batra, R., Ramaswamy, V., Alden, D. L., Steenkamp, J-B.E.M., Ramachander, S. (2000). Effects of Brand Local and Nonlocal Origin on Consumer Attitudes in Developing Countries. *Journal of Consumer Psychology*, 9 (2), 83-95

[7] Beverland, M. (2006). The 'Real Thing': Branding Authenticity in the Luxury Wine Trade. *Journal of Business* Research, 59 (2), 251-258

[8] Bian, Q., Forsythe, S. (2012). Purchase Intention for Luxury Brands: A Cross Cultural Comparison. *Journal of Business Research, 65*, 1443-1451

[9] Bilkey, W., Nes, E. (1982). Country of Origin Effects on Product Evaluations. *Journal of International Business Studies*, 13, 89-99

[10] Cervino, J., Sanchez, J., Cubillo, J. M. (2005). Made in Effect, Competitive Marketing Strategy and Brand Performance: an Empirical Analysis for Spanish Brands. *Journal of American Academy of Business, 6 (2),* 237-243

[11] Christodoulides, G., Michaelidou, N., Li, C. H. (2009). Measuring Perceived Brand Luxury: An Evaluation of the BLI Scale. *Brand Management*, 16 (5/6), 395-405

[12] D'Astous, A., Ahmad, A. S. (1999). The Importance of the Country Images in the Formation of the 'Consumer Product Perceptions. *International Marketing Review*, *16 (2),* 108-125

[13] Dubois, B., Duquesne, P. (1993). The Market for Luxury Goods: Income versus Culture. *European Journal of Marketing, 27,* 35-44

[14] Eng, T-Y., Bogaert, J. (2010). Psychological and Cultural Insights into Consumption of Luxury Western Brands in India. *Journal of Customer Behaviour, 9 (1),* 55-75

[15] Erickson, G. M., Johansson, J. K., Chao, P. (1984). Image Variables in Multi-Attribute Product Evaluations: Country-of-Origin Effects. *Journal of Consumer Research*, 11, 694-669

[16] Fionda, A. M., Moore, C. M. (2009). The Anatomy of the Luxury Fashion Brand. *Brand Management*, *16 (5/6),* 347-363

[17] Han, C. M. (1989). Country Image: Halo or Summary Construct?. *Journal of Marketing Research*, *24,* 222-229

[18] Jackson, T. B. (2004). *International Retail Marketing*. Oxford: Elsevier Butterworth-Heinemann

[19] Khan, H., Bamber, D. (2008). Country of Origin Effects, Brand Image, and Social Status in an Emerging Market. *Human Factors and Ergonomics in Manufacturing, 18 (5),* 580-588

[20] Kinra, N. (2006). The Effect of Country-of-Origin in Foreign Brand Names in the Indian Market. *Marketing Intelligence & Planning*, *24 (1),* 15-30

[21] Koubaa, Y. (2008). Country of Origin, Brand Image Perception, and Brand Image Structure. *Asia Pacific Journal of Marketing and Logistics*, 20 (2), 139-155

[22] Mandel, N., Petrova, P. K., Cialdini, R. B. (2006). Images of Success and the Preference for Luxury Brands. *Journal of Consumer Psychology*, 16 (1), 57-69

[23] Melnyk, V., Klein, K., Völckner, F. (2012). The Double-eged Sword of Foreign Brand Names for Companies from Emerging Countries. *Journal of Marketing, 76*, 21-37

[24] Meng, J., Nasco, S. A., Clark, T. (2007). Measuring Country-of-Origin Effects in Caucasians, African-Americans and Chinese Consumers for Products and Services. *Journal of International Consumer Marketing*, 20 (2), 17-31

[25] Nagashima, A. (1970). A Comparison of Japanese and U.S. Attitudes toward Foreign Products. *Journal of Marketing, 34,* 68-74

[26] Nueno, J. L., Quelch, J. A. (1998). The Mass Marketing of Luxury. *Business Horizons*, *November/December,* 61-68

[27] Okonkwo, U. (2009). The Luxury Brand Strategy Challenge. *Journal of Brand Management, 16,* 287-289

[28] Ozretic-Dosen, D., Skare, V., Krupka, Z. (2007). Assessments of Country of Origin and Brand Cues in Evaluating a Croatian, Western and Eastern European Food Product. *Journal of Business Research*, 60 (2), 130-136

[29] Pappu, R., Quester, P. G., Cooksey, R. W. (2006). Consumer-based Brand Equity and Country-of-Origin Relationships: Some Empirical Evidence. *European Journal of Marketing, 40 (5/6),* 696-717

[30] Park, H-J., Rabolt, N. J., Jeon, K. S. (2008). Purchasing Global Luxury Brands among Young Korean Consumers. *Journal of Fashion Marketing and Management, 12,* 244-259

[31] Pecotich, A., Ward, S. (2007). Global Branding, Country of Origin and Expertise: An Experimental Evaluation. *International Marketing Review*, *24 (3),* 271-296

[32] Piron, F. (2000). Consumers' Perceptions of the Country-of-Origin Effect on Purchasing Intentions of (In)Conspicuous Products. *Journal of Consumer Marketing*, *17 (4),* 308-325

[33] Salciuviene, L., Ghauri, P. N., Streder, R. S., de Mattos, C. (2010). Do Brand Names in a Foreign Language Lead to Different Brand Perceptions?. *Journal of International Marketing, 26 (11/12)*, 1037-1056

[34] Seringhaus, F. H. R. (2005). Comparison of Website Usage of French and Italian Luxury Brands. *Journal of Euromarketing*, *14 (4),* 5-34

[35] Shukla, P. (2012). The Influence of Value Perceptions on Luxury Purchase Intentions in Developed and Emerging Markets. *International Marketing Review, 29 (6)*, 574-596

[36] Snaiderbaur, S. (2009). Made in Italy'' in China: From Country of Origin to Country Concept Branding. *Journal of Brand Management*, 4 (3/4), 63-74

[37] Thakor, M. V., Kohli, C. S . (1996). Brand Origin: Conceptualization and Review. *Journal of Consumer Marketing, 13 (3),* 27-42

[38] Thakor, M. V., Lavack, A. N. (2003). Effect of Perceived Origin Associations on Consumer Perceptions of Quality. *The Journal of Product and Brand Management*, 12 (6/7), 394-407

[39] Truong, Y., Simmons, G., McColl, R., Kitchen, P. J. (2008). Status and Conspicuousness - Are They Related? Strategic Marketing Implications for Luxury Brands. *Journal of Strategic Marketing*, 16 (3), 189-203

The Summarized Weighted Mean of Maxima Defuzzification and Its Application at the End of the Risk Assessment Process

Tamás Portik[1] and László Pokorádi[2]

[1] University of Debrecen, Faculty of Engineering, Ótemető utca 2-4, Debrecen, Hungary, portik@eng.unideb.hu

[2] Óbuda University, Donát Bánki Faculty of Mechanical and Safety Engineering, Népszinház u. 8, Budapest Hungary, e-mail: pokoradi.laszlo@bgk.uni-obuda.hu

Abstract: Nowadays, it is indispensible to take into consider the risk in modern technical management especially in hydraulic systems under different circumstances. The reliability of a hydraulic system is a well investigated area by researchers because of system damages or crash down can cause cost losses, human injuries or death. If the hydraulic system gives back imprecise or vague data and the reports of expert given by linguistic variables during the inspection then the risk assessment must be calculated with fuzzy mathematics. In this paper, the authors propose a modified fuzzy rule based risk assessment method for the risk assessment of hydraulic systems. The difference between the proposed and the original methods is the defuzzification sub process only. This defuzzification sub process is called Summarized Defuzzification (SDF).

Keywords: hydraulic system, fuzzy risk assessment, Summarized Defuzzification

1 Introduction

Pokorádi had written a paper about fuzzy rule based hydraulic risk assessments [13] in which some questions are raised. Before fuzzyfying, the averaging of opinions of experts does not reflect the full spectrum of expert opinions. Therefore authors work out a modified fuzzy rule based risk assessment method to handle extreme opinions on input data because of statistical averaging skips them immediately. The motivation was to handle these extreme opinions in another way because they are opinions from experts so they contain information from the investigated area even if experts do not have longtime experiences. On the other hand if one does not have enough data, moreover they are imprecise and uncertainty, from questionnaires then the statistical averaging is not the best way to handle extreme opinions. The expert opinions, knowledge are mainly reported in linguistic variables.

The first publication is written by Zadeh in [20], where he introduced the notion of fuzzy sets. To have a good introduction into fuzzy set and logic theory, the authors recommend the Ross' book [15] specialized for engineers. To measure imprecise and uncertainty data and notions one must use fuzzy set and logic theory which can handle the linguistic variables and logically inconsistent statements [15].

Fuzzy risk matrix is a very useful tool for semi-quantitative risk assessment which is provided by Markowski & Mannan to handle process activities of hazardous materials [7]. Therefore they developed low-cost, standard and high-cost matrixes.

In Markowski et. al.' paper, they introduced the risk correction index to take into consideration uncertainty concerned with the identification of representative accident scenario which provided more realistic results [8].

Shi et. al.'s analyzed results interpret the effectiveness of the Gaussian-mixed fuzzy clustering model on valence-arousal-related fMRI data-set for feature exracting operations including power spectrum density, spline, shape-preserving and cubic fitting methods [16].

The authors cite another application of fuzzy set and logic theory like Johanyák shows the clonal selection algorithm for tuning up the fuzzy inference system which was tested in case of SISO and MISO systems for which it is successfully usable [4].

Nagy et. al. proposed a stochastic approach for fuzzy control which is extremely fast, robust and simulation of the inverse pendulum makes validate it [9].

Rezaei et. al. proposed analytic hierarchy process with application for fuzzy multi-criteria decision making analysis which takes into consideration more important factors in decision making if it has fuzzy enviroment [14].

Singh & Markaset' paper represents a risk based inspection planning for oil and gas pipe systems based on fuzzy framework [17]. The inspected rate of corrosion and the efficiency of inspection are taken into consideration as fuzzy variables to calculate Trust in Inspection Results and Trust in Predicted Results which are combined together for estimated corrosion for carbon steel pipes.

Liu et. al. have written a comprehensive literature review about computing with words using fuzzy mathematics [6]. They survey the papers for ordinal linguistic approach, fuzzy rule based, fuzzy number and fuzzy extension of typical probabilistic risk assessment and some miscellaneous applications.

Cai represents a literature review about failure-oriented view to holistic view of system failure engineering combined with fuzzy methodology [1]. In his paper, there are shown a lot of engineering case-studies and fuzzy mathematics problems raised by applications.

Fuzzy risk analysis problems are proposed to solve with arithmetic operator – which satisfy all properties for trapezoidal fuzzy numbers according to Xu et. al.

[18] – and investigated the similarity of the trapezoidal fuzzy numbers. Moreover they made an experiment with 30 sets of fuzzy trapezoidal number for calculation risk which results are acceptable because they are close to human thinking.

Accuracy of agricultural natural disaster risk assessment is improved, which work efficiency is about 28% higher then the histogram method, with developed information distribution and interior-outer-set model for calculating fuzzy probability distribution to show imprecision by Huang in [2].

Jamshidi et. al. designed the Mamdani algorithm for pipeline risk assessment based on fuzzy logic toolbox of MATLAB to use one of the most popular techniques in pipeline risk assessment namely relative risk score methodology [3]. Moreover, they made a comparison between the traditional risk assessment and the fuzzy methodology based risk assessment on a case-study in which the last one provides more accurate and precise results.

Li et. al. proposed the fuzzy human error risk assessment for determining human error risk importance as function of human error probability, error-effect probability and error consequence severity and they provide the Mamdani techniques in fuzzy toolbox of MATLAB [5]. The results of case-studies are more realistic, practicable and valuable.

Wang et. al. proposed fuzzy failure mode and effect analysis with weighted geometric mean for prioritization of failure modes by fuzzy risk priority numbers [19]. The alpha-level sets centroid defuzzification method is used for ranking the failure modes.

Pokorádi wrote about the reliability and fuzzy rule based risk assessment with an application in a special helicopter mission for reliability and risk in [10]. Pokorádi’s book [12] speaks about system and process modeling in engineering which contains introduction in fuzzy set and logic theory and its applications specialized for engineers like fuzzy decision-making and fuzzy failure mode and effect analysis. Pokorádi has applied the fuzzy rule based risk assessment in military science in [10].

All of cited papers above show how any kind of fuzzy risk assessment is important on a lot of part of science like engineering, informatics, agriculture and so on.

The general aim of the authors is to work out new fuzzy rule based risk assesment methods which consists new defuzzification where the opinion of experts goes through the process till the composition part during the decision-making without averaging the experts’ opinions at the beginning of the process. During suggested method averaging occurs at the end of the assessment process. Therefore the total spectrum of experts’ opinions has more effect on results. Because of different intermediate results arise in inference and composition subprocesses in case of different experts’ opinions.

This means more carefully quantified experts' opinion as result which reflect more carefully the experts' opinions. The aim of this paper is to present a solution to handle extreme opinions without averaging at the beginning of fuzzy rule based risk assessment. This assessment process uses called Summarized DeFuzzification (SDF).

The rest part of the paper is organized as follows: Section 2 provides a short overview about fuzzy risk assessment and introduces the theoretical background of the SDF method. Section 3 shows an easy case-study about a hydraulic system for its risk computation. In the last section, authors give a summary of their work.

2 Fuzzy Rule-based Risk Assessment

In this section authors proved a short overview about traditional fuzzy rule-based risk assessment and they introduce the summarized defuzzification.

2.1. Traditional Fuzzy Rule-based Risk Assessment

The traditional fuzzy decision process has next parts: fuzzyfication, inference, composition and defuzzification (Figure 1). Sometimes it is joined together the composition and defuzzification process.

Fuzzyfication sub process means that the input data are fuzzyfied with predefined fuzzy membership functions for further computations.

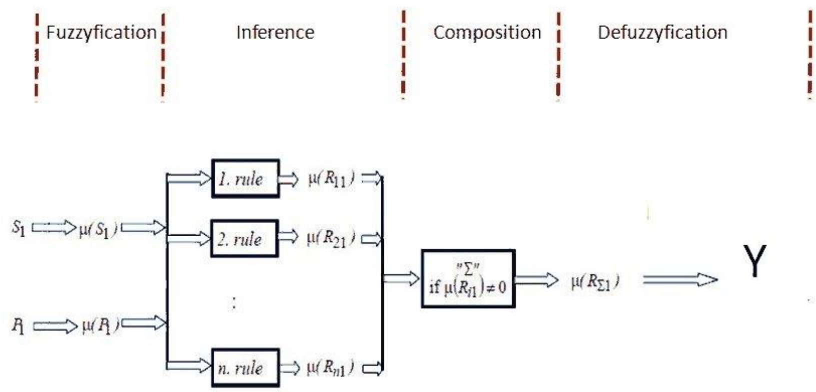


Figure 1

Traditional fuzzy process flow-chart (source: [12])

Inference sub process means if the predicates are joined together with logical AND then the data are combined together with t-norm. If the predicates are connected with logical OR then the data are handled with s-norm. The If-Then rules are represented in the Risk Assessment Matrix (RAM) for easy handling (like Figure 10).

Composition sub process means each category which has more values, is composed with s-norm – usually used maximum operator. One takes only the non zero inference into consider in composition sub process.

Defuzzification sub process means that the fuzzy result is converted in a crisp value. This is necessary especially in technical management because decision makers like managers, leaders can not handle the fuzzy results.

There is a lot of theoretical and practical method for defuzzification process. The authors choose the Weighted Mean of Maximum (WMM) – because of the "base" of case study [13] used that method – where of course it has many different types. In this paper it is defined by equation (1).

$$R_{WMM} = \frac{\sum_{i=1}^{n} \mu_i \cdot z_i}{\sum_{i=1}^{n} \mu_i} \tag{1}$$

n – the number of conclusions which are different from 0,

μ_i – the value of i-th fuzzy membership function,

z_i – the weighted value of the i-th fuzzy membership function.

The weighted value of fuzzy membership function means either the point on abscissa where fuzzy membership function takes its maximum value at only one point or the midpoint of the interval if this maximum value is on an interval. This method is common used in literatures i.e. [12, 15].

2.2. Summarized Defuzzification (SDF)

The fuzzy rule based risk assessment using Summarized Defuzzification (SDF) method is the next: one takes the traditional fuzzy process till the composition part, in other words one let the input opinions, where some of expert teams can have extreme opinion, through fuzzyfication, inference and composition sub process.

One collects all this fuzzy results. This means one has converted all extreme opinions to fuzzy risk assessment. Then one can defuzzyfy for all fuzzy risk opinions, if crisp value of risk is needed, where the SDF is shown on Figure 2. To get it clear, the definition of summarized defuzzification is in equation (2).

$$R_{SWMM} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{ij}\cdot z_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{ij}} \tag{2}$$

R_{SWMM} - Crisp value of summarized defuzzification,

m – the number of opinions (input data),

n – the number of fuzzy membership function,

μ_{ij} – the value of j-th fuzzy membership function which belongs to i-th opinion (input data),

z_{ij} – the weighted value of the i-th fuzzy membership function in case of the j-the opinion.

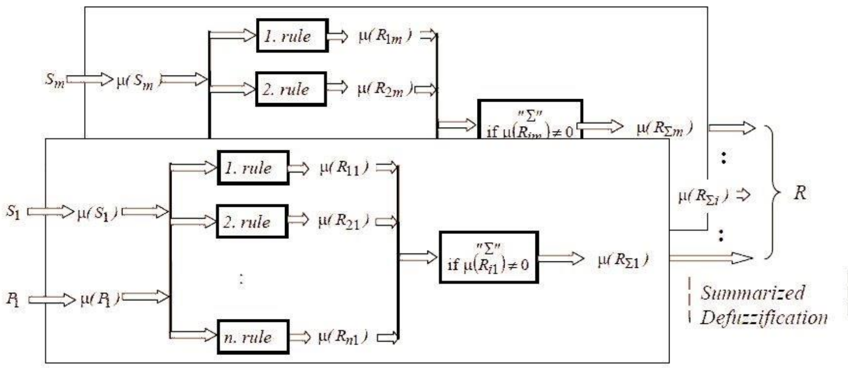


Figure 2
Summarized defuzzification flow-chart

One remark again: The weighted value of fuzzy membership function has either only one point where fuzzy membership function takes its maximum value or the midpoint of the interval if this maximum value is over an interval. This method can easily and quickly to be computed. The equation 2. express that the experts' opinion are taken into considaration at the end of defuzzification process. In this paper all computation is made by Maple in which a module is developed by authors.

3 Case-Study

In this case-study the authors investigate an easy sample for a hydraulic system. This sample was investigated in an earlier paper with the traditional fuzzy rule based method which is written in [13] when Pokorádi used opinions of two expert teams that investigated a hydraulic system from different points of view. During the calculation of case-study, the authors use triangular and trapezoidal fuzzy membership functions on linear (equations (3)-(4) and Figures 3-4) and logarithmic scaled abscissa (equations (5)-(6) and Figures 5-6).

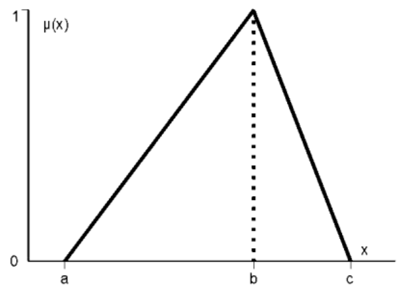


Figure 3
Triangular fuzzy membership function

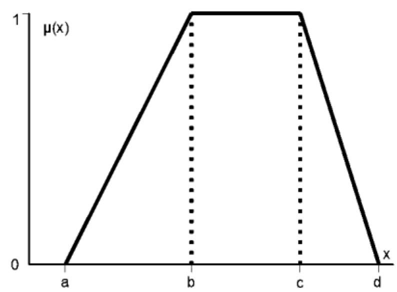


Figure 4
Trapezoidal fuzzy membership function

$$f(x)=\begin{cases}\dfrac{x-a}{b-a} & , \quad a<x\le b\\ \dfrac{c-x}{c-b} & , \quad b<x\le c\\ 0 & , \quad otherwise\end{cases} \tag{3}$$

$$Sign: f:(a,b,c).$$

$$f(x)=\begin{cases}\dfrac{x-a}{b-a} & , \quad a<x\le b\\ 1 & , \quad b<x\le c\\ \dfrac{d-x}{d-c} & , \quad c<x\le d\\ 0 & , \quad otherwise\end{cases} \tag{4}$$

$Sign:\ f:(a,b,c,d)$

$$\varphi(x)=\begin{cases}\dfrac{\lg(x)-\lg(a)}{\lg(b)-\lg(a)} & , \quad a<x\le b\\ \dfrac{\lg(c)-\lg(x)}{\lg(c)-\lg(b)} & , \quad b<x\le c\\ 0 & , \quad otherwise\end{cases} \tag{5}$$

$Sign:\varphi:(a,b,c).$

Firstly the severity, probability, and risk fuzzy sets categories (Table 1, Figures 7-9) and the RAM (Figure 10) should be defined according to experts and their experiences and knowledge.

$$\varphi(x)=\begin{cases}\dfrac{\lg(x)-\lg(a)}{\lg(b)-\lg(a)} & , \quad a<x\le b\\ 1 & , \quad b<x\le c\\ \dfrac{\lg(d)-\lg(x)}{\lg(d)-\lg(c)} & , \quad c<x\le d\\ 0 & , \quad otherwise\end{cases} \tag{6}$$

$Sign:\ \varphi:(a,b,c,d)$

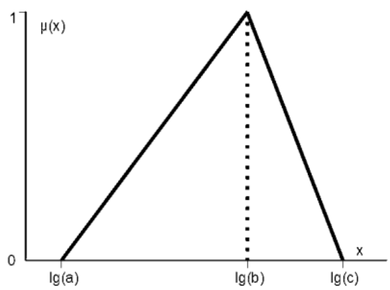


Figure 5

Triangular membership function in case of logarithmic scaled abscissa

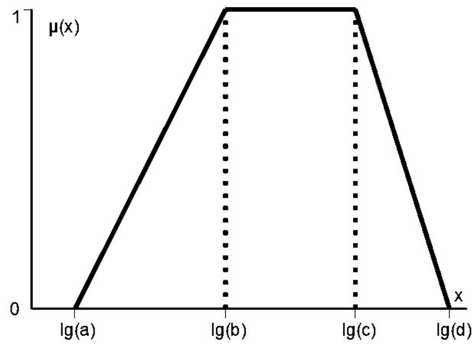


Figure 6

Trapezoidal membership function in case of logarithmic scaled abscissa

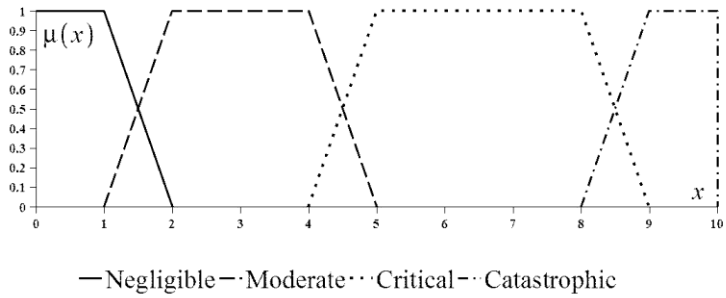


Figure 7

Fuzzy membership function of severity categories

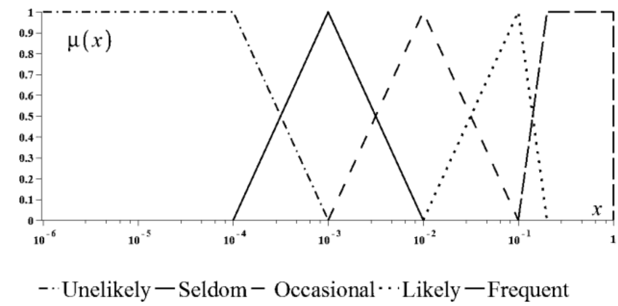


Figure 8

Fuzzy membership function of probability categories

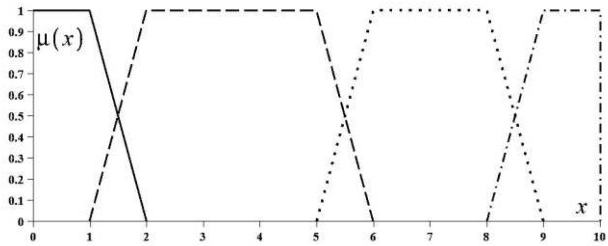


Figure 9
Fuzzy membership function of risk categories

Table 1
The severity-, probability- and risk categories

Severity	Negligible	$\mu_{Neg} = f:(0,0,1,2)$
	Moderate	$\mu_{Mod} = f:(1,2,4,5)$
	Critical	$\mu_{Crit} = f:(4,5,8,9)$
	Catastrophic	$\mu_{Cat} = f:(8,9,10,10)$
Probability	Unlikely	$\mu_{Unl} = \varphi:(0,0,10^{-4},10^{-3})$
	Seldom	$\mu_{Sel} = \varphi:(10^{-4},10^{-3},10^{-2})$
	Occasional	$\mu_{Occ} = \varphi:(10^{-3},10^{-2},10^{-1})$
	Likely	$\mu_{Lik} = \varphi:(10^{-2},10^{-1},0.2)$
	Frequent	$\mu_{Freq} = \varphi:(10^{-1},0.2,1,1)$
Risk	Low	$\mu_{Low} = f:(0,0,1,2)$
	Medium	$\mu_{Med} = f:(1,2,5,6)$
	High	$\mu_{H} = f:(5,6,8,9)$
	Extra High	$\mu_{EH} = f:(8,9,10,10)$

Two types of hazard are investigated in our hydraulic system namely burst in return pipe and pump failure (Table 2). The results of traditional fuzzy process are in “traditional” columns. The data are gathered from investigated failure rate with statistical methods for probability and the scale is from 0 to 1 for it. The experts/engineers have been divided into two expert groups. Let them A and B. The opinion about severity, is given by team A, are 3.8 for burst return pipe and 8 for pump failure on scale from 0 to 10. The opinion about severity, is given by team B, are 4.6 for burst return pipe and 9 for pump failure again on scale from 0 to 10. The averages are 4.2 for burst in return pipe and 8.5 for pump failure. The results of average input mean the traditional result which is published in [13].

