Using data mining to improve digital library services

Ana Kovacevic
Faculty of Security Studies, University of Belgrade, Belgrade, Serbia
Vladan Devedzic
Department of Software Engineering, FON – School of Business Administration, University of Belgrade, Belgrade, Serbia, and
Viktor Pocajt
Faculty of Technology and Metallurgy, University of Belgrade, Belgrade, Serbia

Received 2 September 2009
Revised 11 November 2009
Accepted 16 November 2009

Abstract

Purpose – This paper aims to propose a solution for recommending digital library services based on data mining techniques (clustering and predictive classification).

Design/methodology/approach – Data mining techniques are used to recommend digital library services based on the user's profile and search history. First, similar users were clustered together, based on their profiles and search behavior. Then predictive classification for recommending appropriate services to them was used. It has been shown that users in the same cluster have a high probability of accepting similar services or their patterns.

Findings – The results indicate that *k*-means clustering and Naive Bayes classification may be used to improve the accuracy of service recommendation. The overall accuracy is satisfying, while average accuracy depends on the specific service. The results were better for frequently occurring services.

Research limitations/implications – Datasets were used from the KOBSON digital library. Only clustering and predictive classification was applied. If the correlation between the service and the institution were higher, it would have better accuracy.

Originality/value – The paper applied different and efficient data mining techniques for clustering digital library users based on their profiles and their search behavior, i.e. users' interaction with library services, and obtain user patterns with respect to the library services they use. A digital library may apply this approach to offer appropriate services to new users more easily. The recommendations will be based on library items that similar users have already found useful.

Keywords Digital libraries, Databases, Serbia, Data handling, Service delivery

Paper type Research paper

1. Introduction

Although we are today overwhelmed with data, techniques for finding appropriate information are mostly based on syntax search or low-level multimedia features. For improving search results in interaction with digital libraries (DLs), other more intelligent techniques should be used, based on both top-down knowledge creation (e.g. ontologies, user modeling) and bottom-up automated knowledge extraction (e.g. data mining, web mining) (Chen, 2003).

Valuable information extracted from the collection of DL data can be integrated into the library's strategy, and can be used to improve library search (Chang and Chen, 2006). For an effective design of systems and particularly to help users to find

The Electronic Library
Vol. 28 No. 6, 2010
pp. 829-843

0264-0473
DOI 10.1108/02640471011093525

information more easily, it is crucial to understand how people perform searches. This is especially important in continuous development of technologies. By exploring users' behavior we try to understand better the users themselves and their information needs and provide them with better user-oriented applications. To achieve this goal, we can anticipate a specific user's needs and problems in advance, by using experience of other similar users.

The idea of a recommender system is to help users by advising them on relevant products/information by predicting in advance their interest in a product; this prediction is based on various types of information, e.g. users' past purchases and product features (Huang *et al.*, 2002). For a DL, user recommendations may be very helpful (Geisler *et al.*, 2001, Liao *et al.*, 2009).

To help DL users obtain useful information more easily, we can use data mining techniques. Since data mining techniques are very popular, many researchers have applied them in various domains. However, few are focused on the domain of DLs. Our main objective is to use data mining techniques to recommend specific services to DL users.

We have developed the REKOB system to support the users of a specific digital library called KOBSON (http://kobson.nb.rs). In REKOB, we apply different and efficient data mining techniques for clustering DL users based on their profiles and their search behavior. We do not apply data mining to the library documents, but to its services; thereafter we recommend an appropriate service to a new user. By services, we assume online journal services such as Science Direct, Springer, and Blackwell among others.

The paper is organized as follows: a review of related work is discussed in section 2; section 3 describes the REKOB system architecture; experimental results and evaluation are provided in section 4; and finally, we draw our conclusions and plans for future work in section 5.

2. Related work

Data mining is an iterative process of searching for new, previously hidden information in large volumes of data (Kantardzic, 2003; Devedzic, 2001). We apply data mining according to the CRISP-DM standard (Cross-industry Standard Process for Data Mining, www.crisp-dm.org) (Chapman *et al.*, 2000); it specifies the following phases in the process of data mining (see Figure 1):

- *Business understanding*. After defining the project objectives and requirements, formulate data mining problem definition and prepare initial strategy for achieving these goals.
- *Data understanding*. After collecting the data, analyze it to familiarize with it and discover initial insight; also evaluate the quality of data.
- *Data preparation*. Prepare the final data set from the initial raw data that will be used in the process. Select cases and variables that are appropriate for analyses and perform the necessary data transformations.
- *Modeling*. Apply appropriate modeling techniques and calibrate the model to optimize the results. The results show patterns (i.e. the model) discovered for the data analyzed. If necessary, loop back to the preparation phase to bring the form of the data in line with the specific requirements for particular data mining techniques.