	Frequent	**Likely**	**Occasional**	**Seldom**	**Unlikely**
Catastrophic	EH		H		**M**
Critical				**M**	
Moderate	**H**	**M**		**L**	
Negligible	**M**				

EH — Extra High; **H** — High; **M** — Medium; **L** — Low.

Figure 10

Risk Assessment Matrix (RAM) (source: [13])

The authors emphasize not to do averaging because of all expert has notable knowledge about the investigated hydraulic system even if some of them have not too much experiences.

Table 2

The results of traditional and summarized defuzzification

	Name of Hazard	Burst in return pipe			Pump failure		
		Traditional	A	B	Traditional	A	B
Input data	Severity	4.2	3.8	4.6	8.5	8	9
	Probability	0.0002			0.005		
Severity	Catastrophic	0	0	0	0.5	0	1
	Critical	0.2	0	0.6	0.5	1	0
	Moderate	0.8	1	0.4	0	0	0
	Negligible	0	0	0	0	0	0
Probability	Frequent	0			0		
	Likely	0			0		
	Occasional	0			0.699		
	Seldom	0.301			0.301		
	Unlikely	0.699			0		
Inference & composition	Extra high	0	0	0	00	0	0
	High	0	0	0	0.5	0.699	0.699
	Medium	0.2	0	0.301	0.301	0.301	0
	Low	0.699	0.699	0.6	0	0	0
Defuzzification		1.284	1.205		5.685	6.38	

The first step is fuzzyfying input data, this step can be seen for probability and severity on Table 2. The severity and probability values are joined together with minimum operator according to risk assessment matrix (e.g. Table 3 for B group at burst in return pipe). If some risk categories have many values then they are joined together with maximum operator in composition part (e.g. Table 2) which is represented at Table 3. as well.

The fuzzy result is L – 0.699 after the calculation for opinion A at burst in return pipe. The fuzzy result is M — 0.301 and L — 0.6 after the calculation for opnion B at burst in return pipe. The fuzzy result is L – 0.699 after the calculation for opinion A at burst in return pipe. The fuzzy result is M — 0.301 and L — 0.6 after the calculation for opnion B at burst in return pipe.

Table 3

The RAM & composition result

<table>
<tr><td colspan="2"></td><td>Frequent
0</td><td>Likely
0</td><td>Occasional
0</td><td colspan="2">Seldom
0.301</td><td colspan="2">Unlikely
0.699</td></tr>
<tr><td rowspan="2">Catastrophic</td><td>A, 0</td><td rowspan="2">EH, 0</td><td rowspan="2">EH, 0</td><td rowspan="2">H, 0</td><td colspan="2" rowspan="2">H, 0</td><td colspan="2" rowspan="2">M, 0</td></tr>
<tr><td>B, 0</td></tr>
<tr><td rowspan="2">Critical</td><td>A, 0</td><td rowspan="2">EH, 0</td><td rowspan="2">H, 0</td><td rowspan="2">H, 0</td><td rowspan="2">M</td><td>0</td><td rowspan="2">L</td><td>0</td></tr>
<tr><td>B, 0.6</td><td>0.301</td><td>0.6</td></tr>
<tr><td rowspan="2">Moderate</td><td>A, 1</td><td rowspan="2">H, 0</td><td rowspan="2">M, 0</td><td rowspan="2">M, 0</td><td rowspan="2">L</td><td>0.301</td><td rowspan="2">L</td><td>0.699</td></tr>
<tr><td>B, 0.4</td><td>0.301</td><td>0.4</td></tr>
<tr><td rowspan="2">Negligable</td><td>A, 0</td><td rowspan="2">M, 0</td><td rowspan="2">L, 0</td><td rowspan="2">L, 0</td><td colspan="2" rowspan="2">L, 0</td><td colspan="2" rowspan="2">L, 0</td></tr>
<tr><td>B, 0</td></tr>
<tr><td colspan="9">After composition A opinion about risk for burst in return pipe: L — 0.699
After composition B opinion about risk for burst in return pipe: M — 0.301, L — 0.699</td></tr>
</table>

To do SDF one must use equation (2). The number of opinions n is 2, the number of fuzzy membership function at risk categories are m = 4. So using the equation (2) one gets the result of the fuzzy opinion of experts for crisp value of risk 1.205 on a scale from 0 to 10 by burst in return pipe. At pump failure one gets the result for crisp value of risk 6.38. Both of results are different with SDF from the traditional fuzzy based computation.

After using of modified risk assessment, the Authors showed its complete results (not only these hazards shown above) to participators. Their general opinions were, the modified results are more carefully quantified of joint expertise according to the rate of experts' group. But the participators could not estimate the difference by a numerical crisp value.

Summary

In this paper, authors have proposed a new point of view for handling extreme experts' opinions during fuzzy rule based risk assessment. This method uses summarized defuzzification which gives better crisp values for risk in risk assessment like traditional method according to Pokorádi. Of course like every risk predictions which are used in decision-making process, have been provided responsibilities of managers or leaders. The future aims to work out other SDF methods and to use them in general practice for technical management.

References

[1] K.-Y. Cai, System Failure Engineering and Fuzzy Methodology An Introductory Overview, Fuzzy Sets and Systems, Vol. 83, Issue 2, 8 October 1996, p. 113-133

[2] C. Huang, An Application of Calculated Fuzzy Risk, Information Sciences, Vol. 142, Issues 1-4, May 2002, pp. 37-56

[3] A. Jamshidi, A. Yazdani-Chamzini, S. H. Yakhchali, S. Khaleghi, Developing a New Fuzzy Inference System for Pipeline Risk Assessment, Journal of Loss Prevention in the Process Industries, Vol. 26, Issue 1, January 2013, pp. 197-208

[4] Zs. Cs. Johanyák, Clonal Selection-based Parameter Optimization for Sparse Fuzzy Systems, in Proc. of IEEE 16th International Conference on Intelligent Engineering Systems (INES 2012), June 13-15, 2012, Lisbon, Portugal, pp. 369-373

[5] P.-C. Li, G.-H. Chen, L.-C. Dai, Z. Li, Fuzzy Logic-based Approach for Identifying the Risk Importance of Human Error, Safety Science, Vol. 48, Issue 7, August 2010, pp. 902-913

[6] J. L. Liu, H. Martinez, R. M. Rodríguez, V. Novozhilov, Computing with Words in Risk Assessment, International Journal of Computational Intelligence Systems, Vol. 3, Issue 4. pp. 396-419 (2010)

[7] A. S. Markowski, M. S. Mannan, Fuzzy Risk Matrix, Journal of Hazardous Materials, Vol. 159, Issue 1, 15 November 2008, pp. 152-157

[8] A. S. Markowski, M. S. Mannan, A. Kotynia (Bigoszewska), D. Siuta, Uncertainty Aspects in Process Safety Analysis, Journal of Loss Prevention in the Process Industries, Vol. 23, Issue 3, May 2010, pp. 446-454

[9] K. Nagy, Sz. Divéki, P. Odry, M. Sokola,V. Vujičić, A Stochastic Approach to Fuzzy Control, Acta Polytechnica Hungarica, Vol. 9, No. 6, (2012) pp. 29-48

[10] L. Pokorádi, Fuzzy Logic-based Risk Assessment, AARMS, Academic and Applied Research in Military Science, Volume 1, Issue 1 (2002) pp. 63-73

[11] L. Pokorádi, Application of Fuzzy Set Theory for Risk Assessment, Journal

of KONBiN, No 2,3 (14, 15) 2010, Warsaw, pp. 195-204

[12] L. Pokorádi, Rendszerek és folyamatok modellezése, Campus Kiadó, Debrecen, Hungary (2008)

[13] L. Pokorádi, Risk Assessment based upon Fuzzy Set Theory, 15th "Building Services Mechanical and Industrial Days" International Conference, 15-16 October 2009, Debrecen, Hungary, pp. 311-318

[14] P. Rezaei, K. Rezaie, S. Nazari-Shirkouhi, M. R. J. Tajabadi, Application of Fuzzy Multi-Criteria Decision Making Analysis for Evaluating and Selecting the Best Location for Construction of Underground Dam, Acta Polytechnica Hungarica, Vol. 10, No. 7 (2013) pp. 187-205

[15] T. J. Ross, Fuzzy Logic with Engineering Applications, John Wiley & Sons Ltd., Singapore (2010)

[16] F. Shi, P. McC. Bush, A Gaussian-mixed Fuzzy Clustering Model on Valence-Arousal-related fMRI Data-Set, Acta Polytechnica Hungarica, Vol. 10, No. 8 (2013) pp. 85-104

[17] M. Singh, T. Markeset, A Methodology for Risk-based Inspection Planning of oil and Gas Pipes Based on Fuzzy Logic Framework, Engineering Failure Analysis, Vol. 16, Issue 7, October 2009, pp. 2098-2113

[18] Z. Xu, S. Shang, W. Qian, W. Shu, A Method for Fuzzy Risk Analysis Based on the New Similarity of Trapezoidal Fuzzy Numbers, Expert Systems with Applications, Vol. 37, Issue 3, 15 March 2010, pp. 1920-1927

[19] Y.-M. Wang, K.-S. Chin, G. K .K. Poon, J.-B. Yang, Risk Evaluation in Failure Mode and Effects Analysis using Fuzzy Weighted Geometric Mean, Expert Systems with Applications, Vol. 36, Issue 2, Part 1, March 2009, pp. 1195-1207

[20] L. Zadeh, Fuzzy Sets, Information and Control, Vol. 8 (1965) pp. 338-353

Cloud Computing Support to University Business Processes in External Collaboration

Imre Petkovics[1], Pere Tumbas[1], Predrag Matković[1], Zoltan Baracskai[2]

[1] University of Novi Sad, Faculty of Economics Subotica, Department of Business Informatics, Segedinski put 9-11, 24000 Subotica, Serbia
e-mail: peti@ef.uns.ac.rs, ptumbas@ef.uns.ac.rs, pedja@ef.uns.ac.rs

[2] Doctus Co. Csikihegyek út 8, 1118 Budapest, Hungary
e-mail: zoltan@doctus.hu

Abstract: Implementation of the business process management concept at universities has crucially contributed to abandoning traditional and introducing innovative processes in recent years. In particular, information technologies serve as not only means of automation but also a method of altering the approach to conducting business operations. Information technologies either facilitate or hinder the use faster, better and more cost-effective solutions for implementing business processes. This article analyses the possibility of supporting business processes in universities' external collaboration with cloud computing technologies, and their influence on the process' input-output data quality dimensions.

Keywords: cloud computing; business process; innovative university

1 Introduction

Information technologies (IT) have featured as a key factor in the implementation of the business process redesign (BPR) concept ever since its advent. They have been changing the nature of business processes, and this will continue in the future as well. The motives for the application of this radical concept in process redesigning in the organisation have been diverse, and the most common ones include the need to accelerate process implementation, reduce required resources, improve productivity and efficiency, and improve the organisation's competitiveness. Information technologies were originally used for accelerating and facilitating office work, rather than changing or transforming them [1]. However, BPR and IT have the potential to create more flexible, team oriented, coordinative and communication-based work capability [2]. Information technologies represent much more than a process automation toolkit. They enable full-scale reformation and redesign of a process. Researching the logical and natural relation between BPR and IT has not yet been fully completed [3]. Information technologies are the most effective enabling technologies for BPR, aiding the accomplishment of redesign objectives in three ways: by providing information across functional levels and establishing easy communication, improving the process performance, and helping reengineering effort by modelling, optimizing and assessing its consequences [1].

Over the past decade, business processes at universities have also been changing and acquiring characteristics in accordance with changes occurring in higher education systems. Faced with challenges and dilemmas such as: globalisation vs. national needs, government administration vs. institutional autonomy, harmonisation vs. diversity, public vs. private sectors, basic vs. applied research, and competition vs. cooperation [4], universities respond by selecting and changing their processes. Thus, in their development, they undergo transformations through various organisation models, the most common being the Entrepreneurial University or the Innovative University model.

An Innovative University is a university of the entrepreneurial type, which is characterized by scientific and educational activities based on innovative technologies and principles of management [5]. The creation and implementation of innovation is the main goal and main content of the innovative activities. The system of innovative activities includes not only the development and creation of innovations but also the management of the innovative processes and the provision of conditions for their successful implementation [5].

Innovation at universities implies a need for them to reconsider their role and position in at least three areas, and define their own place in the contemporary higher education system development trends. These include areas of innovative teaching and learning process, innovative research process and enabling processes or organisational innovation processes.

The key factors for becoming an innovative university are most commonly structured in four categories: regulations and policies, external relations, internal stakeholders and infrastructure [6]. In particular, in the infrastructure category, the following IT infrastructure elements should be highlighted:

- reliable, fast and free internet connection, including wireless service;
- current software packages and use of open source software;
- well equipped classrooms, labs and libraries with digital boards, video projectors, etc.;
- sufficient available computers for all students;
- access to communication networks, and
- access to databases, online libraries, and online academic journals.

The research in this article is guided by the following research question: What is the supporting capability of cloud computing technologies in conducting universities' business processes in terms of improving universities' external collaboration? Seeking an answer to this question implied elaborating the concept, main characteristics, models of cloud computing technology and their advantages and disadvantages. This was followed by identifying and analysing universities' business processes from the aspect of input-output quality dimensions. Particular attention was devoted to analysing the quality of data that are consumed and produced by the identified processes. At the end of this research, conclusions were drawn on the possibility of applying particular types of service models and types of cloud computing technologies in the conduct of the universities' external collaboration and support of various innovation processes.

The following research hypotheses were set in the paper:

H1. Cloud computing technology improves university business processes.

H2. Integrated (Hybrid) cloud computing is the best way to implement information technology.

H3. Cloud computing technology contributes to the development of the Triple Helix concept of innovation through external collaboration.

H4. Applying cloud computing technology in performing universities' collaborative processes influences their input-output quality dimensions.

2 Research Background

2.1 Cloud Computing – Definitions, Features, and Models

Cloud computing concept broadly refers to using computer resources (hardware, software, platform, database servers, application, software packages etc.) over a network (the Internet) and for a vast majority of those involved in service-level understanding of the processing and payment [7].

The originator of the idea of cloud computing is John McCharty. He does not actually use the term "cloud computing" by concept formulation, but as computing utility, to denote the new paradigm. He argued that, in the future, computer time-sharing technology might result in the fact that computing power and even specific applications could be sold through the utility business model [8].

The emergence of the Internet in 1969 relaunched the idea of computation services: "As of now, computer networks are still in their infancy. But as they grow up and become more sophisticated, we will probably see the spread of 'computer utilities' which, like present electric and telephone utilities, will service individual homes and offices across the country" [9][10]. We had to wait for almost fifty years to see his ideas realised.

The word "cloud" itself was borrowed from the telephone companies, where the term "cloud" is applied to a situation where the contracted bandwidth is not specific wire pairs to deliver, but free phone lines activated at the time of the transfer to ensure guaranteed bandwidth. The word "cloud" is then applied to the state when it is still impossible to pinpoint the specific transmission lines.

The definition of cloud computing by the National Institute of Standards and Technology (NIST) reads as follows: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [11]. This definition primarily reflects the provider's point of view of cloud computing. The consumer side of cloud computing describes this phenomenon as a new cost-efficient computing paradigm in which applications, platforms, networks and computer power can be accessed from a Web browser by consumers without any information or care about the location of these resources, and all these resources/services are available on a pay-per-use basis. An academic definition of cloud computing was first given by Armbrust et al [12].

According to the NIST, cloud computing have five fundamental characteristics, involve three service models and four deployment models. The five main characteristics of cloud computing benefits are: (a) on-demand self-service, (b) ubiquitous, broad network access, (c) location-independent resource pooling, (d) rapid elasticity and (e) measured pay-per-use service.

Apart from the fundamental ones, there are some other significant features of cloud computing [7], [20]:

- Multitenancy – a single instance of an application serves multiple consumers (tenants) with the possibility (for the tenants) to customize some details (GUI or business rules) of the software applications;
- Identity management and access control – application of advanced techniques in identity checks and access control in order to provide legitimate data use;
- Security of services – implementation of the most stringent cutting-edge security measures for service provision;
- Increased reliability, because errors are constantly removed and repaired;
- Centralization of infrastructure and lower expenses of operation and maintenance;
- Diverse platform support, where cloud providers often provide services with built-in support for an abundant compilation of client platforms to reach a broader base of users;
- Faster development - cloud computing platforms with a large number of services and preset templates for application development considerably facilitate the development process.

NIST differentiates between three service models:

1. *Infrastructure as a Service (IaaS*). IaaS is the lowest layer of cloud computing. At this level, the subject of service includes hardware resources (processor, memory, bandwidth and HDD storage) leased to the consumer. By means of leasing, consumers create the opportunity for installing and using virtual machines and various applications used in their business operations. The key advantages of IaaS for the consumer are:
   - scalability - IaaS enables the consumer to simply lease additional resources for newly arisen needs on the pay-per-use principle; and
   - portability – IaaS enables consumers to transfer their own virtual machines to some other hardware resources.

   Iaas providers often mention the following advantages in their marketing activities as well: freedom from limits, power and control, predictability, expert guidance, cost savings, system acquisition bypass, enhanced agility and innovative business ricing.

2. *Platform as a Service (PaaS).* The NIST describes PaaS as "the capability provided to the consumer to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider." PaaS providers offer the consumers a development platform as a service, which they can use for their own software solution and deploy it on cloud infrastructure. In addition to the obligation to maintain hardware and replicate data, PaaS provider is responsible for managing operative systems, network resources, servers and storage, whereas the PaaS consumer is obliged to manage the deployed application and customise the application-hosting environment. Based on the above, it is obvious that PaaS is a higher layer of cloud computing in comparison with the IaaS. The advantages of PaaS for the consumer are:
   - a development platform prepared for fine tuning and use – PaaS offers consumers platforms on which they can quickly develop and deploy their own software solutions; and
   - consumers do not have to maintain hardware and network infrastructure; it is the PaaS that performs these tasks for consumers, so that the latter only need to perform their core activity – developing software solutions.

   In addition to these key advantages, PaaS providers also point out reduced costs, faster time to market, lower development failure risk and increased security. Of course such a mode of cloud computing has its disadvantages as well, one of the most significant being the fact that PaaS providers demand that consumers sign long-term PaaS usage agreements.

3. *Software as a Service (SaaS).* SaaS is provision of software by providers to consumers on a pay-per-use service basis. The consumer uses the services for business purposes, through the Internet as a communication medium. While PaaS providers offer some kind of development platform, and IaaS provides infrastructure, SaaS offers any software solution uploaded to the cloud. SaaS

providers may act as both IaaS or PaaS tenants and SaaS providers. By leasing SaaS, the consumer gets a complete and functional solution, with the only obligation to customize and use it, whereas the provider undertakes all other obligations. Such a way of meeting consumers' needs for software solutions frees the consumer from the obligation to buy software licenses and hardware required for the purchased licenses to function. The user benefits of SaaS are:

- pay-per-use – leasing the SaaS, the consumer pays the lease cost for a given time period, with a constantly open possibility to change the SaaS package;
- infrastructure maintenance and software updates – when leasing the SaaS, the consumer undertakes to configure and use the rented SaaS, and the provider undertakes all the obligations inherent to the lower IaaS Cloud layer, and the obligation to maintain and update the rented SaaS.

In addition to these key advantages, the SaaS providers also point out easier administration, compatibility, easier collaboration and global accessibility. The key disadvantage of SaaS is data lock-in, a problem most often manifested when attempting to transfer the service to another SaaS provider. The problem of data, i.e. their transferability, often emerges in such cases.

According to many sources [13], [14], [15] and [16] the list of the above services is not complete. There are many classifications of cloud computing services along different layers. Youseff et al. [16] were the first to suggest the presentation of cloud computing as a layered model with associated services. The new services on this model are:

1. *Data Storage as a Service (DaaS)*. This is on-demand data storage that allows consumer to store their data via the Internet on remote disks and access them anytime from any place with any Internet accessible devices.
2. *Communication as a Service (CaaS)*. This service provides communication encryption, network monitoring, dedicated bandwidth, guaranteed message delay, virtual overlays for traffic isolation, etc. (VoIP, audio conferencing and video conferencing) [17]
3. *Hardware as a Service (HaaS)*. This is an on-demand data centre that includes the actual physical hardware and switches that form the backbone of the cloud by means of clusterisation.

From the aspect of deployment of Cloud Computing there are three (not exclusive) classifications, which are based on: (a) publicity of usage, (b) locality of realization and (c) organization of maintenance [17].

From the aspect of publicity of usage, the deployment models are:

- *Private cloud*: The cloud infrastructure is dedicated for exclusive use to a single organization with a large number of workers/end users.
- *Community cloud*: The type of cloud service that provides for a specific group of persons or organizations with common interest.

- *Public cloud*: The type of cloud deployment that provides offer to any consumer for open use over the public Internet.
- *Hybrid cloud*: This design of cloud deployments present the mix or composition of two or three types of cloud deployment models (private, community, or public).

Some sources mention new deployment model forms, such as consumer cloud (like Facebook), although it is a typical form of public cloud, or government cloud (in the USA and South Korea), to emphasise the higher level of quality and innovation in retail service [18].

From the point of view of locality of implementation, the deployment models are [19]:

- *Internal cloud*, where services are offered by an internal IT organization (in-house implementation) for internal usage (usually) (in-house computer clouds, physically located within organisation premises and protected by the organisation's firewall), where access is provided through the intranet, and the clouds are for internal use only;
- *External cloud*, where cloud computing services are not hosted by an internal IT organization (external implementation). An external cloud can be public or private, but must be implemented by an external IT organization (external computer clouds organised outside the user organised organisation by the cloud provider within its ecosystem).

From the aspect of organisation of maintenance, the deployment models are:

- in-house maintenance of cloud computing services, and
- outsourced maintenance of cloud computing services.

The emergence of cloud computing, of course, would have been impossible without the discovery of the Internet and a whole range of innovation and other technical and organisational novelties in the area of hardware miniaturisation, with a drastic decrease of energy consumption for work. On the other hand, the financial crisis lasting more than five years already gives rise to and dictates huge cost-cutting campaigns (i.e. severe cost limitations) in corporate functioning. The companies, therefore, do not invest major funds in IT nowadays (apart from PCs, laptops, notebooks or smart phones), but rather, when they need a more intense data processing environment, they opt for applying cloud computing rather than purchasing.

2.2 Business Process Redesign – Concept and Process Quality

BPR is a management technique applied for transforming organizations radically in order to improve them radically [22], [23]. According to most authors, the critical role in change initialisation is played by information technologies. As stated above, the IT infrastructure helps organisations not only automate business process activities, but also reshape and redesign business processes [24].

The primary features of BPR include: complete process redesign by means of radical change, with high risk and long implementation period where collaboration between business and IT is optional. The BPR process is cost-intensive, and the outcome is a drastic change in one or several business processes.

In order to meet consumer demands, companies have to design business processes in an appropriate way [21]. In particular, four essential process competencies feature as a basis for consideration, i.e. discussion: process cost, process flow time, process flexibility, and process quality [25]. It is essential to point out that the quality dimension of a process is often neglected in business process modelling and design. Normally, process quality has two groups of dimensions: the function quality dimensions and input-output quality dimensions.

A function is a basic building block in a business process that corresponds to an activity (task, process step) which needs to be executed [22]. Heravizadeh et al. identified thirteen quality dimensions, which can be relevant for a function, but not all are applicable for every function. These are: suitability, accuracy, security, reliability, understandability, learnability, time efficiency, resource utilisation, effectiveness, productivity, safety, user satisfaction and robustness [21]. Using these dimensions, analysts can identify quality aspects of particular functions of an individual process very precisely.

Table 1

Input output data quality dimensions

Input-output quality dimensions	Description
Accessibility	Data accessibility refers to the extent to which data is easily accessible (obtainable) when needed.
Amount of data	Amount of data relates to a sufficient data volume for our needs.
Completeness	Completeness is the extent to which the data includes all necessary values.
Comprehensiveness	Data comprehensiveness - means that all the required data elements are included or that the record is complete.
Consistency	Data consistency - means that the data are reliable. Data consistency is the synchronization of data objects across the company.
Currency	Data currency/data timeliness - means that data should be up-to-date and recorded at or near the time of the event or observation.
Reliability	Data should reflect stable and consistent data collection processes across collection points and over time.
Security	Data security is the extent to which data is protected against unauthorized access.
Timeliness	Data timeliness refers to the extent to which the data is sufficiently current.
Transparency	Data transparency is the ability to trace back data to its origin and find out its real world meaning.

The input and output of functions within a business process include physical and informational objects that are consumed and produced by it. Given the research subject, individual attention was devoted solely to data and information quality.

Two categories of data quality are identified in [26] - data product quality and data service quality. Data product quality includes the quality aspects related to data itself, whereas service quality includes aspects related to the service delivery process of the information to consumers. Data quality attributes identified in [27], and modified in [21] make the focus of this research. A detailed analysis shortlisted ten out of twenty quality dimensions in correlation with cloud computing technology. A list of those and a brief description is given in Table 1. The chosen data quality dimensions were used for assessing the eligibility of universities' collaborative business processes for the implementation of cloud computing technology.

3 Research Methodology

The basis for the analysis of university business processes was the „Project Task of Implementation of a Unified Information System of the University of Novi Sad", completed within the Tempus project „Governance and Management Reform in Higher Education in Serbia" (GOMES) in 2009. The project included identification of business processes of Serbian universities, which were then redesigned based on the experiences of partner universities from the EU (Italy, Greece, Poland, and Czech Republic) in business process implementation.

In view of the research question formulated in the article, the key activity in the remaining part of the research is gathering and analysing reference literature, which provides the overview of existing knowledge on the subject issue. For this reason, a comprehensive study of literature was carried out, followed by a detailed and rigorous verification of its validity and reliability. To provide transparency and repeatability of the entire procedure, the authors made a meticulous selection of academic databases, types of publication, key words and time interval.

The following academic bases were chosen for the purpose of this research: EBSCO Academic Search Premier, EBSCO Business Source Premier, Directory of Open Access Journals, Emerald Group Publishing, Science Direct and Wiley InterScience. Peer reviewed articles published in academic journals and proceedings of the scientific conferences from 2008 till 2013 were searched in them. The following search strings: "cloud computing", "cloud", "education" and "university" – in various combinations – were used in the search. They were searched in article titles, abstracts, and keywords. The results of research are presented in Table 2 and Table 3.

Table 2
Search results for search strings "cloud computing", "cloud computing"+education, and "cloud computing"+university

Academic database	Provider	"cloud computing"		"cloud computing" + education		"cloud computing" + university	
		Title	Abstr	Title	Abstr	Title	Abstr
Academic Search Premier, Business Source Premier	EBSCO publishing	226	691	4	37	0	38
Directory of Open Access Journals	Lund University	407	709	5	34	0	15
Emerald Group Publishing	Emerald Group Publishing	7	21	0	3	0	1
Science Direct	Elsevier	143	566	2	8	0	11
Wiley InterScience	Wiley and Blackwell	54	717	0	6	0	4

After a detailed analysis of search results, due to the multidisciplinary character of a large number of matches containing the key word “cloud” in Table 3, these were eliminated from further consideration.

A total of 1005 papers were downloaded, and 114 papers related to the research objective were selected after detailed examination. 88 of these were from scientific journals, and 26 from proceedings of scientific conferences. The dispersion of academic scientific journals where the papers were published is very large.

Table 3
Search results for search strings "cloud", "cloud+education", and "cloud+university"

Academic database	Provider	cloud		cloud + education		cloud + university	
		Title	Abstr	Title	Abstr	Title	Abstr
Academic Search Premier, Business Source Premier	EBSCO publishing	4652	13629	11	117	7	240
Directory of Open Access Journals	Lund University	1593	3759	12	48	0	65
Emerald Group Publishing	Emerald Group Publishing	78	143	0	7	1	4
Science Direct	Elsevier	1357	4308	5	17	0	53
Wiley InterScience	Wiley and Blackwell	1680	4511	0	13	1	46

4 The Research Results

According to the research elaborated in the Project task [28], the university business process model hierarchy comprises five hierarchy levels: mega processes, major processes, sub-processes, activities and tasks [31]. The highest level processes identified by the university are following mega processes: the learning and teaching process, the research process, the enabling process, and the planning and governance process. Table 4 gives an overall review of identified major processes, but only the sub-processes and activities possessing the innovative and collaborative component were selected and included in the analysis of possible support by cloud computing technologies.

Innovation with elements of entrepreneurship in university processes has been the subject of consideration over the past ten years. The characteristics most often highlighted in articles as the basis of innovation are: [5] [6] [30]

- international student education and student mobility including joint educational programmes;
- development of innovative education with using interdisciplinary, problem- and project-oriented technologies;
- cooperation and collaboration between universities, business and government through the Triple Helix i.e. integration of scientific research and education;
- development of an infrastructure for interaction with the environment; and
- creation of a multi-channel financing system and appropriate university organisation and management structure.

Table 4

Overview of identified mega and major university business processes

Learning and teaching process	**Research process**
• study programme accreditation	• research planning
• teaching process preparation and realisation	• research preparation
• teaching process outcomes monitoring	• research conduct
• teaching process assessment	• research outcomes monitoring
• student and teacher mobility realisation	• research evaluation
Enabling processes	**Planning and governance processes**
• student administration services	• organisation management services
• library services	• change and business process management
• staff provision and development services	• plan development
• finance and accounting services	• budget and funds planning
• marketing, sale and distribution services	• performance assessment
• procurement services	

In the evolution of universities from traditional to innovative, their openness is one of the most significant elements of development. Survival and progress of the university as a self-governing and independent organisation is absolutely inconceivable nowadays. The university's business processes are the driving force of innovation, especially processes extending beyond its organisational boundaries. Through these processes, universities collaborate with the environment, stimulating not only their own development, but also the development of society as a whole.

From the technical point of view, cloud computing is a technology usually deployed in the resources in a cloud, technologically ready to support the multitenancy principle. Such a technology facilitates collaboration in comparison with other technologies as one of its key characteristics. Collaboration is the reason for cloud computing being a technology that provides high-quality support to the conduct of universities' business processes reaching beyond its boundaries. Such processes are identified and analysed, and the applicability of cloud computing technology in their implementation is elaborated in the following section.

Student and Teacher Mobility Realisation

Realising student and teacher mobility is one of the identified major processes in the learning and teaching mega process, in which students attend a part of their study programmes outside their home universities, at other universities, usually outside the country where the headquarters of the home university seat is, whereas teachers go to other universities to perform teaching activities.

Although student and teacher mobility is an intrinsically open process, it is performed as a set of closed activities during its implementation. Only by integrating the activities into a single unified process does it acquire sense, of course, devoid of all the benefits that could be reached by opening its constituents. The analysis of the input-output quality dimensions of the student and teacher mobility process points to the possible benefits achievable by deploying the process into a cloud. Given the significance of the segments of this process in the universities' functioning, the following section is devoted to student mobility only.

The following data categories are used and created through student mobility programmes: data on universities, data on the programmes and courses available to students participating in the mobility process, data on teachers, data on students participating in the mobility process, data on mobility applications, data on learning agreements and data on students' academic performance.

Under the assumption of applying traditional computing, evaluating earlier categories of data from the viewpoint of defined input-output quality dimensions, the following characteristics improvable by means of cloud computing can be identified:

- Low accessibility: All the above mentioned data are on the university's internal resources, and therefore hard to access. Many of the data are even exchanged as paper documents, which are subsequently recorded in the universities' information systems.

- High limitation of data amount: The student mobility process uses and generates large amounts of data, originating from various sources, and highly heterogeneous by form. Traditional computing uses limited resources by its nature, so that it can be freely said that the limitation of data amount is high.
- Low completeness and comprehensiveness: Mobility process is realised through a set of completely closed activities integrated into a whole. Such a mode of work causes data fragmentation, different standards for data, and their difficult integration. Completeness and comprehensiveness of data is difficult to achieve in such a mode, and requires high skill levels of university administrative staff.
- Low consistency: Different standards at universities result in low consistency, so that very often the same data have completely different denotation at different universities, and, similarly, in different sections of the same process.
- Fuzzy currency: Currency of a great majority of the above mentioned data depends on the currency of universities, and one set of data is, by nature, of questionable currency. This set of data refers to achieved results in the exchange, which are, in most cases, distributed by traditional mailing systems, or, in the best scenario, by e-mail. The reason for this should be sought in the existence of various university systems and lack of any connection between them.
- Fuzzy timeliness: Although process organisation defines the rules affecting the timely availability of data, delays and bottlenecks are very common in the mobility process, thus slowing it down. It must be emphasised that the process is slowed down by the existence of rules, due to the nature of data, which are often distributed through classical media.
- Average transparency: Data from the exchange process mostly stay in university archives, excluding the data whose transparency is insisted on through the rules of the exchange, such as Erasmus rules. Technologically, even if the rules of the process demand, transparency is not easy to reach in traditional computing. In traditional computing, it is achieved by combining various technologies for overcoming the limitations of the basic technology.

In addition to the above, it is necessary to pay attention to input-output quality dimensions, whose characteristics must be maintained when introducing cloud computing technology:

- High reliability: Reliability is a dimension of data quality with a positive value in the student mobility process, and is reached by means of traditional security measures, applied in processes generating important documents.
- High security: Data used and created in the student mobility process are characterised by high-quality in terms of security. They are stored in internal university bases and are subject to standards that the domicile students' data are subject to.

Such requirements can be met by using a hybrid cloud, by placing sensitive functionalities and data on a private cloud, and the remaining ones on the community cloud.

Joint Study Programmes Preparation, Realisation and Monitoring

This is a special set of sub-processes from major processes in learning and teaching mega process. Joint study programmes are programmes prepared, conducted and evaluated by two or more universities. The outcome of joint programmes can be a joint degree or a double degree. What is characteristic of joint programmes is the participation of teachers from two or more universities in the conduct of a study programme, which is joint property of two or more universities. Sub-processes related to joint study programmes include: joint programme preparation and realisation, and joint programme monitoring.

The following groups of data are used and/or created when conducting joint study programmes: data on the study programme, data on courses, data on student workload, data on teachers, data on students, data on institutions, data on students' progress and data on degrees received.

Under the assumption of using traditional computing, the following characteristics of the data categories can be identified, from the standpoint of defined input-output quality dimensions, which can be enhanced by means of cloud computing:

- Low accessibility: Only one part of data is located on a university's internal resources, and other data cannot be accessed in real time. Data are often exchanged as paper documents or electronically.
- High limitation of data amount: like in the mobility process, a large amount of data is generated and stored on the limited resources of traditional computing.
- Low completeness and comprehensiveness: This process is accomplished through a set of completely closed activities integrated into a whole. Such a mode of work causes data fragmentation, different standards for data, and their difficult integration. Like in the case of student mobility process, completeness and comprehensiveness of data is difficult to achieve in such a mode, and requires high skill levels of university administrative staff.
- Low consistency: As joint study programmes are conducted at several universities as a rule, their different standards result in low consistency, so that very often the same data have completely different denotation at different universities, and, similarly, in different sections of the same process.
- High transparency: Like in the case of the mobility process, transparency in this process is also achieved by combining various technologies overcoming the limitations of traditional technologies. The joint programmes process is a rigidly formalised one, and, as such, demands high transparency of data. Any different characteristic of this quality dimension could compromise the entire process. However, efforts invested in attaining transparency in traditional computing can be significantly reduced for this process by using cloud computing.

- Fuzzy currency: Analysis of timeliness of data points to a group of data which must be current: data on the study programme, data on courses, data on student workload, data on teachers, data on students, data on institutions, data on students' progress and data on degrees received. Similarly to the student mobility process, data on students' progress are often not current and relevant due to the existence of various university systems and lack of connection between them.
- Fuzzy timeliness: Data timeliness for most identified data groups is high, except for the data group related to student progress. Problems with timeliness mostly occur at the completion of studies, when it is necessary to record all the data at a unified place.

In addition to the above, it is necessary to pay attention to input-output quality dimensions, whose characteristics must be maintained when introducing cloud computing technology:

- High reliability and security: reliability and security are quality dimensions to which high degree of attention is devoted in traditional computing. As already stated in the description of these quality dimensions, data for student mobility process are stored in internal university bases and are subject to standards to which the data of each of the university's domicile study programmes are subject to. In the case of joint programmes, the positive assessment of these quality dimensions is further strengthened by data consolidation.

Joint Research Planning, Preparation, Realisation, Monitoring and Evaluation

This is a special set of sub-processes from **major processes** in research mega process. The research process is a process comprised of the following major processes: research planning, research preparation, research conduct, monitoring research outcomes and research evaluation. If the research process is viewed in the context of the universities' external collaboration, then it pertains solely to joint research. Joint research is research between universities and entities belonging to any of the elements of the Triple Helix (industry, government, intermediary). The research is conducted as research subsidised by national and international research support funds, commissioned by a corporate entity interested in using the research results, and the research initiated to develop the researcher's own resources.

Any form of research includes the following data classes: data on calls for proposals, data on researchers, data on funds, data on research institutions and data on research outcomes. The analysis of characteristics of the identified data categories was conducted for the joint research process as well, from the standpoint of earlier defined input-output quality dimensions, under the assumption of applying traditional computing:

- Low accessibility: although this is an open process, except in the case of commissioned research, traditional technologies make a negative impact on data accessibility. This is of particular interest for data that should be available when establishing research partnerships. Cloud computing technology creates predispositions not only for significant enhancement of accessibility, but also for managing it depending on the type of research.

- High limitation of data amount: quite identically as in the above discussed processes, the very nature of traditional computing and limitations of its resources produces high limitation of data amount.
- Low completeness and comprehensiveness: considering the nature of the research process itself, completeness and comprehensives feature as a vital requirement. Still, traditional computing makes a negative impact on the completeness and comprehensiveness of data, and very often results in their fragmentation and difficult integration.
- Average consistency: the consistency of the data used and/or created in the joint research process should be highly consistent. Harmonising the standards at the level of European Research Area raised the consistency of data to a much higher level than earlier. Still, traditional technologies often disrupt data consistency, due to implementation of local criteria and their subsequent harmonisation with the standards of the European Research Area.
- Fuzzy transparency: some of the data classes in the joint research process are characterised by high transparency, whereas some (mostly data on current research) demand low transparency. Still, traditional computing negatively affects the transparency of data classes for which a high value of this dimension is extremely important.
- Average currency and timeliness: in an ideal process realisation, data classes of the research process should be highly current and timely. However, although the research process by its nature implies current and timely data, still, in the real world, technological limitations of traditional computing result in the diminishment of these quality dimensions.

Like in the previously described processes, it is necessary to pay attention to input-output quality dimensions, whose characteristics must be retained when introducing cloud computing technology:

- High reliability and security: similar to the previous processes, the joint research conducted using traditional computing results in attaining high reliability and security. Conducting the process using cloud computing will tend to diminish it, and all technological measures must be taken in order to keep these dimensions on a high level.

Having considered the input-output quality dimensions of universities' collaborative processes based on traditional computing, Table 5 shows the influence of cloud computing on the same dimensions. The dark arrow represents an increase or decrease in the observed input-output quality dimension for all data classes, whereas a blank arrow represents a slight increase or reduction of the observed input-output quality dimensions of some or all data classes.

In addition to the above described business processes of universities' external collaboration, it is also necessary to mention the collaborative processes through which universities joins various networks such as business, professional, alumni or social networks. At the same time, one must also mention collaboration conducted through the membership of universities in various associations (business, professional, expert, and general).

Table 5

The impact of cloud computing on input-output quality dimensions of universities' processes

Processes	**Technology for process implementation**	**accessibility**	**limitation of data amount**	**completeness**	**comprehensiveness**	**transparency**	**consistency**	**currency**	**timeliness**	**reliability**	**security**
Student and teacher mobility realisation	traditional computing	L	H	L	L	M	L	F	F	H	H
	cloud computing	⬆	⬇	⬆	⬆	⬆	⬆	⇧	⇧	⬇	⬇
Joint study programmes preparation, realisation and monitoring	traditional computing	L	H	L	L	H	L	F	F	H	H
	cloud computing	⬆	⬇	⬆	⬆	⇧	⬆	⇧	⇧	⬇	⬇
Joint research planning, preparation, realisation, monitoring and evaluation	traditional computing	L	H	L	L	F	M	M	M	H	H
	cloud computing	⬆	⬇	⬆	⬆	⇧	⬆	⬆	⬆	⬇	⬇

Legend: *L - Low, M -Medium, H - High, F - Fuzzy*

Conclusion

The article proposes the use of cloud computing technologies, which, except in the case of relationships based on consulting and commercialisation foundations, significantly contribute to the attainment of expected outcomes in the universities' business processes. Given the specifics of the required data quality dimensions of certain processes at universities, such as the student and teacher mobility realisation, the joint study programmes preparation, realisation and monitoring, and the joint research planning, preparation, realisation, monitoring and evaluation, which it can fulfil, cloud computing technology is presented as successful in external collaboration processes. In particular, hybrid cloud imposes itself as the optimum solution for the implementation of cloud computing for such major processes at universities.

Cloud computing technology makes a significant impact on input-output data quality dimensions of universities' collaborative processes. This technology strongly increases accessibility, completness, comprehensiveness, consistency, and decreases limitation of data amount, reliability, and security. Also, cloud computing slightly increases currency, timeliness and transparency.

Cooperation and collaboration between universities, business i.e. industry, government and the community as a whole is a major plank in considering innovation systems. Stronger interactions and relationships between the Triple Helix entities should enhance their productivity, performance, and competitiveness. To this end, external collaboration must be formalised and enable knowledge transfer relationship in one of three relationship modes: market based contractual relationships, network based mutual relationships or integrated organisational relationships.

Implementing the cloud computing technology enables a more efficient and efficient execution of external cooperation and collaboration processes in developing a modern concept of innovative university. Furthermore, new possibilities are created in the modes of connecting the entities of the Triple Helix concept, i.e. academia, industry and government, which should provide universities with a changed status and key role in connecting the learning and teaching process with the research process.

References

[1] Attaran, M.: Exploring the Relationship between Information Technology and Business Process Reengineering, *Information & Management*, 41(5), 2004, pp. 585-596

[2] Whitman, M. E.: IT Divergence in Reengineering Support: Performance Expectations vs. Perceptions, *Information & Management*, 30(5), 1996, pp. 239-250

[3] Grover, V., Jeong, S. R., Kettinger, W. J., Teng, J. T. C.: The Implementation of Business Process Reengineering, *Journal of Management Information Systems*, 12(1), 1995, pp. 109-144

[4] Guri-Rosenblit, S., Sebkova, H., Teichler, U.: Massification and Diversity of Higher Education System: Interplay of Complex Dimensions, *Higher Education Policy, 20*, 2007, pp. 373-389

[5] Berestova, T. V.: From Innovative Projects to an Innovative University, *Scientific and Technical Information Processing*, 36(3), 2009, pp. 180-185

[6] Barluenga, M.: Becoming an Innovative University, 1st Training Module: Strategic Implementation of University-Industry Cooperation, Universidad de Alicante, 2010

[7] Petkovics, I., Tumbas, P.: Temptation of the Past in the Shape of Clouds – Cloud Computing, Conference Proceedings Informatika Korszerű Technikái IKT, Dunaújváros - Hungary, 2010, pp. 190-199

[8] Garfinkel, S. L.: The Cloud Imperative, MITTechnology Review Magazine, October, 2011

[9] Tugend, T.: UCLA to be First Station in Nationwide Computer Network. UCLA Press Realise, 1969

[10] Kleinrock, L.: A Tugend, T.: A Vision for the Internet, *St. Journal of Research* 2(1) Networked Multimedia, 2005, pp. 5-6

[11] Mell, P., Grance, T.: The NIST Definition of Cloud Computing, NIST Special Publication 800-145, September 2011

[12] Armbrust, M., et al.: Above the Clouds: A Berkeley View of Cloud Computing, Technical Report No. UCB/EECS-2009-28

[13] Naik, A. B., Ajay, A. K., Kolhatkar, S. S.: Applicability of Cloud Computing in Academia, *Indian Journal of Computer Science and Engineering* 4(1), 2013, pp. 11-15

[14] Leimeister, S., Böhm, M., Riedl, C., Krcmar, H.: The Business Perspective of Cloud Computing: Actors, Roles, and Value Networks, ECIS 2010 Proceedings, paper 56

[15] Böhm, M., Leimeister, S., Riedl, C., Krcmar, H.: Cloud Computing - Outsourcing 2.0 or a new Business Model for IT Provisioning?, Application Management: Challenges - Service Creation – Strategies, Technische Universität München (TUM), Germany, 2011, pp. 31-56

[16] Youseff, L., Butrico, M., Da Silva, D.: Toward a Unified Ontology of Cloud Computing. Grid Computing Environments Workshop, 2008, pp. 1-10

[17] Petkovic, I.: CRM in the Cloud, Proceedings of the International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 2010

[18] Esayas, S. Y.: A Walk in to the Cloud and Cloudy it Remains: The Challenges and Prospects of 'Processing' and 'Transferring' Personal Data, *Computer Law & Security Review*28(6), 2012, pp. 662-678

[19] Petkovič, I., Trninić, J., Đurković, J.: Information Technology Support of Sustainable Evolution, Proceeding of International Scientific Conference Strategic Management and Decision Support Systems in Strategic Management, Subotica-Palic, Serbia, 2013, pp. 447-460

[20] Seres, L., Horvat, L., Horvat, I.: A Thunder in the Cloud, *International Journal of Management Information Systems* 6(1), 2011, pp. 16-19

[21] Heravizadeh, M., Mendling, J., Rosemann, M.: Dimensions of Business Processes Quality, In Business Process Management Workshops, Springer Berlin Heidelberg, pp. 16-19

[22] Hammer, M, Champy, J.: Reengineering the Corporation: a Manifesto for Business Revolution, McGraw Hill, New York, 1993

[23] Davenport, T. H.: Process Innovation: Reengineering Work Through Information Technology, Harvard Business School Press, Boston, 1993

[24] Bhatt, G. D.: Exploring the Relationship between Information Technology, Infrastructure and Business Process Re-Engineering, *Business Process Management Journal* 6(2), 2000, pp. 139-163

[25] Anupindi, R., Chopra, S., Deshmukh, S., Mieghem, J., Zemel, E.: Managing Business Process Flows. Prentice Hall, 1999

[26] Kahn, B., Strong, D., Wang, R.: Information Quality Benchmarks: Product and service performance. *Communications of the ACM* 45(4), 2002, pp. 184-192

[27] Wang, R. Y., Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information System* 12(4), 1996, pp. 5-34

[28] Tumbas, P., Matkovic, P.: Projektni zadatak implementacije jedinstvenog informacionog sistema Univerziteta u Novom Sadu, Novi Sad, 2009

[29] Mircea, M., Andreescu, A. I.: Using Cloud Computing in Higher Education: A Strategy to Improve Agility in the Current Financial Crisis, Communications of the IBIMA 2011(32), 2011

[30] Christensen, C. M., Eyring, H. J.: The Innovative University, Jossey-Bass A Willey Imprint, San Francisco, 2011

[31] Tonchia, S., Tramontano, A.: Process Management for the Extended Enterprise, Springer-Verlag Berlin - Heidelberg, 2004

On the Reachability of Linear Time Varying Systems

Sándor Molnár

Department of Informatics, Institute for Mathematics and Informatics, Szent István University, Páter Károly utca 1, H-2100 Gödöllő, Hungary
e-mail: molnar.sandor@gek.szie.hu

Abstract: In this paper system properties of generalized linear time varying (LTV) systems are discussed where, in addition to the control, its certain derivatives also appear both in the dynamics and the observation equation. Developing an adequate version of the Cauchy formula, a necessary and suffiecient condition for complete reachability of generalized LTV systems is obtained in terms of a generalized Gram matrix. Starting from the expansion of coefficient functions in the corresponding Lie algebra basis, we derive an appropriate condition of persistent excitation. The latter leeds to a general condition of complete reachability in terms of quasi-polynomials of the solution of the Wei-Norman equation and differential polynomials of the coefficient functions of the generalized LTV system. Also applying the well-known duality theory of LTV systems, other basic system properties such as controllability, reconstructability and observability can be also treated.

Keywords: Controllability and reachability; Differential algebra; Linear time varying systems; Matrix Lie algebra, persistent excitation

1 Introduction

Definition 1 We call

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t) \end{aligned} \tag{1}$$

a classical linear time varying system in canonic form, where functions

$$A\ : [0,T] \to \mathbb{R}^{n\times n}, \qquad B\ : [0,T] \to \mathbb{R}^{n\times k}$$
$$C\ : [0,T] \to \mathbb{R}^{l\times n}, \qquad D\ : [0,T] \to \mathbb{R}^{l\times k},$$

are continuous on a fixed interval $[0,T]$.

Remark 1 R. Kalman solved all fundamental problems of such systems (see [9]). He proved the duality both of reachability and observability and of controllability and reconstructability. He also showed the equivalence of reachability-controllability and observability-reconstructability condition pairs for continuous time systems of (1). Therefore we are interested in only one system property, reachability.

Consider the initial value problem

$$\dot{x}(t) = A(t)x(t), \quad x(\tau) = I \tag{2}$$

in $\mathbb{R}^{n\times n}$. If the coefficient matrix is continuous, then it has a unique solution

$$t \mapsto \Phi(t,\tau) \in \mathbb{R}^{n\times n}$$

defined on the whole interval $[0,T]$, which is continuously differentiable as a two variable function of (t,τ), and $\Phi(t,\tau)$ is invertible for all pairs (t,τ). Consider the solution $t \mapsto \Psi(t,\tau)$ to the initial value problem

$$\dot{Y}(t) = -Y(t)A(t), \quad Y(\tau) = I$$

defined on the whole interval $[0,T]$. Then

$$\begin{aligned}\frac{d}{dt}\Big(\Psi(t,\tau)\Phi(t,\tau)\Big) &= \dot{\Psi}(t,\tau)\Phi(t,\tau) + \Psi(t,\tau)\dot{\Phi}(t,\tau) = \\ &= \Big(-\Psi(t,\tau)A(t)\Phi(t,\tau) + \Psi(t,\tau)\Big(A(t)\Phi(t,\tau)\Big)\Big) = 0,\end{aligned}$$

that is, $\Psi(t,\tau)\Phi(t,\tau) = I$ which implies $\Psi(t,\tau) = \Phi\ (t,\tau)^{-1}$ in $\mathbb{R}^n$. Moreover, $\Phi(t,\tau) = \Phi(t,0)\Phi(\tau,0)^{-1}$ because $t \mapsto \Phi(t,0)\Phi(\tau,0)^{-1}$ is a solution to the equation $\dot{x}(t) = A(t)x(t)$ and $\Phi(\tau,0)\Phi(\tau,0)^{-1} = I$. Interchanging t and τ,

$$\Phi(\tau,t) = \Phi(\tau,0)\,\Phi(t,0)^{-1} = \Phi(\tau,0)\,\Psi(t,0),$$

that is,

$$\begin{aligned}\frac{d}{dt}\Phi(\tau,t) &= \Phi(\tau,0)\frac{d}{dt}\Psi(t,0) = \Phi(\tau,0)\Big(-\Psi(t,0)A(t)\Big) = \\ &= -\Phi(\tau,0)\Phi(t,0)^{-1}A(t) = -\Phi(\tau,t)A(t), \Phi(\tau,\tau) = I.\end{aligned}$$

Therefore

$$\Phi(t,\tau)^{-1} = \Phi(\tau,t).$$

Definition 2 We call

$$G_{\mathrm{Re}a}[0,T]=\int_0^T \Phi(T,t)B(t)B(t)^*\Phi(T,t)^*\,dt$$

the reachability Kalman-Gram matrix.

Theorem 1 (Kalman's reachability theorem) The system (1) is reachable from state 0 if and only if the reachability Kalman-Gram matrix is invertible, or equivalently, positive definite.

Remark 2 A similar theorem holds for controllability. The controllability Kalman-Gram matrix is defined by

$$G_C[0,T]=\int_0^T \Phi(0,t)B(t)B(t)^*\Phi(0,t)^*\,dt.$$

If we define the dual system of (1) as

$$\begin{aligned}\dot{x}(t) &= A(t)^* x(t)+C(t)^* u(t)\\ y(t) &= B(t)^* x(t)+D(t)^* u(t),\end{aligned} \tag{1^*}$$

then the input in (1^*) could be denoted by *y* and the output by *u*, indicating the exchange of their roles.

Theorem 2 (Kalman's duality theorem) (1) is controllable if and only if (1^*) is reconstructable, and (1) is reachable if and only if (1^*) is observable.

Remark 3 Since the dual of the dual system is the original system, (1) is observable if and only if (1^*) is reachable, moreover (1) is reconstructable if and only if (1^*) is controllable.

The observability Kalman-Gram matrix is

$$G_0[0,T]=\int_0^T \Phi(T,t)^* C(t)^* C(t)\Phi(T,t)dt,$$

and reconstructability is equivalent to the invertability (or positive definiteness) of the reconstructability Kalman-Gram matrix

$$G_{\mathrm{Re}}[0,T]=\int_0^T \Phi(0,t)^* C(t)^* C(t)\Phi(0,t)dt.$$

In the following we investigate the reachability of the generalisation of system (1).

Definition 3 A system will be called generalised linear time varying system in canonic form where all functions are assumed to sufficiently smooth:

$$\dot{x}(t) = A(t)x(t) + \sum_{j=0}^{J} B_j(t)u^{(j)}$$
$$y(t) = C(t)x(t) + \sum_{j=0}^{J} D_j(t)u^{(j)}. \tag{3}$$

Several publications (see for example [4], [6], [7]) deal with the system (1). Based on the above we focus on the general canonic system (3), especially the theoretical construction of the persistent excitation condition, which plays an important role in the existence of a Kalman-like rank condition for linear time-varying systems.

First we recall some basics on Lie algebras.

Definition 4 Let L be a vector space over $\mathbb{R}$ endowed with a multiplication-like operation, the so-called Lie multiplication or Lie bracket: If $l_1, l_2 \in L$ then $[l_1, l_2] \in L$, $l_1 \mapsto [l_1, l_2]$ and $l_2 \mapsto [l_1, l_2]$ are linear mappings and

1) $[l, l]=0$ for all $l \in L$

2) $[l_1, l_2]+[l_2, l_1]=0$ for all $l_1, l_2 \in L$,

3) $[l_1,[l_2,l_3]]+[l_2,[l_3,l_1]]+[l_3,[l_1,l_2]]=0$ for all $l_1, l_2, l_3 \in L$.

endowed with a Lie multiplication is called a Lie algebra.

Remark 4 In the above definition

condition 2 means anticommutativity $[l_1,l_2]=-[l_2,l_1]$, and

condition 3 measures non-associativity.

Indeed,

$$[l_1,[l_2,l_3]] = -[l_2,[l_3,l_1]] - [l_3,[l_1,l_2]] = \\ [[l_1,l_2],l_3] - [l_2,[l_3,l_1]],$$

because if $[l_2,[l_3,l_1]]=0$ then the remaining equation

$$[l_1,[l_2,l_3]] = [[l_1,l_2],l_3]$$

means associativity.

Examples

1) Let $L = \mathbb{R}^{n\times n}$. If we define $[A,B] = AB - BA$ then $\left(\mathbb{R}^{n\times n},[\cdot,\cdot]\right)$ is a Lie algebra.

2) Let $\Omega \subset \mathbb{R}^n$ be an open set and consider the vector space $A(\Omega)$ of analytic functions $f\colon \Omega \to \mathbb{R}^n$. Let the Lie bracket be defined by

$$[f, g](x) = f'(x)g(x) - g'(x)f(x). \tag{4}$$

Then we obtain the Lie algebra $\left(A(\Omega), [\cdot,\cdot]\right)$.

3) Similarly define the real vector space of infinite times differentiable functions $C^\infty(\Omega)$ on an open set $\Omega \subset \mathbb{R}^n$ and define the Lie multiplication as in (4). Then we obtain the Lie algebra $\left(C^\infty(\Omega), [\cdot,\cdot]\right)$.

Definition 5 Consider system (3). The sub-Lie-algebra $L \subset \mathbb{R}^{n\times n}, \left(L, [\cdot,\cdot]\right)$ generated by

$$\left\{A(t):\ t \in [0,T]\right\} \subset \mathbb{R}^{n\times n}$$

is defined as the smallest Lie-algebra for which $\left\{A(t):\ t \in [0,T]\right\} \subset L$ holds.

Remarks 5

Such a sub-Lie-algebra exists because the set of the containing sub-Lie-algebras is non-empty, $\mathbb{R}^{n\times n}$ is an element, and the intersection of these is the minimal sub-Lie-algebra generated by $\left\{A(t)\right\}$.

Since $\mathbb{R}^{n\times n}$ has finite dimension (n^2), the sub-Lie-algebra $L \subset \mathbb{R}^{n\times n}$ is also finite dimensional.

Let $A_1, A_2, \ldots, A_I \in L$ be a basis of L. In this basis

$$A(t) = \sum_{i=1}^{I} a_i(t) A_i,$$

and the Lie bracket $[A_i, A_j] \in L$

$$[A_i, A_j] = \sum_{k=1}^{I} \Gamma_{ij}^k A_k.$$

Since $X \mapsto [A_i, X] = AdA_i(X)$ is a linear mapping on the vector space L (which is also a Lie algebra), the matrix representation of AdA_i in the basis $A_1, A_2, \ldots, A_I$ can be expressed with the help of the numbers Γ_{ij}^k.

Let $X = \sum_{j=1}^{I} x_j A_j$. Then

$$[A_i, X] = \left[A_i, \sum_{j=1}^{I} x_j A_j\right] = \sum_j x_j \left(\sum_{h=1}^{I} \Gamma_{ij}^h A_h\right),$$

$\sum_j \Gamma_{ij}^h x_j$ is the h^{th} component of the matrix-vector product

$$\begin{pmatrix} \Gamma_{i1}^1 & \Gamma_{i2}^1 & \dots & \Gamma_{il}^1 \\ \Gamma_{i1}^2 & \Gamma_{i2}^2 & \dots & \Gamma_{il}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{i1}^l & \Gamma_{i2}^l & \dots & \Gamma_{il}^l \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_l \end{pmatrix} = \Gamma_i \mathbf{x}$$

and because of the the correspondences $AdA_i X \leftrightarrow \Gamma_i \mathbf{x},\ X \leftrightarrow \mathbf{x}$, Γ_i is the representation of the matrix AdA_i in the basis $A_1, A_2, ..., A_l \in L$.

According to the Cauchy formula the solution to system with initial condition $x(0) = \xi$ is

$$x(t) = \Phi(t,0)\xi + \int_0^t \Phi(t,\tau)B(\tau)u(\tau)d\tau,$$

and similarly in the case of the generalized system (3) in canonic form:

$$x(t) = \Phi(t,0)\xi + \int_0^t \Phi(t,\tau)\left(\sum_j B_j(\tau)u^{(j)}(\tau)\right)d\tau.$$

For systems with constant coefficients the basic solutions $\Phi(t,\tau)$ are the solutions to

$$\dot{x}(t) = Ax(t), \qquad x(\tau) = I$$

i.e.,

$$\Phi(t,\tau) = \exp A(t-\tau).$$

Moreover, if the system's structure matrix $A(t)$ has the form $A(t) = a(t)A$ then the basic solutions are $\Phi(t,\tau) = \exp A\int_\tau^t a(s)ds$.

In the case of system (3)

$$\dot{x}(t) = \sum_{i=1} a_i(t)A_i x + \sum_j B_j(t)u^{(j)}(t)$$

the basic solutions take the form

$$\Phi(t,\tau) = \exp A_1 g_1(t,\tau)\exp A_2 g_2(t,\tau)\dots\exp A_l g_l(t,\tau).$$

Again, we assume that $A_1, A_2, \ldots, A_I$ is a basis in the Lie algebra L generated by $A(t)$ and also that matrix $\Gamma_i \in \mathbb{R}^{IxI}$ is the representation of AdA_i. Then the existence of the above representation is guaranteed by the Wei-Norman theorem.

Wei-Norman Theorem

Let $\gamma(t) = g(t,\tau) \in \mathbb{R}^k$ *be a solution to the so-called Wei-Norman nonlinear differential equation*

$$\left(\sum_{i=1}^{I} \left(\exp \Gamma_1 \gamma_1 \exp \Gamma_2 \gamma_2 \ldots \exp \Gamma_{i-1} \gamma_{i-1} \right) \cdot E_{ii} \right) \dot{\boldsymbol{\gamma}} = \mathbf{a} \qquad (5)$$
$$\gamma(\tau) = 0$$

where vector $\mathbf{a}$ is composed from functions a_i and E_{ii} is the (0,-1) matrix with single 1 entry at the i^{th} diagonal element.

Then

$$\Phi(t,\tau) = \exp A_1 g(t,\tau) \exp A_2 g(t,\tau) \ldots \exp A_I g(t,\tau).$$

Remarks 6 The solution locally exists, because the initial condition $\gamma(\tau) = 0$ implies that the matrix to be inverted at $\tau = 0$ is the identity, which is invertible, thus also invertible in an appropriate neighbourhood of τ, and so can be made explicit.

It is well-known (cf. [3]) that

$$\exp A_i g_i(t,\tau) = \sum_{j=0}^{n-1} q_{ij} \left(g_i(t,\tau) \right) A_i^j \qquad (6)$$

is a polynomial of A_i with maximal degree $n-1$, a quasipolynomial of $g_i(t,\tau)$ that is, a polynomial of $g_i(t,\tau)$, $\sin \alpha_l g_i(t,\tau)$, $\cos \beta_l g_i(t,\tau)$ and exponential of $\lambda_l g_i(t,\tau)$, where λ_l are the real parts and α_l, β_l are the imaginary parts of the eigenvalues of A_i. (The basic results can be found in the classical monographs of matrix theory and ordinary differential equations, such as [3]).

Substituting (6) into the exponential product we obtain

$$\Phi(t,\tau) = \sum_{\mathbf{n}} Q_{\mathbf{n}} \left(\mathbf{g}(t,\tau) \right) A_1^{n_1} A_2^{n_2} \ldots A_I^{n_I},$$

where $Q_{\mathbf{n}}$ is a quasipolynomial of $\mathbf{g}(t,\tau) = \left(g_1(t,\tau), g_2(t,\tau), \ldots, g_I(t,\tau) \right)$ (a certain product of quasipolynomials $g_{ij}\left(g_i(t,\tau) \right)$).

2 Reachability of the Canonic System

Lemma 1 Consider the final states of the generalized linear system with time-varying coefficients (3) and with the initial condition $x(0)=0$. Then with appropriate integer constants $C_{\alpha,\beta}$

$$x(T)=\int_0^T \Phi(T,t)\left[\sum_{j=0}^{J}\left(\sum_{|\boldsymbol{\alpha}+\boldsymbol{\beta}|=j-\hat{j}} C_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(A^{(\boldsymbol{\alpha})}(t)\right)^{\boldsymbol{\beta}}\right)B_j^{(\hat{j})}(t)\right]u(t)dt.$$

Here we apply the following notation:

for integers $0\le \alpha_1,\alpha_2,\dots,\alpha_\gamma$, $1\le \beta_1,\beta_2,\dots,\beta_\gamma$,

$$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\dots,\alpha_\gamma),\quad \boldsymbol{\beta}=(\beta_1,\beta_2,\dots,\beta_\gamma),$$

$$|x|=\sum|\alpha|,$$

$$\left(A^{(\boldsymbol{\alpha})}(t)\right)^{\boldsymbol{\beta}}=\left(A^{(\alpha_1)}(t)\right)^{\beta_1}\left(A^{(\alpha_2)}(t)\right)^{\beta_2}\dots\left(A^{(\alpha_\gamma)}(t)\right)^{\beta_\gamma}.$$

Proof By the Cauchy formula,

$$x(T)=\int_0^T \Phi(T,t)\left(\sum_j B_j(t)u^{(j)}(t)\right)dt. \tag{7}$$

If the highest order derivative of u that appears is $u^{(J)}(t)$ then we assume that for all $j=0,1,2,\dots,J-1$ the boundary conditions $u^{(j)}(0)=0$, $u^{(j)}(T)=0$ hold. This assumption does not affect the subspace of the reachable final states in $\mathbb{R}^n$.

$$\begin{aligned}x(T)&=\int_0^T\Phi(T,t)B_0(t)u(t)dt+\sum_{j=1}\int_0^T\Phi(T,t)B_j(t)u^{(j)}(t)dt=\\&=\int_0^T\Phi(T,t)B_0u(t)dt+\sum_{j\ge1}[\Phi(T,t)B_j(t)u^{(j-1)}(t)]_0^T-\\&-\sum_{j=1}\int_0^T\frac{d}{dt}\left(\Phi(T,t)B_j(t)\right)u^{(j-1)}(t)dt=\int_0^T\Phi(T,t)B_0(t)u(t)dt+\\&+\int_0^T\Phi(T,t)\left(A(t)B_1(t)-B_1'(t)\right)u(t)dt+\sum_{j\ge2}\int_0^T\Phi(T,t)\Big(A(t)B_j(t)-B_j'(t)\Big)u^{(j-1)}(t)dt.\end{aligned}$$

Repeating this for the last term we obtain the equations of the next step:

$$\sum_{j\geq 2}\int_0^T \Phi(T,t)\Big(A(t)B_j(t)-B_j'(t)\Big)u^{(j-1)}(t)dt=\sum_{j\geq 2}\Big[\Phi(T,t)\Big(A(t)B_j(t)-B_j'(t)\Big)u^{(j-2)}(t)\Big]_0^T-$$

$$-\int_0^T \frac{d}{dt}\Big(\Phi(T,t)\Big(A(t)B_2(t)-B_2^{'}(t)\Big)\Big)u(t)dt-\sum_{j\geq 3}\int_0^T \frac{d}{dt}\Big(\Phi(T,t)\Big(A(t)B_j(t)-B_j^{'}(t)\Big)\Big)$$

$$u^{(j-2)}(t)dt=\int_0^T \Phi(T,t)\Big(A(t)^2B_2(t)-2A(t)B_2'(t)-A'(t)B_2(t)+B''(t)\Big)u(t)dt+$$

$$+\sum_{j\geq 3}\int_0^T \Phi(T,t)\Big(A(t)^2B_j(t)-2A(t)B_j'(t)-A'(t)B_j(t)+B_j''(t)\Big)u^{(j-2)}(t)dt.$$

Again, integrating by parts in the last term, we have a similar equation:

$$\sum_{j\geq 3}\int_0^T \Phi(T,t)\Big(A(t)^2B_j(t)-2A(t)B_j'(t)-A'(t)B_j(t)+B_j''(t)\Big)u^{(j-2)}(t)dt=$$

$$=\int_0^T \Phi(T,t)\Big(A(t)^3B_3(t)-2A(t)A(t)'B_3(t)-A'(t)A(t)B_3(t)-3A(t)^2B_3'(t)+3A'(t)B_3'(t)+$$

$$+3A(t)B_3''(t)-B_3'''(t)\Big)u(t)dt+$$

$$+\sum_{j\geq 4}\int_0^T \Phi(T,t)\Big(A(t)^3B_j(t)-3A(t)^2B_j^{'}(t)-2A(t)A'((t))B_j(t)+$$

$$+3A(t)B_j''(t)-A'(t)A(t)B_j(t)+3A'(t)B_j'(t)+A''(t)B_j(t)-B_j'''(t)\Big)u^{(j-3)}(t)dt.$$

After the j^{th} step, no derivative of $u(t)$ appears in the integral. Then (applying the above notation)

$$x(T)=\int_0^T \Phi(T,t)B_0(t)u(T)dt+\int_0^T \Phi(T,t)\Big(A(t)B_1(t)-B_1'(t)\Big)u(t)dt+$$

$$+\int_0^T \Phi(T,t)\Big(A(t)^2-A'(t)B_2(t)-2A(t)B_2'(t)+B_2''(t)\Big)u(t)dt+$$

$$+\int_0^T \Phi(T,t)\Big(\Big(A(t)^3-2A(t)A'(t)-A'(t)A(t)\Big)B_3(t)-$$

$$-\Big(3A(t)^2-3A'(t)\Big)B_3'(t)+3A(t)B_3''-B_3'''\Big)u(t)dt+$$

$$+\ldots=$$

$$=\int_0^T \Phi(T,t)\left[\sum_{j=0}^{J}\left(\sum_{|\alpha+\beta|=j-\hat{j}} C_{\alpha,\beta}\left(A^{(\alpha)}(t)\right)^{\beta}\right)B_j^{(\hat{j})}(t)\right]u(t)dt,$$

which was to be proved. ■

Definition 8 The reachability Kalman-Gram matrix for generalised systems is defined as

$$G_{\mathrm{Re}a}[0,T] =$$
$$\int_0^T \Phi(T,t)\left[\sum_{j=0}^{J}\left(\sum_{|\boldsymbol{\alpha}+\boldsymbol{\beta}|=j-\hat{j}} C_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(A^{(\boldsymbol{\alpha})}(t)\right)^{\boldsymbol{\beta}}\right)B_j^{(\hat{j})}(t)\right]\cdot$$
$$\cdot\left[\sum_{j=0}^{J}\left(\sum_{|\boldsymbol{\alpha}+\boldsymbol{\beta}|=j-\hat{j}} C_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(A^{(\boldsymbol{\alpha})}(t)\right)^{\boldsymbol{\beta}}\right)B_j^{(\hat{j})}(t)\right]^{Tr}\Phi(T,t)^{Tr}\,dt.$$

Theorem 3 *The general linear time-varying system on the interval* $[0,T]$ *is completely reachable if and only if the Kalman-Gram matrix* $G_{\mathrm{Re}a}[0,T]$ *is invertible, or equivalently, positive definite.*

Proof Based on our first lemma (Lemma 1) the proof is similar to the case of classical time-varying linear systems.

Starting from the expansion of coefficient functions in the corresponding Lie algebra basis, we derive an appropriate condition of persistent excitation. The latter leeds to a general condition of complete reachability in terms of quasi-polynomials of the solution of the Wei-Norman equation and differential polynomials of the coefficient functions of the generalized LTV system. Also applying the well-known duality theory of LTV systems, other basic system properties such as controllability, reconstructability and observability can be also treated.

Let L be the Lie algebra generated by $\{A(t):\ t\in[0,T]\}\subset\mathbb{R}^{n\times n}$ and let $A_1, A_2, \dots, A_I \in L$ be a basis. In this basis

$$A(t)=\sum_{i=1}^{I} a_i(t)A_i.$$

Similarly, for the matrices $B_0(t), B_1(t), \dots, B_J(t)$, if

$$V=V\{B_0(t), B_1(t), \dots, B_J(t):\ t\in[0,T]\}\subset R^{n\times k}$$

is the subspace of the vector space $\mathbb{R}^{n\times k}$ spanned by $B_j(t)$ then a basis $B_1, B_2, \dots, B_{\hat{I}}$ can be chosen in V such that

$$B_j(t)=\sum_{\hat{i}=1}^{\hat{I}} b_{j\hat{i}}(t)B_{\hat{i}}.$$

Now $\Phi(T,t)$ can be written as a polynomial of $A_1, A_2, \dots, A_I$, a quasi-polynomial of the solutions $g_1, g_2, \dots, g_I$ to the Wei-Norman equation [8], and the kernel function $\sum_{j=0}^{J}\left(\sum_{|\boldsymbol{\alpha}+\boldsymbol{\beta}|=j-\hat{j}} C_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left(A^{(\boldsymbol{\alpha})}(t)\right)^{\boldsymbol{\beta}}\right)B_j^{(\hat{j})}(t)$ in the integral form of $x(T)$ can be written as a polynomial of $A_1, A_2, \ldots, A_I$ and $B_1, B_2, \ldots, B_{\hat{I}}$ and a differential polynomial of $a_i(t)$ and $b_{j\hat{i}}(t)$ (with integer coefficients and first degree $B_{\hat{i}}$). Exchanging adjacent terms, the powers of A_i can be arranged in the natural order (having the form $A_1^{m_1}, A_2^{m_2}, \ldots, A_I^{m_I}$, where all m_i satisfy $0 \le m_i \le n_i$, or $\mathbf{0} \le \mathbf{m} < \mathbf{n}$) using the equations $A_{i_1}A_{i_2} = A_{i_2}A_{i_1} + \sum_{h=1}^{I}\Gamma_{i_1 i_2}A_h$ and $A_i^n = \sum_{\hat{i}=0} C_{i\hat{i}} A_i^{\hat{i}}$ assuming that the characteristic polynomial of A_i has the form $\lambda^n = \sum_{i=0}^{n-1} C_{i\hat{i}}\lambda^i$.

Then we have

$$x(T) = \sum_{\mathbf{0}\le\mathbf{m}<\mathbf{n}}\sum_{\hat{i}=0}^{\hat{I}} A_1^{m_1}\cdot A_2^{m_2}\cdot\ldots\cdot A_I^{m_I} B_{\hat{i}}\int_0^T P_{\mathbf{m},\hat{i}}$$
$$\left(\mathbf{g}(T,t), \mathbf{a}^{[\infty)}(t), \mathbf{b}_1^{[\infty)}(t), \ldots, \mathbf{b}_J^{[\infty)}(t)\right)u(t)dt.$$

Here $P_{\mathbf{m},\hat{i}}\left(\mathbf{g}(T,t), \mathbf{a}^{[\infty)}(t), \ldots, b_j^{[\infty)}(t), \ldots\right)$ are quasi-polynomials of $g_1(T,t), g_2(T,t), \ldots, g_I(T,t)$ and differential polynomials of $a_1(t), \ldots, a_I(t)$, and $b_{11}(t), b_{12}(t), \ldots, b_{1\hat{I}}(t)$, $b_{21}(t), \ldots, b_{2\hat{I}}(t), \ldots, b_{J1}(t), b_{J2}(t), \ldots, b_{J\hat{I}}(t)$, and the following notation is used

$$\mathbf{a}^{[\infty)} = (a, a', a'', \ldots, a^{(\mu)}),$$
$$\mathbf{b}^{[\infty)} = (b, b', b'', \ldots, b^{(\mu)}),$$

where μ, ν are arbitrary (but finite) nonnegative integers.

This implies that the reachable subspace of the general system on [0,T] must be a subset of the image space

$$\mathrm{Im}\left\{\ldots, A_1^{m_1}\cdot A_2^{m_2}\cdot\ldots\cdot A_I^{m_I}\cdot B_{\hat{i}}, \ldots\right\}$$

which is similar to the case of classical canonical systems.

The reachability subspace is extended because the derivatives can also be inputs, therefore let

$$V_0 = V\left\{B_0(t);\ t\in[0,T]\right\} \subset \mathbb{R}^{n\times k},$$

$$V_J = V\left\{B_0(t), B_1(t), \ldots, B_J(t);\ t\in[0,T]\right\} \subset \mathbb{R}^{n\times k}$$

be the subspaces generated by the corresponding $B_J(t)$ matrices. Choose a basis of V_J such that the first $\hat{I}_0$ elements form a basis of V_0:

$$V_0 = V\left\{B_1, B_2, \ldots, B_{\hat{I}_0}\right\},$$
$$V_J = V\left\{B_1, B_2, \ldots, B_{\hat{I}_0},\ B_{\hat{I}_0+1}, B_{\hat{I}_0+2}, \ldots, B_{\hat{I}_J}\right\}$$

From this it is obvious that for the general system, the image of the corresponding generalized Kalman-matrices (briefly Kalman matrices in the following) contains the image of the general Kalman matrices of the classical system.

From the proofs for the classical system one can deduce the persistent excitation condition which guarantees that the reachability subspace of the general system coincides with the image of the image of the general Kalman-matrix of the system.

Suppose that ξ is a vector in the image space of the Kalman-matrix,

$$K_{\text{gen}} = \left\{\ldots, A_1^{m_1} \cdot A_2^{m_2} \cdot \ldots \cdot A_I^{m_I} \cdot B_{\hat{i}}, \ldots\right\}.$$

K_{gen} does not equal the reachable subspace of the general system over $[0,T]$ if and only if there exists $\xi \neq 0$, $\xi \in \text{Im}(K_{\text{gen}})$, such that $\langle \xi, x(T)\rangle = 0$ for all possible inputs $u(t)$.

This means

$$0 = \langle \xi, x(T)\rangle =$$
$$= \left\langle \xi, \sum_{\mathbf{0}\le\mathbf{m}<\mathbf{n}} \sum_{\hat{i}=0}^{\hat{I}} A_1^{m_1} A_2^{m_2} \ldots A_I^{m_I} B_{\hat{i}} \int_0^T P_{\mathbf{m},\hat{i}}\left(\mathbf{g}(T,t), \mathbf{a}^{[\infty]}(t), \ldots, \mathbf{b}_j^{[\infty]}, \ldots\right) u(t)dt \right\rangle =$$
$$= \int_0^T \left\langle \sum_{\mathbf{0}\le\mathbf{m}<\mathbf{n}} \sum_{\hat{i}=0}^{\hat{I}} P_{\mathbf{m},\hat{i}}\left(\mathbf{g}(T,t)\mathbf{a}^{[\infty]}(t)\ldots \mathbf{b}_j^{[\infty]}\ldots\right)\left(A_I^T\right)^{m_I}\left(A_{I-1}^T\right)^{m_{I-1}}\ldots\left(A_1^T\right)^{m_1} B_{\hat{i}}^T \xi, u(t)\right\rangle dt.$$

By the classical Lagrange lemma, if the above holds for every "nice", e.g. continuous function u then

$$\sum_{\mathbf{0}\le\mathbf{m}<\mathbf{n}} \sum_{\hat{i}=0}^{\hat{I}} P_{\mathbf{m},\hat{i}}\left(g(T,t), \mathbf{a}^{[\infty]}(t), \ldots, \mathbf{b}_j^{[\infty]}, \ldots\right) \left(A_I^T\right)^{m_I}\left(A_{I-1}^T\right)^{m_{I-1}}\ldots\left(A_1^T\right)^{m_1} B_{\hat{i}}^T \xi = 0 \qquad (8)$$

The analytic functions $\exp\lambda g, \cos\alpha g, \sin\alpha g$ in the quasi-polynomials $P_{\mathbf{m},\hat{i}}$ can be replaced by new variables $\overline{g} = \exp\lambda g$, $\hat{g} = \cos\alpha g$, $\breve{g} = \sin\alpha g$. The corresponding differential equations are:

$$\begin{aligned} \dot{\overline{g}} &= \lambda\dot{g}\exp\lambda g = \lambda\dot{g}\overline{g}, \quad \text{and} \\ \dot{\hat{g}} &= -\alpha\dot{g}\sin\alpha g = -\alpha\dot{g}\breve{g}, \\ \dot{\breve{g}} &= \alpha\dot{g}\cos\alpha g = \alpha\dot{g}\hat{g}. \end{aligned} \tag{9}$$

To make these differential equations explicit consider the Wei-Norman differential equation

$$\left(\sum_{i=1}^{I}\exp\Gamma_1 g_1 \exp\Gamma_2 g_2 \dots \exp\Gamma_{i-1} g_{i-1} E_{ii}\right)\dot{\mathbf{g}} = \mathbf{a}, \ \mathbf{g}(0) = \mathbf{0}.$$

The exponential products are exponents in the multiplication table of the Lie algebra $\Gamma_1, \Gamma_2, \dots, \Gamma_{I-1}$ (the Christoffel symbols). Again, we can introduce the non-polynomial terms $\exp\lambda g, \cos\alpha g, \sin\alpha g$ as new variables, which means adding more differential equations of the type (9) that are polynomial. Thus the Wei-Norman equation becomes polynomial but non-explicit.

The equation can be made explicit in the original derivatives $\dot{g}$:

$$\dot{g} = \left(\sum_{i=1}^{I}\exp\Gamma_1 g_1 \exp\Gamma_2 g_2 \dots \exp\Gamma_{i-1} g_{i-1} E_{ii}\right)^{-1}\mathbf{a},$$

and equations (9) also become explicit with fractional denominators:

$$\det\left(\sum_{i=1}^{I}\exp\Gamma_1 g_1 \exp\Gamma_2 g_2 \dots \exp\Gamma_{i-1} g_{i-1} E_{ii}\right).$$

Multiplying the system of explicit equations by these, in the end we obtain an implicit polynomial differential equation with variables $\mathbf{g}, \overline{\mathbf{g}}, \overset{\frown}{\mathbf{g}}, \mathbf{g}$, where each equation contains only one derivative, that is, a regular differential equation which can be made explicit in the derivatives (with fractional right hand sides).

Thus the quasi-polynomials $P_{\mathbf{m},\hat{i}}\left(\mathbf{g}(T,t), \mathbf{a}^{[\infty)}(t), \mathbf{b}_1^{[\infty)}(t), \mathbf{b}_2^{[\infty)}(t), \dots, \mathbf{b}_I^{[\infty)}(t)\right)$ can be replaced by polynomials

$$\overline{P}_{\mathbf{m},i}\left(\mathbf{g}(T,t), \overline{\mathbf{g}}(T,t), \hat{g}(T,t), \breve{g}(T,t), \mathbf{a}^{[\infty)}(t), \mathbf{b}_1^{[\infty)}(t), \mathbf{b}_2^{[\infty)}(t), \dots, \mathbf{b}_I^{[\infty)}(t)\right)$$

of variables $\mathbf{g}, \overline{\mathbf{g}}, \hat{\mathbf{g}}, \breve{\mathbf{g}}$ and differential polynomials of functions $\mathbf{a}(t)$ $\mathbf{b}_0(t), \mathbf{b}_1(t), \dots, \mathbf{b}_J(t)$.

Denote the variables $\mathbf{g}, \overline{\mathbf{g}}, \hat{\mathbf{g}}, \breve{\mathbf{g}}$ by $\mathbf{x} = (x_1, x_2, \dots, x_N)$ and by $\mathbf{u} = (u_1, u_2, \dots, u_K)$, and rewrite the above implicit polynomial differential equation as

$$F(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \ddot{\mathbf{u}}, \dots) = 0.$$

We also rewrite equation (8) using $\mathbf{x}, \mathbf{u}$

$$\sum_{\mathbf{0} \le \mathbf{m} < \mathbf{n}} \sum_{\hat{i}=0}^{I} \overline{P}_{bm,\hat{i}}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}, \dots) \left(A_I^T\right)^{m_I} \left(A_{I-1}^T\right)^{m_{I-1}} \dots \left(A_1^T\right)^{m_1} B_{\hat{i}}^T \xi = 0.$$

Define the output equation

$$y = \sum_{\mathbf{0} \le \mathbf{m} < \mathbf{n}} \sum_{\hat{i}=0}^{I} \overline{P}_{\mathbf{m},\hat{\mathbf{i}}}(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, \dots,) \left(A_I^T\right)^{m_I} \left(A_{I-1}^T\right)^{m_{I-1}} \dots \left(A_1^T\right)^{m_1} B_{\hat{i}}^T \xi = G\left(\mathbf{x}, u, \dot{u}, \dots, \xi\right).$$

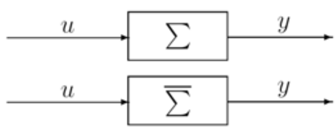
Thus we have an input-output system

$$\begin{aligned} F\left(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, ,\dots, \xi\right) &= 0 \\ \mathbf{y} &= G(\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, ,\dots, \xi), \end{aligned} \qquad (\Sigma)$$

which is polynomial and implicit in the derivatives $\dot{x}$, with the regularity condition $\partial_{\dot{x}} F(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, \ddot{\text{u}}, \dots, \xi) \neq 0$. Here **u** are the inputs, **x** are the states and **y** are the outputs. Consider another representation with possibly different states but with the same inputs and outputs

$$\begin{aligned} \overline{F}\left(\overline{\mathbf{x}}, \dot{\overline{\mathbf{x}}}, \mathbf{u}, \dot{\mathbf{u}}, ,\dots, \xi\right) &= 0 \\ y &= \overline{G}(\overline{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, ,\dots, \xi). \end{aligned} \qquad (\overline{\Sigma})$$

Let $(\sum)$ and $(\overline{\sum})$ be input-output systems. We call them equivalent if for every input-output pair (u, y), has a solution $\mathbf{x}$ if and only if has a solution $\overline{\mathbf{x}}$. In this case the two systems



behave in the same way. $\sum$ and $\overline{\sum}$ can be written more briefly also allowing the derivatives of the outputs y

$$J(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}, \dots, \mathbf{y}, \dot{\mathbf{y}}, \ddot{\mathbf{y}}, \dots, \xi) = 0, \qquad (10)$$

and

$$\overline{J}(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}, \dots, \mathbf{y}, \dot{\mathbf{y}}, \ddot{\mathbf{y}}, \dots, \xi) = 0.$$

Diop [1] proved the existence of a finite purely algebraic algorithm which gives differential polynomials

$$\hat{J}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi)$$
$$\hat{G}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi)$$

such that the system (10) is equivalent to the input-output system $\mathbf{u} \mapsto \dot{\mathbf{y}}$ defined by the implicit equation and the non-equality condition

$$\begin{aligned} \hat{J}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi) = 0 \\ \hat{G}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi) \neq 0. \end{aligned} \tag{11}$$

The latter has no state variable $\mathbf{x}$ thus we can call the Diop algorithm a state elimination algorithm.

Definition 9 (10) and (11) define equivalent input-output systems if for any input-output pair (u, y), (10) has a solution with respect to the state $\mathbf{x}$, i.e. the triple (x,u,y) is a solution to (10) if (u, y) is a solution to the polynomial equation

$$\hat{J}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi) = 0$$

and

$$\hat{G}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi) \neq 0$$

holds.

Remark 7 We get this latter result by dividing by a differential polynomial in each step of the algorithm, and since the divisor obviously cannot be *0*, this must be assumed. Their products form the differential polynomials $\hat{G}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\ldots,\mathbf{y},\dot{\mathbf{y}},\ddot{\mathbf{y}},\ldots,\xi)$. If this product $\neq 0$ then neither of its factors can be 0.

Now returning to the original input-output system $(\sum)$, taking the "input" $\mathbf{u} = (\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J)$ ordered continuously in a row, we have

$$F\big(\mathbf{x},\dot{\mathbf{x}},(\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_j),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_j),\ldots,\xi\big) = 0,$$
$$G\big(\mathbf{x},(\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_j),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_j),\ldots,\xi\big) = 0.$$

Substituting the input-output pair $\big((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),0\big)$ into the state eliminated system obtained from system $(\sum)$, the equivalence of the systems yields the equation and non-equality

$$\begin{aligned}&\hat{J}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,0,\ldots,\xi\right)=0\\&\hat{G}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,0,\ldots,\xi\right)\neq 0\end{aligned}\tag{12}$$

which give a sufficient condition that for all inputs $\mathbf{u}$, the end state $x(T)$ is orthogonal to the given vector

$$\xi \in \operatorname{Im}\left\{\ldots, A_1^{m_1} A_2^{m_2} \ldots A_I^{m_I} B_j, \ldots\right\}. \tag{13}$$

Definition 10 We say that the time-variant coefficients $\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J$ persistently excite the system if the subspace of the reachable states coincides with the image space of the generalized Kalman-matrix, i.e., it is the largest possible subspace. According to our equations, if the coefficients satisfy the conditions (12) then state ξ must be 0.

The most interesting special case is when the image space of the generalized Kalman-matrix is the whole space $\mathbb{R}^n$. Then the coefficients $\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J$ persistently stimulate the system if and only if the system is totally reachable on the interval $[0,T]$.

Thus if the coefficients do not persistently excite the system then

$$\begin{gathered}\hat{J}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,0,0,\ldots,\xi\right)=0\\\hat{G}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,0,0,\ldots,\xi\right)\neq 0\\\xi\neq 0\end{gathered}$$

can be solved. Regarding the equation as an implicit function of ξ, it can be solved for ξ,

$$\boldsymbol{\xi}=\hat{f}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,\right). \tag{14}$$

Writing this into the two non-equalities we have that the condition of "persistent non-excitation" is the paralel fulfillment of the two non-equalities:

$$\begin{aligned}0\neq\hat{G}\Big(&(\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,\ldots,\\&\hat{f}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots\right)\\0\neq\hat{f}&(\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J)\ldots\Big)\end{aligned}$$

Negation of these statements gives the condition for persistent excitation

$$0 = \hat{G}(\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots,0,0,\ldots,$$
$$\hat{f}\left((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J),\ldots\right)$$

or

$$0 = \hat{f}\Big((\mathbf{a},\mathbf{b}_0,\mathbf{b}_1,\ldots,\mathbf{b}_J),(\dot{\mathbf{a}},\dot{\mathbf{b}}_0,\dot{\mathbf{b}}_1,\ldots,\dot{\mathbf{b}}_J)\ldots\Big).$$

Returning to the solvability of the implicit function of ξ we can obtain (14).

Again Diop's elimination theorem (algorithm) can be applied. Regard the vector ξ as a state that can be eliminated. For this we would need a state equation, a dynamics as a differential equation for ξ. But since ξ is a constant the dynamics is simply $\dot{\boldsymbol{\xi}} = 0$.

References

[1] Diop, S.: Elimination in Control Theory, *Math Control Signals Systems*, Signals, 4, 17-32, 1991

[2] Fliess M.: Controllability Revisited, in Mathematical System Theory: The Influence of R. E. Kalman, A. C. Antoulas (Ed.), *Springer-Verlag*, Berlin, 1991

[3] Gantmacher F. R., The Theory of Matrices, Chelsea Publishing Company, New York, N.Y., 1974

[4] Molnár S. and Szigeti F.: A Generalisation of Fuhrmann's Rank Condition for Discrete Dynamic Systems, *Int. J. System of Systems Engineering,* Vol. 2(4), pp. 279-289, 2011. DOI: 10.1504/IJSSE.2011.043864

[5] Silverman L. M.: Controllability and Observability in Time-Variable Linear Systems, *SIAM Journal on Control*, Vol. 5(1), pp. 64-73, 1967, DOI: 10.1137/0305005

[6] Szigeti F.: A Differential-Algebraic Condition for Controllability and Observability of Time Varying Linear Systems, Decision and Control, Proceedings of the 31st IEEE Conference on, pp. 3088-3090, 1992, Tucson, AZ, USA, DOI: 10.1109/CDC.1992.371050

[7] Szigeti F.: Kalman's Rank Conditions for Infinite Dimensional Time Dependent Linear Systems. In: Proc. Conf. EQUADIFF. pp. 927-931, 1992, Barcelona, Spain

[8] Wei, J. and E. Norman: On Global Representations of the Solutions of Linear Differential Equations as a Product of Exponentials, Proc. Amer. Math. Soc. 15(12), pp. 327-334, 1964

[9] Kalman R. E., Falb P. L., Arbib M. A.: Topics in Mathematical Systems Theory, McGraw Hill Book Company, New York, Toronto, Sydney, 1969

A New Sliding Mode Speed Observer of Electric Motor Drive Based on Fuzzy-Logic

Chiheb Ben Regaya, Abderrahmen Zaafouri, Abdelkader Chaari

Unit C3S, High School of Sciences and Techniques of Tunis (ESSTT), 5 Av. Taha Hussein, BP 56, 1008 Tunis, Tunisia

E-mails: chiheb.benregaya@enseignant.edunet.tn; abderrahmen.zaafouri@isetr.rnu.tn; assil.chaari@esstt.rnu.tn

Abstract: In this paper, the speed of the induction machine is controlled by a variable structure controller. To eliminate speed sensor we use a sliding mode observer based on fuzzy logic "FSMSO". The control algorithm and observation is emphasized by simulation tests and it is comapred with the classical sliding mode speed observer"SMSO". Analysis of the results shows the characteristic robustness to disturbances of the load and the speed variation for the FSMSO than SMSO.

Keywords: induction motor; fuzzy sliding mode speed observer; sliding mode control; sliding mode speed observer; vector control

1 Introduction

Vector control is the most common technique used to control the induction motor, since it offers performance very close to that of the dc motor. But it remains sensitive to parametric variation. For this reason a lot of research has proposed nonlinear control laws with parameter identification and state estimation [1-4]. Among them, the back stepping control, the sliding mode control. The latter has a good performance and is insensitive to parametric variations [5, 6].

But the use of sign functions in control law or observer allows the appearance of the chettering phenomenon that can exite the high frequencies qnd the latter can damage the system [7, 8]. To alleviate this drawback, several solutions have been proposed, such as using a transition band around the sliding surface, or an integrator block in the controller output [9, 10]. In both solutions the phenomenon of chattering is surely reduced, but the tracking error remains.

To make the observer more robust to to disturbances of the load and the speed variation while reducing the chattering phenomenon, we propose in our paper a sliding mode speed control with a new sliding mode speed observer based on fuzzy logic.

In this paper, in addition to the sliding mode speed control, a speed observation system based on fuzzy logic comprising a current observer, a rotor flux observer and a rotor speed observer, is presented for an indirect vector control of an induction motor drive. The present paper is organized as follows: in Section 2, the model of induction motor is defined. The sliding mode control of speed is established in Section 3. In Section 4, the classical sliding mode speed observer is developed. The fuzzy sliding mode speed observer is the subject of the fifth part. We close our paper with a conclusion.

2 Induction Motor Modeling

In a reference connected to the stator, we can express the model of the induction machine by the following state equation [1, 3]:

$$\frac{d}{dt}\begin{bmatrix} i_s \\ \phi_r \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} i_s \\ \phi_r \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} v_s \qquad (1)$$

$$i_s = Cx$$

Where:

$i_s = \begin{bmatrix} i_{\alpha s} & i_{\beta s} \end{bmatrix}^{\mathrm{T}}$:Stator current;

$v_s = \begin{bmatrix} v_{\alpha s} & v_{\beta s} \end{bmatrix}^{\mathrm{T}}$:Stator voltage;

$\phi_r = \begin{bmatrix} \phi_{\alpha r} & \phi_{\beta r} \end{bmatrix}^{\mathrm{T}}$:Rotor flux;

R_s , R_r : Resistance of Stator and Rotor;

L_s , L_r : Stator and Rotor inductance;

M : Mutual inductance;

$\sigma = 1 - \left(M^2 / L_s L_r \right)$: Total leakage factor;

$\tau_r = L_r / R_r$: Rotor time constant;

ω_r : Angular rotor speed;

$$A_{11} = \left[(-1/\sigma)\left((R_s/L_s) + (1 - \sigma/\tau_r) \right) \right] I$$

$$A_{12} = \left(M/\sigma L_s L_r\right)\left[\tau_r^{-1} I - \omega_r J\right]$$

$$A_{21} = \left(M/\tau_r\right) I \; ; A_{22} = \omega_r J - \left(1/\tau_r\right) I$$

$$B_1 = \left(1/\sigma L_s\right) I \; ; C = \left[I \quad 0\right]$$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} ; \; J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

3 Sliding Mode Speed Control

The sliding mode control is a strategy to reduce the state trajectory to the sliding surface and to move it to the balance point [5].

The design of this type of control requires an appropriate choice of the switching surface, the state of convergence and the calculation of the control law.

The model of the induction machine can be represented by the following mechanical equation:

$$\begin{aligned} & j \frac{d}{dt} \omega_r = C_{em} - C_r - f \omega_r \\ & C_{em} = \frac{3}{2} p \frac{M}{L_r} \left(\phi_{dr} i_{qs} - \phi_{qr} i_{ds} \right) \end{aligned} \tag{2}$$

For representation in the state space, we set:

$$\begin{cases} x_1 = \omega^* - \omega_r \\ x_2 = \dfrac{d}{dt} \omega_r = -\dot{x}_1 \end{cases} \tag{3}$$

The speed error is processed by a variable structure controller for generating the control signal U. To cancel the static and reduce the effect of chattering, we add an integrator to the sliding mode regulator with a response time $\boldsymbol{\tau}$. The simplified model can be written as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & -\dfrac{f}{j} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{1}{j\tau} \end{bmatrix} U \tag{4}$$

To have a response to the system fast without overshoot, the right commutation or sliding surface can be chosen as follows:

$$S_{(x)} = \upsilon x_1 + \dot{x}_1 \tag{5}$$

With υ the slope of sliding line.

The control law is defined as follows:

$$U = \psi_1 x_1 + \psi_2 \dot{x}_1 \tag{6}$$

Where $\psi_1$and ψ_2 are coefficient of the sliding mode to be designed.

With:

$$\psi_1 = \begin{cases} \alpha_1 & if\ S_{(x)} x_1 > 0 \\ \beta_1 & if\ S_{(x)} x_1 < 0 \end{cases} \quad \psi_2 = \begin{cases} \alpha_2 & if\ S_{(x)} \dot{x}_1 > 0 \\ \beta_2 & if\ S_{(x)} \dot{x}_1 < 0 \end{cases} \tag{7}$$

The convergence condition is defined by the Lyapunov equation in order to make the area attractive and invariant [5, 9].

$$S_{(x)} \dot{S}_{(x)} = S_{(x)} \left[\left(\upsilon - (f/j) - (1/j\tau) \right) \dot{x}_1 - (1/j\tau) x_1 \right] < 0 \tag{8}$$

Condition (8) is true if:

$$S_{(x)} \left(\upsilon - (f/j) - (1/j\tau) \right) \dot{x}_1 < 0 \text{ and } -S_{(x)} (1/j\tau) x_1 < 0 \tag{9}$$

From equation (9) we can write the expression of $\alpha_{1,2}$ and $\beta_{1,2}$ as follows:

$$\begin{array}{l} f \begin{cases} S_{(x)} x_1 > 0 \rightarrow \alpha_1 > 0 \\ S_{(x)} \dot{x}_1 > 0 \rightarrow \alpha_2 > \left(\upsilon - (f/j) \right) \cdot j\tau \end{cases} \\ if \begin{cases} S_{(x)} x_1 < 0 \rightarrow \beta_1 < 0 \\ S_{(x)} \dot{x}_1 < 0 \rightarrow \beta_2 < \left(\upsilon - (f/j) \right) \cdot j\tau \end{cases} \end{array} \tag{10}$$

Then

$$\begin{array}{c} \beta_1 < 0 < \alpha_1 \\ \beta_2 < \left(\upsilon - (f/j) \right) \cdot j\tau < \alpha_2 \end{array} \tag{11}$$

4 Sliding Mode Speed Observer

It is an observer of flux and current based on the sliding mode method. This observer has the advantage of not requiring input speed and rotor time constant unlike other observers. Thus, any variation of these quantities will not affect the estimation of current and flux [6, 9].

In addition, the use of sliding mode technical methods for the design of the observer ensures on the one hand the robustness with respect to various disturbances, and on the ohter the good dynamic performance.

The system of equation (1) can be described as follows:

$$\begin{aligned} \frac{d}{dt} i_s &= \Gamma\left(\Lambda\phi_r - A'_{11} i_s\right) - A''_{11} i_s + \beta_1 v_s \\ \frac{d}{dt}\phi_r &= -\left(\Lambda\phi_r - A'_{11} i_s\right) \end{aligned} \tag{12}$$

Where:

$$\Lambda = \begin{bmatrix} \frac{1}{\tau_r} & \omega_r \\ -\omega_r & \frac{1}{\tau_r} \end{bmatrix}, \quad \Gamma = \frac{M}{\sigma L_s L_r}$$

$$A'_{11} = \frac{M}{\tau_r}, \quad A''_{11} = \frac{1}{\sigma\tau_s}, \quad -\left(A'_{11} + A''_{11}\right) I = A_{11}$$

S matrix is defined by:

$$S = \left(\Lambda\phi_r - A'_{11} i_s\right) = \left(\begin{bmatrix} \frac{1}{\tau_r} & \omega_r \\ -\omega_r & \frac{1}{\tau_r} \end{bmatrix} \begin{bmatrix} \phi_{\alpha r} \\ \phi_{\beta r} \end{bmatrix} - \frac{M}{\tau_r} \begin{bmatrix} i_{\alpha s} \\ i_{\beta s} \end{bmatrix} \right) \tag{13}$$

We note that the matrix S appears simultaneously in the equations of flux and currents. This implies that the design of the observer for current and flux can be based on the replacement of the common term, which is the matrix S by the same sliding function ϑ:

$$\begin{bmatrix} \vartheta_1 \\ \vartheta_2 \end{bmatrix} = \hat{\vartheta} = \hat{S} = \begin{bmatrix} \frac{1}{\tau_r} & \hat{\omega}_r \\ -\hat{\omega}_r & \frac{1}{\tau_r} \end{bmatrix} \begin{bmatrix} \hat{\phi}_{\alpha r} \\ \hat{\phi}_{\beta r} \end{bmatrix} - \frac{M}{\tau_r} \begin{bmatrix} \hat{i}_{\alpha s} \\ \hat{i}_{\beta s} \end{bmatrix} \tag{14}$$

The observers of currents and flux can be written as follows:

$$\begin{aligned} \frac{d}{dt}\hat{i}_s &= \Gamma\hat{\vartheta} - A''_{11}\hat{i}_s + \beta_1 v_s \\ \frac{d}{dt}\hat{\phi}_r &= -\hat{\vartheta} \end{aligned} \tag{15}$$

Where:

$$\begin{aligned} \vartheta_1 &= -u_0 sign\left(S_\alpha\right) \\ \vartheta_2 &= -u_0 sign\left(S_\beta\right) \end{aligned} \tag{16}$$

$$\hat{i}_s = \begin{bmatrix} \hat{i}_{\alpha s} & \hat{i}_{\beta s} \end{bmatrix}^T, \quad \hat{\phi}_r = \begin{bmatrix} \hat{\phi}_{\alpha r} & \hat{\phi}_{\beta r} \end{bmatrix}^T$$

$$S_\alpha = \overline{i}_{\alpha s} = \hat{i}_{\alpha s} - i_{\alpha s}, \quad S_\beta = \overline{i}_{\beta s} = \hat{i}_{\beta s} - i_{\beta s}$$

With

$$sign(S_\alpha) = \begin{cases} 1 & if\ S_\alpha > 0 \\ -1 & if\ S_\alpha < 0 \end{cases}$$

$$sign(S_\beta) = \begin{cases} 1 & if\ S_\beta > 0 \\ -1 & if\ S_\beta < 0 \end{cases}$$

$\hat{i}_{\alpha s}, \hat{i}_{\beta s}$ and $i_{\alpha s}, i_{\beta s}$ are respectively the components observed and measured stator current. Flux estimation is a simple integration of sliding mode functions when the estimated current converges to the measured current without the need to know the speed or the rotor time constant. The selection of u_0 will guarantee the convergence of the current observation by Lyapunov stability analysis.

It is worth noting that we have supposed that the equivalent control of sliding mode observer is achieved by a simple low-pass filtering the discontinuous control [9].

$$\vartheta_{1,2}^{eq} = \frac{1}{\mu s + 1} \vartheta_{1,2} \tag{17}$$

μ is the time constant of the low pass filter.

Now the rotor flux can be estimated by the following equation:

$$\begin{bmatrix} \dot{\hat{\phi}}_{\alpha r} \\ \dot{\hat{\phi}}_{\beta r} \end{bmatrix} = -\begin{bmatrix} \vartheta_1^{eq} \\ \vartheta_2^{eq} \end{bmatrix} \tag{18}$$

The angle of the rotor flux can be calculated as follows:

$$\hat{\theta} = \tan^{-1}\left(\hat{\phi}_{\beta r} / \hat{\phi}_{\alpha r} \right) \tag{19}$$

Using equation (14), we can calculate the estimate rotor speed as follows:

$$\hat{\omega}_r = \frac{1}{\left|\hat{\phi}_r\right|} \left(\hat{\phi}_{\beta r} \left(\vartheta_1^{eq} + \frac{M}{\tau_r} \hat{i}_{\alpha s} \right) - \hat{\phi}_{\alpha r} \left(\vartheta_2^{eq} + \frac{M}{\tau_r} \hat{i}_{Bs} \right) \right) \tag{20}$$

The block diagram of the sliding mode observer of the currents, flux and speed is given in the figure below:

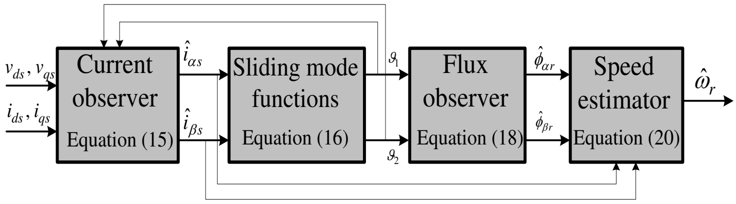


Figure 1
Schematic diagram of sliding mode observer

The stability of the observer can be proved by the Lyapunov stability theory. The Lyapunov function is chosen as follows [6]:

$$L = \frac{1}{2} I_n^T I_n \tag{21}$$

Where

$$I_n = \begin{bmatrix} \bar{i}_{\alpha s} & \bar{i}_{\beta s} \end{bmatrix}$$

The Lyapunov function V is positive definite, which satisfies the Lyapunov stability. The derivative of L is:

$$\dot{L} = I_n^T \dot{I}_n \tag{22}$$

To satisfy the Lyapunov stability, the second condition must satisfy $\dot{L} < 0$. From Equation (12) and (15), we have:

$$\dot{\bar{I}}_n = \Gamma\left(\vartheta - \Lambda\phi_r\right) - A_{11}\bar{I} \tag{23}$$

Equations (22) become:

$$\dot{L} = \bar{I}^T \Gamma\left(\vartheta - \Lambda\phi_r\right) - A_{11}\bar{I}^T\bar{I} < 0 \tag{24}$$

Note that: $\vartheta = -u_0 sign(\bar{I})$

We can find the expression of constant u_0:

$$u_0 > \frac{\left|\bar{I}^T \Lambda\phi_r\right| - \frac{A_{11}}{\Gamma}\bar{I}^T\bar{I}}{\left|\bar{i}_{\alpha s}\right| + \left|\bar{i}_{\beta s}\right|} \tag{25}$$

With:

$$\left|\bar{I}^T \Lambda\phi_r\right| = \bar{i}_{\alpha s}\left(\frac{\phi_{\alpha r}}{\tau_r} + \omega_r\phi_{\beta r}\right) + \bar{i}_{\beta s}\left(\frac{\phi_{\beta r}}{\tau_r} + \omega_r\phi_{\alpha r}\right)$$

$$A_{11}\bar{I}^T\bar{I} = A_{11}\left(\bar{i}_{\alpha s}^2 + \bar{i}_{\beta s}^2\right)$$

5 Fuzzy-Sliding Mode Speed Observer

In this section, fuzzy logic and sliding mode are combined together to give us a new concept of observer. The observer thus obtained is called fuzzy sliding mode speed observer. It presents the same structure as the SMSO given in Section 4, but the sliding function ϑ will be replaced by a fuzzy controller. So after the calculation of the sliding function ϑ, we can use the Equation 15 to estimate the currents and rotor flux in oder to estimate the rotor speed using the following equation:

$$\hat{\omega}_r = \frac{1}{\left|\hat{\phi}_r\right|}\left(\hat{\phi}_{\beta r}\left(\vartheta_1 + \frac{M}{\tau_r}\hat{i}_{\alpha s}\right) - \hat{\phi}_{\alpha r}\left(\vartheta_2 + \frac{M}{\tau_r}\hat{i}_{Bs}\right)\right) \tag{26}$$

The idea of this combination is inspired by the fact that in the ideal case when $S_{(\alpha,\ \beta)}$ is far from the sliding surface, ϑ must have a very large value and when $S_{(\alpha,\ \beta)}$ is near the sliding surface ϑ must have a small value.

The block diagram of the fuzzy logic calculation of ϑ used in our simulation is given in the figure below:

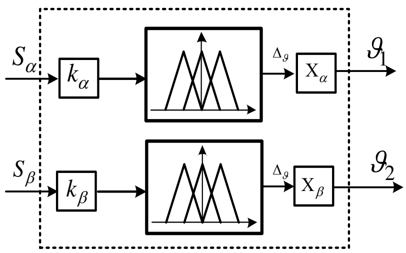


Figure 2

Block diagram of the adaptation mechanism of the sliding function ϑ using fuzzy logic

When choosing the linguistic value, it should be taken into account that the control must be robust and time of calculation adopted by the fuzzy controller should not be high to not slow down the process [11-13]. In this proposed linguistic method we have used 5 rules.

For fuzzyfication, we have chosen triangular fuzzyfication and for deffuzzyfication the centroid deffuzzyfication method is used in the proposed method.

All fuzzy variables have the same universe of discourse (S_α, S_β, Δ_ϑ) and are divided into five fuzzy sets (NB, NS, ZE, PS, PB) for the input variables and (PS, P, ZE, PM, PB). Membership functions are chosen in the form triangular as in Figures 3 and 4.

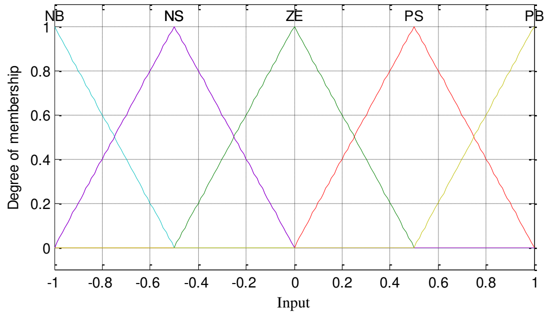


Figure 3
Degree of input membership

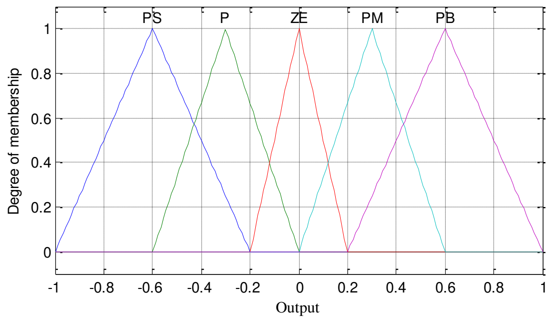


Figure 4
Degree of output membership

In terms of numerical values, the behavior of this mechanism is characterized by action law shown in Figure 5. Indeed, for each input values the mechanism generates a variation of Δ_ϑ which corresponds to the increase or decrease of sliding function ϑ.

Rule 1: If $S_{(\alpha,\beta)}$ is *NB* **So** Δ_ϑ *is PB.*

Rule 2: If $S_{(\alpha,\beta)}$ is *NS* **So** Δ_ϑ *is PM.*

Rule 3: If $S_{(\alpha,\beta)}$ is *ZE* **So** Δ_ϑ *is ZE.*

Rule 4: If $S_{(\alpha,\beta)}$ is *PS* **So** Δ_ϑ *is P.*

Rule 5: If $S_{(\alpha,\beta)}$ is *PB* **So** Δ_ϑ *is PS.*

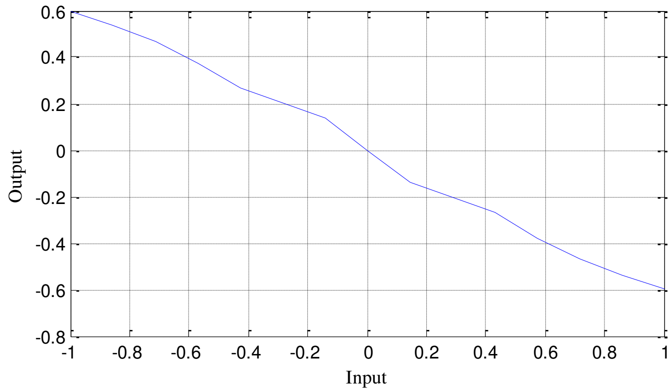


Figure 5

Variation law of fuzzy controller for sliding function ϑ adaptation

5 Simulation Results and Discussion

To assess the proposed algorithm, the solution has been checked using Matlab – Simulink in order to validate the fuzzy sliding mode speed observer for estimating the rotor speed. The parameters of the three phase induction motor are given in Table 1.

Table 1

Parameters of induction motor

Designation	Notations	Rating values
Stator resistance	R_s	2.3 Ω
Rotor resistance	R_r	1.83 Ω
Stator self-inductance	Ls	261 mH
Rotor self-inductance	L_r	261 mH
Mutual inductance	M	245 mH
Moment of inertia	j	0.03 kgm^2
Friction coefficient	f	0.002 Nm
Number of poles	p	2
Rated voltage	V_{sn}	220 V

Figure 6 shows the architecture of the vector control algorithm incorporating the fuzzy sliding mode speed observer.

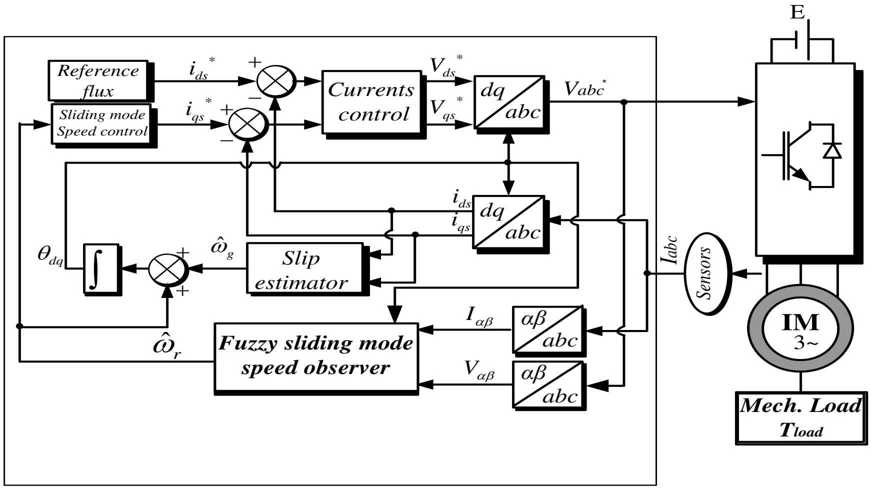


Figure 6

Block diagram of the vector control including speed fuzzy sliding mode observer

Figures 7 and 8 show the test of tracking speed for the machine by considering the nominal conditions. These figures represent the dynamic responses of the electromagnetic torque and the speed. The performance is checked in terms of load torque and speed variations. Very good tracking speed is obtained.

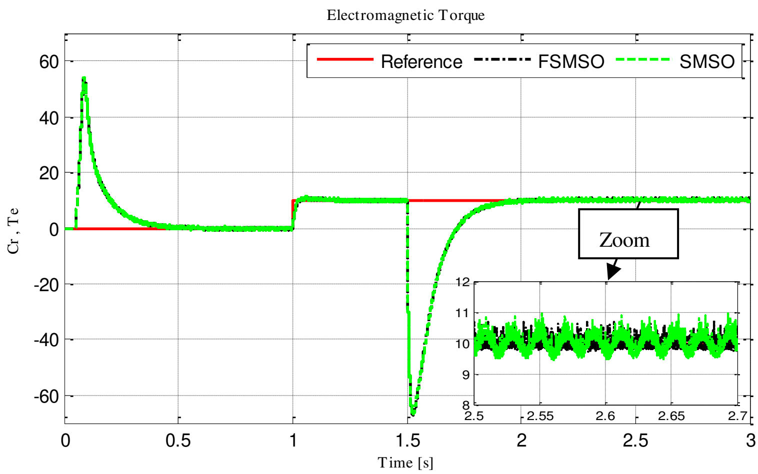


Figure 7

Torque response for step varying of load torque (FSMSO and SMSO)

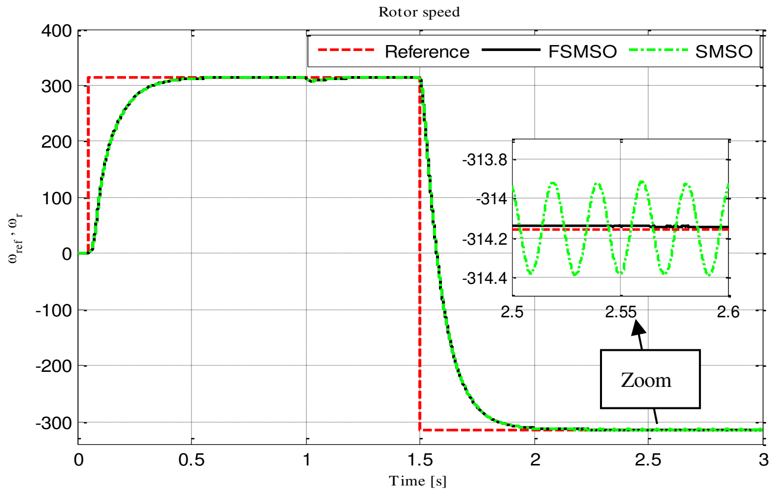


Figure 8

Estimated speed using FSMSO and SMSO

Figures 7 and 8 show the simulation results to a step speed (314rd/s) under a load torque of 10 Nm. The estimated speed by the sliding mode speed observer and the fuzzy sliding mode speed observer are nearly similar (Figure 8). We can see the effect of the chattering phenomenon on the speed estimated reponse, this is due to the nature of the sliding functions. Figure 7 shows the electromagnetic torque and the effect of chattering phenomenon developed during the starting and the stationary phase.

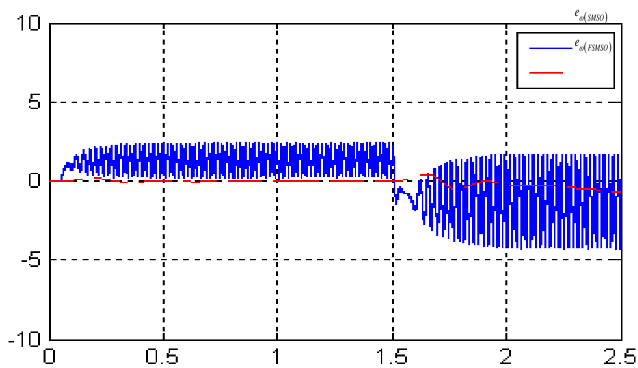


Figure 9

Speed estimation error for the SMSO and FSMSO

Figure 9 shows the speed estimation error for both types of observers. It is clear that the estimation error for FSMSO is smaller than the error for SMSO which is caused by the sliding mode functions.

Results obtained show clearly the superiority of FSMSO compared to SMSO for the speed estimation and minimization of the ripple effect (chattering phenomenon).

Conclusions

A sliding mode control and speed estimation algorithm based on current and flux observers are proposed in this paper. In this algorithm fuzzy sliding mode functions are selected to control and determine rotor speed that is assumed to be unknown. The proposed scheme validated by simulation results shows the superiority of FSMSO compared to SMSO for the speed estimation and reducing of the chattering phenomenon. This algorithm can be improved further by using an adaptive approach for the rotor and stator resistance variations, and will be implemented in the future work on a digital processor (DSP) to validate the proposed scheme.

References

[1] Ben Regaya, C., Farhani, F., Zaafouri, A., Chaari, A. "Comparison between Two Methods for Adjusting the Rotor Resistance", *International Review on Modelling & Simulations (I.RE.MO.S.)* 2012, Vol. 5, N. 2, pp. 938-944

[2] Derdiyok, A. "A Novel Speed Estimation Algorithm for Induction Machines", *Electric Power Systems Research*, 2003, Vol. 64, N. 1, pp. 73-80

[3] Farhani, F., Ben Regaya, C., Zaafouri, A., Chaari, A. "On-Line Efficiency Optimization of Induction Machine Drive with Both Rotor and Stator Resistances Estimation", *International Review on Modelling & Simulations (I.RE.MO.S.)* 2012, Vol. 5, N. 2, pp. 930-9937

[4] Cheng-Hung, T. "CMAC-based Speed Estimation Method for Sensorless Vector Control of Induction Motor Drive", *Electric Power Components and Systems*, 2006, Vol 34, N. 11, pp. 1213-1230

[5] Zhao, Junhui, Caisheng Wang, Feng Lin, Le Yi Wang, and Zhong Chen. "Novel Integration Sliding Mode Speed Controller for Vector Controlled Induction Machines", *IEEE Power and Energy Society General Meeting*, 2011, pp. 1-6

[6] G. R. Arab Markadeh. "A Current-based Output Feedback Mode Control for Speed Sensorless Induction Machine Drive using Adaptive Sliding Mode Flux Observer", *The Fifth International Conference on Power Electronics and Drive Systems*, 2003, PEDS 2003, Vol. 1, pp. 226-231

[7] Karim Khémiri. "Novel Optimal Recursive Filter for State and Fault Estimation of Linear Stochastic Systems with Unknown Disturbances", *International Journal of Applied Mathematics and Computer Science*, 2011, Vol. 21, N. 4, pp. 629-637

[8] Tichy, Michael. "On Algebraic Aggregation Methods in Additive Preconditioning", *Department of Mathematics, University of Würzburg*, 2011

[9] Mezouar, A. "Adaptive Sliding-Mode-Observer for Sensorless Induction Motor Drive using Two-Time-Scale Approach", *Simulation Modelling Practice and Theory*, 2008, Vol. 16, N. 9, pp. 1323-1336

[10] Rehman, H., Gilven, M. K., Derdiyok, A., Longya Xu. "A New Current Model Flux Observer Insensitive to Rotor Time Constant and Rotor Speed for DFO Control of Induction Machine", *Power Electronics Specialists Conference. PESC. 2001 IEEE 32nd Annual*, 2001, Vol. 2, pp. 1179-1184

[11] Kouzi, K., Nait-Said, M. S., Hilairet, M., Bertholt, E. "A Robust Fuzzy Speed Estimation for Vector Control of an Induction Motor", *6th International Multi-Conference on Systems Signals and Devices*, 2009, SSD `09, pp. 1-6

[12] Khan, Md. Haseeb. "Fuzzy Logic-based HPWM-MRAS Speed Observer for Sensorless Control of Induction Motor Drive", *Journal of Engineering & Applied Sciences,* 2010, Vol. 5, N. 4, pp. 86-93

[13] wei-Ling Chiang. "Adaptive Fuzzy Sliding Mode Control for Base-isolated Buildings" *International Journal on Artificial Intelligence Tools*, 2000, Vol. 9, N. 4, pp. 493-503

Conventional PID and Modified PID Controller Design for Roll Fin Electro-Hydraulic Actuator

Fuat Alarçin[1], **Hakan Demirel**[2], **M. Ertugrul Su**[1], **Ahmet Yurtseven**[1]

[1] Faculty of Naval Architecture and Maritime, Yıldız Technical University, Barbaros Bulvari 34349 Beşiktaş İstanbul, Turkey

E-mail: alarcin@yildiz.edu.tr, mesu@yildiz.edu.tr, ahmety@yildiz.edu.tr

[2] Department of Marine Engineering Operations, Bulent Ecevit University, 67300 Ereğli/Zonguldak, Turkey

E-mail: hakandemirel@beun.edu.tr

Abstract: The aim of this study is to decrease a fishing boat's roll motion, which is adversely affected by disturbing hydrodynamic forces, by applying fin roll stabilizer. It is ensured that roll motion with nonlinear damping and restoring moment coefficients reaches the desired level by means of classical PID and modified PID algorithms. By Lyapunov's direct method, the key issue- stability which is needed during the course of sailing was examined considering initial conditions, and it was realized that the system was generally stable. Besides, NACA 0015 model was utilized for the fin roll stabilizer, and flow analysis was conducted by CFD method. According to the simulation results, when the same gains were practiced, modified PID controller algorithms were rather more efficient than conventional PID in the roll fin stabilizer system.

Keywords: modified PID control; fin stabilizer; Lyapunov direct method

1 Introduction

For many years, vessels have remained as a research subject for researchers as they lack fundamentals of stability despite meeting the requirements of current laws. It is known that a ship on route is likely to be off the stable zone due to disrupting hydrodynamic forces of a ship roll motion. Grim [1] modelled the roll equation using the gradual change in the restoring moment among the waves. In another study of his, he analyzed unstable roll equation in a more general equation form using Mathieu equation.

Advanced studies on roll motion were carried out by Dalzell [2], and alternative damping models were put forward. Soliman and Thomson [3] solved the nonlinear differential equation of roll motion through 4^{th} level Runga-Kutta method. How damping effect of roll motion stability changed owing to disrupting forces and frequency was studied. Haddara and Wang [4] pointed out that quadratic damping model of speed for small fishing ships came up with good results. Taylan [5] studied a nonlinear roll motion model and the solution of this model through generalized Krylov-Bogoliubov asymptotic method on a time line. In this model, nonlinear restoring terms were considered as a 3^{rd} level polonium; likewise, nonlinear damping was regarded as a 2^{nd} level term. Taylan [6] studied on generalization of the nonlinear equation form indicating roll motion of a ship sailing amidst waves by means of Duffing method on frequency level. Surendran and Reddy [7] discussed the roll dynamics of a Ro-Ro ship taking into account many types of combination of loads in linear and nonlinear forms.

Safety of voyage has to be assured against the disrupting hydrodynamic effects of passengers as well as cargos, and roll amplitude has to be at an acceptable level. Therefore, a number of applications such as fin roll stabilizers and U-tube have been used in literature [8, 9].

Sgoppo and Persons [10] demonstrated that roll motion can be reduced by 35% through fin roll stabilizer system based on WMEC 901 class ship with 3^{rd} degree of freedom model. Surendran et al. [11] pointed out that nonlinear roll motion could be minimized with the fin roll stabilizer based on PID (Proportional-Integral-Derivative) control algorithm considering a frigate warship. In this study, he acquired lift characteristics of the fin roll stabilizer stemming from hydrodynamic flow using CFD (Computational Fluid Dynamics). In simulations for different occasions, an 80% reduction in roll value was observed.

Guan and Zhang [12], who expressed nonlinear damping terms with the Backstepping Integrator theory, controlled the nonlinear fin roll stabilizer by means of PID and close loop algorithms by ignoring high level nonlinear terms. Ghassemiet al. [13] expressed that fin roll stabilizers are effective components in reducing roll amplitude out of wave movements. It was pointed out that lift forces occurring around the angle of attack and the fin roll stabilizer create a reversed moment. That’s why he showed roll amplitude could be reduced using PID and Neural Network combination to control attack angle. Karakaş et al. [14] designed a controller based on Lyapunov method for fin roll damping systems in beam seas. The roll motion of the ship was considered as a single degree of freedom and 3^{rd} degree nonlinear terms were added to the model. It was mentioned that the likelihood of capsize was reduced to a considerable extent by means of the controller based on Lyapunov method. PID controllers’ versions have been investigated by many researchers Alfaro et al. [24], Precup et al. [25], Hadlovska and Jajcisin [26].

In this paper, the problem of a nonlinear roll motion of fishing vessel is considered. Section 2 deals with mathematical model of nonlinear roll motion. Section 3 analyses stability of roll motion by Lyapunov's Direct Method. Section 4 describes the fin roll stabilizer system and classical PID and the modified PID controller. Nonlinear roll motion of a fishing boat, which will be capable of operating in the Black Sea and the Mediterranean Sea, was controlled by fin stabilizer system. Section 5 discusses simulation results.

2 Mathematical Model of Nonlinear Roll Motion

In order to ease to design mathematical model of ship motions, some important assumptions are made in modelling a fishing boat rolling motion, neglecting all other degrees of freedom of ships, xz-plane symmetry, rigid body and homogenous mass distribution. Considering some simplifications, the following nonlinear expression for the roll equation is obtained,

$$(I+J)\,\ddot{\varphi}+B_1\dot{\varphi}+B_2\,\dot{\varphi}\,|\dot{\varphi}|+\Delta(\,c_1\varphi+c_3\,\varphi^3+c_5\,\varphi^5\,)=\omega_e^2\,\alpha_m\,I\cos(\,\omega_e t\,)-M_f \tag{1}$$

where φ, $\dot{\varphi}$, $\ddot{\varphi}$ are angle, angular velocity and angular acceleration of roll motion, respectively. I and J are the mass moment of inertia and the added mass moment of inertia, respectively. B_1, B_2 are roll damping coefficients, c_1, c_2, c_3 are determined by restoring force coefficients and Δ is the weight displacement of the ship, ω_e is encounter frequency of the wave, α_m is the maximum wave slope, M_f is the control moment of active fins. Dividing the equation (1) throughout by $(I+J)$,

$$\ddot{\varphi}+b_1\,\dot{\varphi}+b_2\dot{\varphi}\;|\dot{\varphi}|+\Delta(c_1\,\varphi+c_3\,\varphi^3+c_5\,\varphi^5\,)=\frac{\omega_e^2\,\alpha_m\,I\cos(\,\omega_e t\,)}{I+J}-\frac{M_F}{I+J} \tag{2}$$

Inertia moment arises due to the ship's reaction to a movement and it is proportional to the acceleration of motion. Added inertia moment is a reaction of sea water to the ship motion. These inertia values can be calculated depending on weight displacement of ship (Δ), breadth (*B*) and the vertical distance of the center of gravity (*KG*),

$$(I+J)=\frac{\Delta}{12\,g}(\,B^2+4\,KG^2\,) \tag{3}$$

The roll damping coefficient for a ship hull form has several contributions. These components are considered as skin friction of the hull, eddy shedding from the hull, free surface waves, lift effect damping and bilge keel damping. Theoretical and semi-empirical methods have been used to evaluate the roll damping by Ikeda et al. [15] and Ikeda [16]. A non-dimensional damping coefficient for different ship types is expressed as follows:

$$B_1 = \frac{2a\sqrt{(I+J)\Delta GM}}{\pi} \tag{4}$$

$$B_2 = \frac{3}{4}b(I+J) \tag{5}$$

These coefficients are directly related to a linear damping coefficient B_1 and a non-linear damping coefficient, B_2 represents quadratic drag [2]. The non-dimensional damping coefficients for fishing boat were obtained as follow (*a=0.1, b=0.0140*). The curve for righting arm has been represented by the polynomial,

$$M(\varphi) = c_1\varphi + c_3\varphi^3 + c_5\varphi^5 + \ldots$$

where $c_1 > 0,\ c_3 < 0,\ c_5 > 0$. The roll restoring moment coefficients are defined by ref. [6]

$$c_1 = \frac{d(GZ)}{d\phi} = GM \quad c_3 = \frac{4}{\phi_v^4}(3A_{\phi v} - GM\phi_v^2) \quad c_5 = -\frac{3}{\phi_v^6}(4A_{\phi v} - GM\phi_v^2) \tag{6}$$

Angle of vanishing stability ϕ_v, area under the *GZ* curve $A_{\phi v}$, and dynamic characteristics of the *GZ* curve such as metacentric height *GM*. Based on the above-mentioned coefficients, numerical calculations were carried out for a fishing boat, whose body plan is given in Fig. 1.

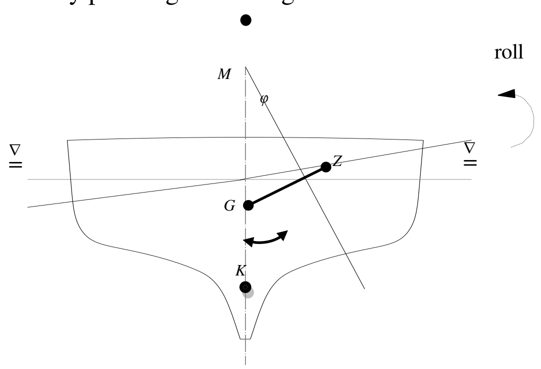


Figure 1
Body plan of the fishing boat

The righting arm curve is a graphical representation of the fishing boat’s stability in Fig. 2 [17]. From this figure, it can be seen that Hydrodynamic lift effect develops in an approximately linear manner with an increasing angle of attack.

The area under the curve is an indication of the fishing boat's ability to counter the capsizing moments acting on the boat.

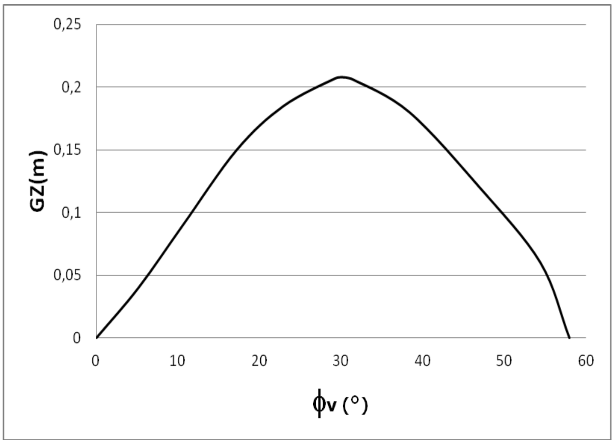


Figure 2
Righting arm curve of fishing boat

Due to the influences of high wave frequency, ships sailing on the sea produce undesirable roll motion. It is expressed the wave moment to consist of the encounter frequency as,

$$M_w = \omega_e^2\, \alpha_m\, I\, cos(\, \omega_e\, t\,) \tag{7}$$

$$\omega_e = \omega_w - \frac{\omega_w^2}{g} V\, cos(\, \mu_w\,)$$

M_w is the wave moment, ω_e is the encounter frequency of the wave, ω_w is the wave frequency, α_m is the maximum wave slope, μ_w is the wave encounter angle of the ship. It can be envisaged that the wave excitation will depend not only on amplitude and frequency of the waves but also on encounter angle and speed.

3 Stability Analysis of Roll Motion by Lyapunov's Direct Method

A fishing boat is called stable when it has enough positive stability to counter the external forces generated by current weather, fishing conditions and it will return to its upright position [18]. Lyapunov's Direct Method was used for stability analyses by Ozkan [19]. This is a very powerful and applicable technique since it does not require any knowledge about the explicit solutions of the equations. By using state variables of equation (1),

$$\dot{\varphi} = \dot{\varphi}_1 = \varphi_2$$

$$\dot{\varphi}_2 = -(\omega_0^2 \varphi_1 + m_3 \varphi_1^3 + m_5 \varphi_1^5) - (b_1 \varphi_2 - b_2 \varphi_2 |\varphi_2|) \tag{8}$$

Lyapunov function V_x satisfying

- $V_x > 0$ positive definite and $V(0)=0$
- $\frac{dV(\vec{x})}{dt} \leq 0$
- $V(x) \to \infty$ as $\|x\| \to \infty$

Lyapunov second method will be used to test the system stability.

$$\dot{V}(t) = \frac{dV(\vec{x})}{dt} = \nabla V^T \dot{\vec{x}}$$

If symmetric coefficients are assumed equal to zero, derivative of the Lyapunov function is negative. Lyapunov function is obtained depending on the non-linear roll damping coefficient. If this value is smaller than zero, non-linear roll motion can be said to be stable.

$$\dot{V}(x) = -\varphi_2^2 (b_1 + b_2 |\varphi_2|) < 0 \tag{9}$$

$$V(x) = \omega_0^2 \frac{\varphi_1^2}{2} + m_3 \frac{\varphi_1^4}{4} + m_5 \frac{\varphi_1^6}{6} \tag{10}$$

4 Controller Design for Roll Fin Actuator

The objective of the control is to generate the input current such that the angular position of the control fin is regulated to the desired position. The motion of a ship can be affected by fin actuators that impart forces and moments. Actuators play a very important role in the control system structure. When the roll fin stabilizers attack to the fluid, it can bee seen that the lifting force caused by the rotation and the angle of attack occurs on the surface of fins. The lift force and the lift in non-dimensional form are as in the following form in Fig. 3 [8],

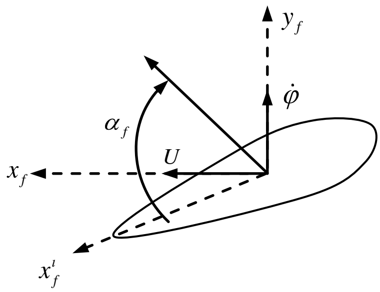


Figure 3
Ship roll fin stabilizer

$$L = \frac{1}{2}\rho V A_F C_L \tag{11}$$

$$C_L = \frac{L}{0.5\,\rho V^2 A_F} \tag{12}$$

where L lifting force (N); ρ density of fluid (t/m^3); A_F fins area (m^2); C_L fins lift coefficient (lift coefficient/rad); V the ship speed (m/s). General formulas of fin roll stabilizer are expressed as the following equations:

$$M_F = \rho V^2 A_F C_L I_F \left(\alpha_F + \frac{\dot{\varphi}}{v} I_F \right) \tag{13}$$

where M_F, fin roll stabilizer moment; I_F the fins force arm; α_F attack angle. The electro-hydraulic system dynamics of fin stabilizer system are assumed to be governed by

$$\dot{\alpha}_F + t_1 \alpha_F = t_2 u_F \tag{14}$$

where α_F is the actuator output (actual fin angle), u is the input to the electro-hydraulic systems. The hydraulic control model presented in Fig. 4. Conrol surfaces are commanded by hydraulic machinery that implement the action demanded by controller.

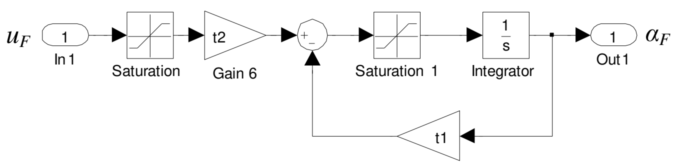


Figure 4
Block diagram of fin stabilizer system

As can be seen in Fig. 4, a saturator has been intermitted between the controller and the fin actuator.

This element models the saturating behaviour of the controller. The foil motion is constrained to move within certain maximum angle.

$$-\alpha_{F\,max} \leq \alpha_F \leq \alpha_{F\,max}$$

The results of hydrodynamic lift coefficient in function of attack angle of fin and roll angle of ship are presented in Fig. 5.

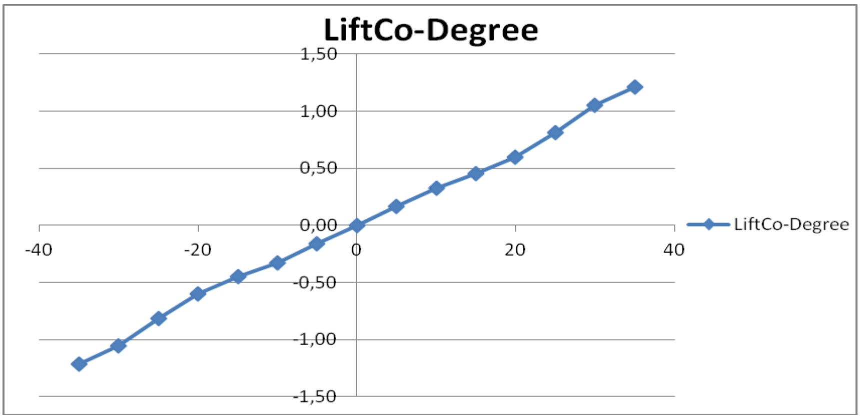


Figure 5
Lift coefficient

Computational fluid dynamics analysis has been carried out with Star CCM + package software, via 8 Parallel Processor and 24 GB of RAM hardware configuration. Preprocessing, running and finishing parts were integrated in the program. At the time of Preprocessing, Trimmer, Surface Remesher and Prism Layer Masher solution mesh properties were executed simultaneously. Hexahedral meshes analyses were conducted. Variations with 170000, 970000 and 1500000 cells were generated respectively for our three dimensional model and afterwards, the solution continued with the one having 970000 cells 2900000 surface via mesh independence. A sample case of free surface shape around fin is presented in Fig. 6.

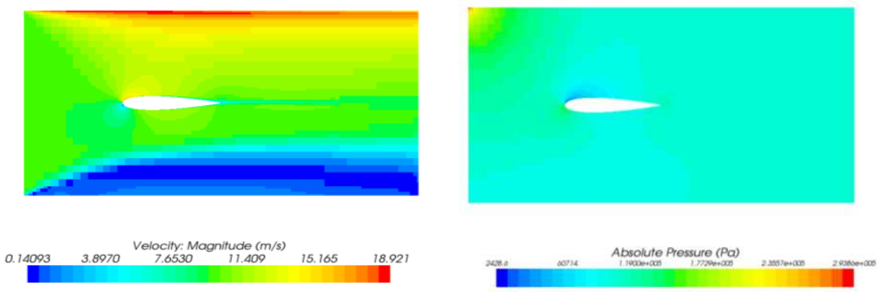


Figure 6
Velocity and pressure gradient cross section (attack angle $+30^o$)

We considered only a magnitude constraint for the mechanical angle of the fins as *30°*. Segregated flow, Reynolds averaged Navier Stokes and k-ε turbulence model were used as the solvers for running process. Convergence conditions were approximately *1E-5* (10^{-5}). Running was carried out in parallel with 8 cores. During the final process, along with the velocity and pressure gradients, the lift force coefficient was directly taken from the model. This coefficient was obtained depending on fin roll stabilizer's attack angle (from -30^o to $+30^o$).

4.1 Conventional PID Controller

The parameters of PID controller lead to different effects on system characteristics. The proportional block provides an overall control action, and the integral block reduces steady-state errors. On the other hand, the derivative block improves transient response. The basic elements of a PID controller for a ship roll motion control system are shown in Fig. 7 [20].

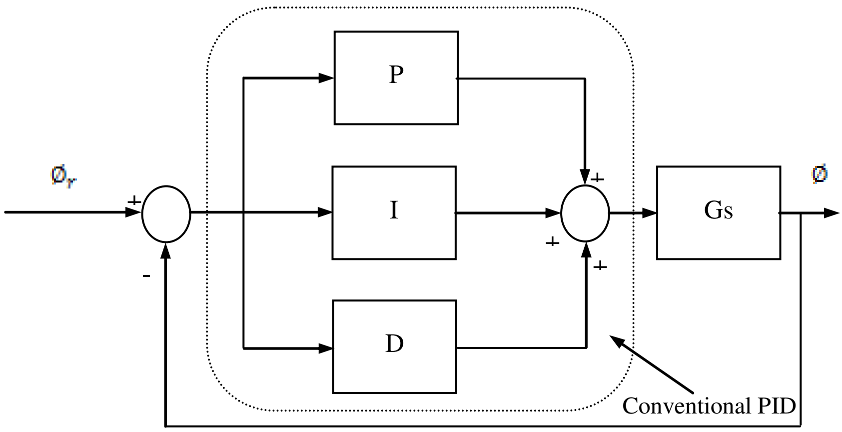


Figure 7

Closed loop system with conventional PID controller

Generally, transfer function of classical PID controller is given below:

$$G_{pid}(t) = K_p\left[e + \frac{1}{\tau_i}\int d(t) + \tau_d \frac{de(t)}{dt}\right] \tag{15}$$

Where $\tau_i = \dfrac{K_p}{K_i}$, $\tau_d = \dfrac{K_d}{K_p}$ and e is the error between the reference and the output of a system, τ_i is the integral time, τ_d is derivative time. K_p, K_i and K_d are proportional, integral and derivative gains, respectively. The conventional PID controller is the simplest form of controllers that utilize the derivative and integration operations in the compensation of control systems. Thanks to its flexibility, it is easier to use this controller in many applications and many control problems [21, 22, 23, 24, 25, 26, 27 and 28]. The model of closed loop system with modified PID is shown in Fig. 8.

4.2 Modified PID Controller

The position of integral action which affects the difference of reference signal and feedback signal has remained unchanged on forward path. However, derivative and proportional actions have moved on feedback path to affect only the output signal. The objective is to force the system's output to follow a given bounded reference value.

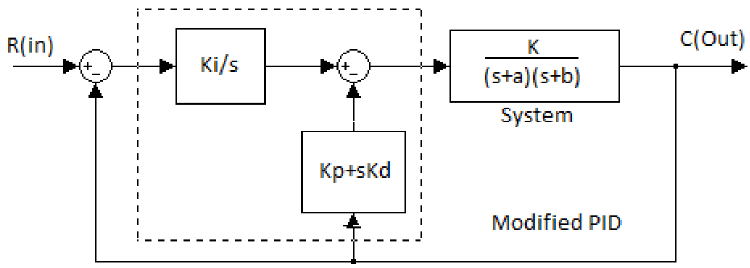


Figure 8

The model of closed loop system with modified PID controller

The modified PID controller can be indicated as follows:

$$G_c(s) = \frac{K_i - sK_p - s^2 K_d}{s} \tag{16}$$

Transfer function of closed loop system is given below:

$$\frac{C(s)}{R(s)} = \frac{K_i K}{s^3 + s^2(a+b) + s(ab + K_i K_d K) + K_i K_p K} \tag{17}$$

There are two zeros in the system with conventional PID controller. It is hard to adjust the system response due to these zeros. Their effect occurs as earlier peak or higher overshoot. The proposed modified PID controller gets over these effects and ameliorates the system response by adding proportional and derivative blocks of PID on feedback path instead of on forward path. Therefore, better solution of system response is achieved in modified PID in comparison with conventional PID.

5 Simulation

In control application, mathematical model was presented to perform numerical simulations of different scenarios, and to obtain the performance of the fin stabilizer system. These numerical simulations were applied with Matlab simulink program. The fishing boat and the fin (NACA 0015) particulars are shown in Table 1.

Table 1
The fishing boat and the fin (NACA 0015) particulars

Principal Particulars	Symbol	Parameter
Length between perpendiculars	L_{bp}	*20 m*
Breadth	B	*5.714 m*
Depth	D	*3.2 m*
Draught	T	*2.285 m*
Displacement	∇	*119.34* m^3
Metacentric height	GM	*0.57 m*
Vertical center of gravity	KG	*2.4 m*
Block coefficient	C_B	*0.457*
Service speed	V	*10 knot*
Fins area	A_F	*2.5* m^2
Fins lift coefficient	C_L	*0.65*
Vanishing angle of stability	ϕ_v	58^0

Non-dimensional damping and restoring moment coefficients for fishing boat are expressed in Table 2.

Table 2
Non-dimensional damping and restoring moment coefficients for fishing boat

b_1	b_2	m_1	m_3	m_5
0.069	0.010	1.204	-1.80	0.61

The flow chart of MPID control system is displayed in Fig. 8. The simulation results for fin roll stabilizer system show roll angle, roll velocity and phase diagram in Figs. 9, 10 and 11. Comparisons of the control performance were made between conventional PID and the MPID controllers. K_p, K_i and K_d control values were obtained by trial method. The values of PID gains K_p = 0.2145, K_d = 1.2288, K_i =2.89, ensured good roll reduction. The modified PID control response of the fin roll stabilizer is better than PID control.

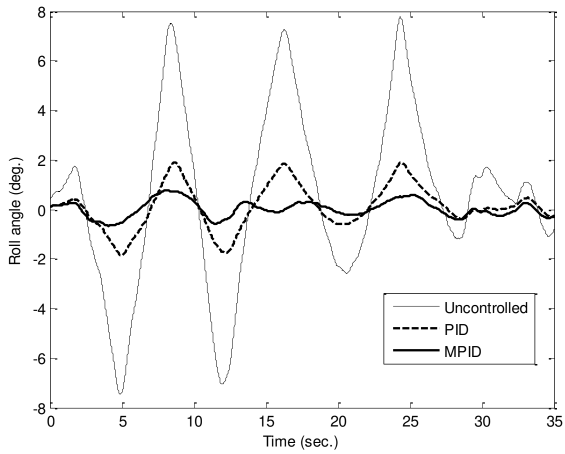


Figure 9
Comparison of roll angle response

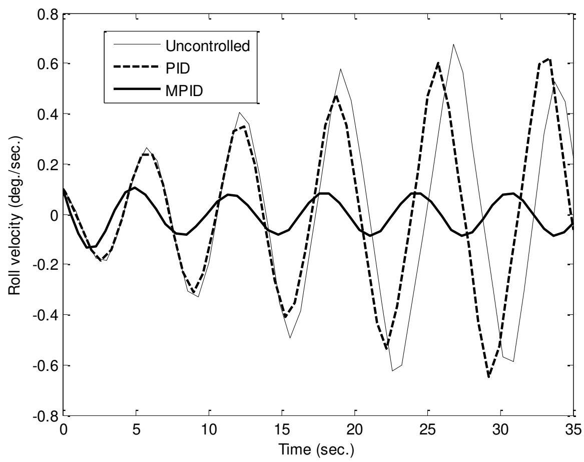


Figure 10
Comparison of roll velocity response

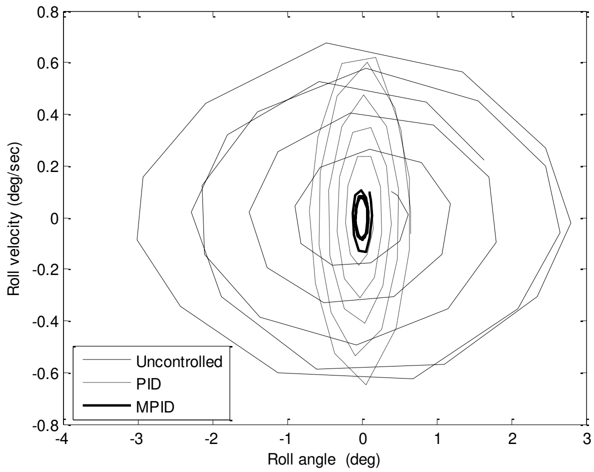


Figure 11

Comparison of phase diagram

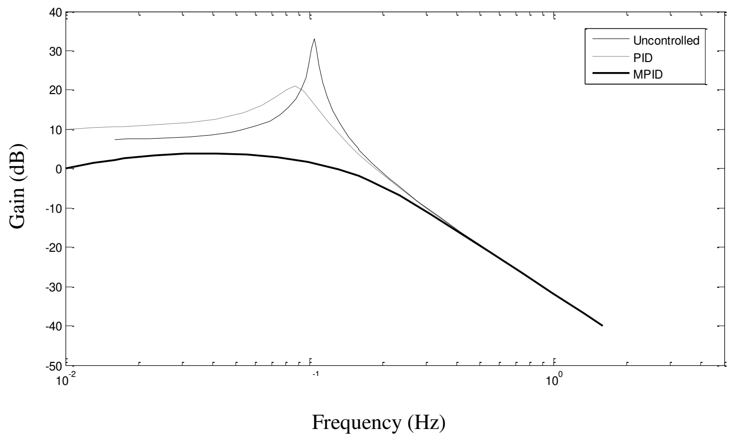


Figure 12

Frequency response of the roll motion for uncontrolled, PID and MPID

Since the aim of the controller was to control the roll motion, the improvement has been realized particularly around 1.1 Hz at low frequencies. This is also demonstrated by plotting the frequency response of controlled and uncontrolled body bounce acceleration in Fig. 12.

The comparison of the two controllers is presented in Table 3, which shows roll angle and roll velocity. The settling time for uncontrolled is longer compared to the time for settlement of MPID.

Table 3
Performance Comparisons

Controller	Max. roll angle (deg.)	Max. roll velocity (deg./s)
Uncontrolled	7.8^0	*0.7*
PID	1.9^0	*0.6*
MPID	0.7^0	*0.1*

Conclusions

This paper introduces a mathematical modeling, and examines controlling of nonlinear roll motion by applying fin stabilizer system. Nonlinear restoring and damping moment coefficients are computed by means of empirical equations. Nonlinear roll motion stability of fishing boat is examined through Lyapunov direct method. In the course of the simulation, it is accepted that the control gains receive the same values for PID and MPID algorithms. According to the results of the simulation, it is notable that MPID controller has shown a considerable amelioration in roll magnitude about *91%.* The performance of MPID controller performance has been *15.4%* greater than conventional PID as shown in Table 3.

References

[1] O. Grim: schwingungen, Stabilitat und Sicherheit in Seegang, Schiffstechnik (1952) pp. 85-100

[2] J. F. Dalzell: A Note on the Form of Ship Roll Damping, Journal of Ship Research, Vol. 22(3) (1978) pp. 178-185

[3] M. Soliman and J. M. T. Thompson: Transient and Steady State Analysis of Capsize Phenomena, Applied Ocean Research, 13 (1991) pp. 82-92

[4] M. Haddara and Y. Wang: Parametric Identification of Maneuvering Models for Ships, Int. Shipbuild. Programs, 445 (1999) pp. 5-27

[5] M. Taylan: Solution of the Nonlinear Roll Model by a Generalized Asymptotic Method, Ocean Engineering, Vol. 26 (1999) pp. 1169-1181

[6] M. Taylan: The Effect of Nonlinear Damping and Restoring in Ship Rolling, Ocean Engineering, Vol. 27 (2000) pp. 921-932

[7] S. Surendran and R., Venkata Ramana Reddy: Roll Dynamics of a Ro-Ro Ship, International Ship Building Progress, Vol. 49, No. 4 (2002) pp. 301-320

[8] T. Perez and G. C. Goodwin: Constrained predictive Control of Ship Fin Stabilizers to Prevent Dynamic Stall, Control Engineering Practice 16, (2008) pp. 482-494

[9] C. Holden and T. I. Fossen: A Nonlinear 7-DOF Model for U-Tanks of Arbitrary Shape, Ocean Engineering 45 (2012) pp. 22-37

[10] J. N. Sgobbo and M. Parsons: Rudder/Fin Roll Stabilization of the USCG WMEC 901 Class Vessel, Marine Technology, 36 (1999) pp. 157-170

[11] S. Surendran, S. K. Lee and S. Y. Kim: Studies on an Algorithm to Control the Roll Motion Using Active Fins, Ocean Engineering 34 (2007) pp. 542-551

[12] W. Guan and X. K. Zhang: Concise Robust Fin Roll Stabilizer Design Based on Integrator Backstepping and CGSA, Systems and Control in Aeronautics and Astronautics (ISSCAA) 2010 3rd International Symposium on , Vol. 3, 8-10 June 2010, pp. 1392-1397

[13] H. Ghassemi, F. H. Dadmarzi, P.Ghadimi and B. Ommani: Neural Network-PID Controller for Roll Fin Stabilizer. Polish Maritime Research 2(65) Vol. 17, (2010) pp. 21-30

[14] Ş. C. Karakas, E. Uçer and E. Pesman: Control Design of Fin Roll Stabilization in Beam Seas Based on Lyapunov's Direct Method, Polish Maritime Research 2 (73), Vol. 19 (2012) pp. 25-30

[15] Y. Ikeda, Y. Himeno and N. Tanaka: New York A Prediction Method for Ship Roll Damping, Report No. 00405 of the Department of Naval Architecture, University of Osaka Prefecture (1978)

[16] Y. Ikeda: Prediction Methods of Roll Damping of Ships and their Application to Determine Optimum Stabilization Devices, Marine Technology, Volume 41, Number 2, 1 April 2004, pp. 89-93(5)

[17] M. Aydın and H. Akyıldız: Assessment of the Intact Stability Characteristics of the Fishing Boats Suitable for Turkish Water. ITU publications Vol. 4 (2005) No. 6

[18] M. A. S. Neves., N. Perez and L. Valerio: Stability of Small Fishing Vessels in Longitudinal waves, Ocean Engineering, 26 (1999) pp. 1389-1419

[19] I. R. Ozkan: Lyapunov's Direct Method for Stability Analysis of Ships, ITU, PhD dissertation (1977)

[20] K. Ogata: Modern Control Engineering, Prentice-Hall, 4th Edition, New Jersey (1990)

[21] T. Hagiwara, K. Yamada, Y. Ando, I. Murakamı, S. Aoyama and S. Matsuura: A Design Method for Modified PID Control Systems For Multiple-Input, multiple-Output Plants To Attenuate Unknown Disturbances, World Automation Congress, Vol. 7 (2010) pp. 1-6

[22] A. Visioli: Modified Anti-Windup Scheme for PID Controllers. IEE Proc.-Cont. Theory App., Vol. 150, No. 1 (2003) pp. 49-54

[23] K. Yamada, N. Matsushima and T. Hagiwara: A Design Method for Modified PID Controllers for Stable Plants and Their Application, ECTI

Transactions on Electrical Eng., Electronics and Communications Vol. 5, No. 1 (2007) pp. 31-40

[24] V. M. Alfaro, R. Vilanova, O. Arrieta: Robust Tuning of Two-Degree-of-Freedom (2-DoF) PI/PID-based Cascade Control Systems, Journal of Process Control, Vol. 19, No. 10 (2009) pp. 1658-1670

[25] R. E. Precup, S. Preitl, E. M. Petriu, J. K. Tar, M. L. Tomescu, C. Pozna: Generic Two-Degree-of-Freedom Linear and Fuzzy Controllers for Integral Processes, Journal of The Franklin Institute, Vol. 346, No. 10 (2009) pp. 980-1003

[26] W. Ji, Q. Li, B. Xu, D. Zhao, S. Fang: Adaptive Fuzzy PID Composite Control with Hysteresis-Band Switching for Line of Sight Stabilization Servo System, Aerospace Science and Technology, Vol. 15, No. 1 (2011) pp. 25-32

[27] A. Hadlovska, S. Jajcisin: Predictive Control Algorithms Verification on the Laboratory Helicopter Model, Acta Polytechnica Hungarica, Vol. 9, No. 4 (2012) pp. 221-245

[28] F. Tahri, A. Tahri, A. Allali, S. Flazi, The Digital Self-Tuning Control of Step a Down DC-DC Converter, Acta Polytechnica Hungarica, Vol. 9, No. 6 (2012) pp. 49-64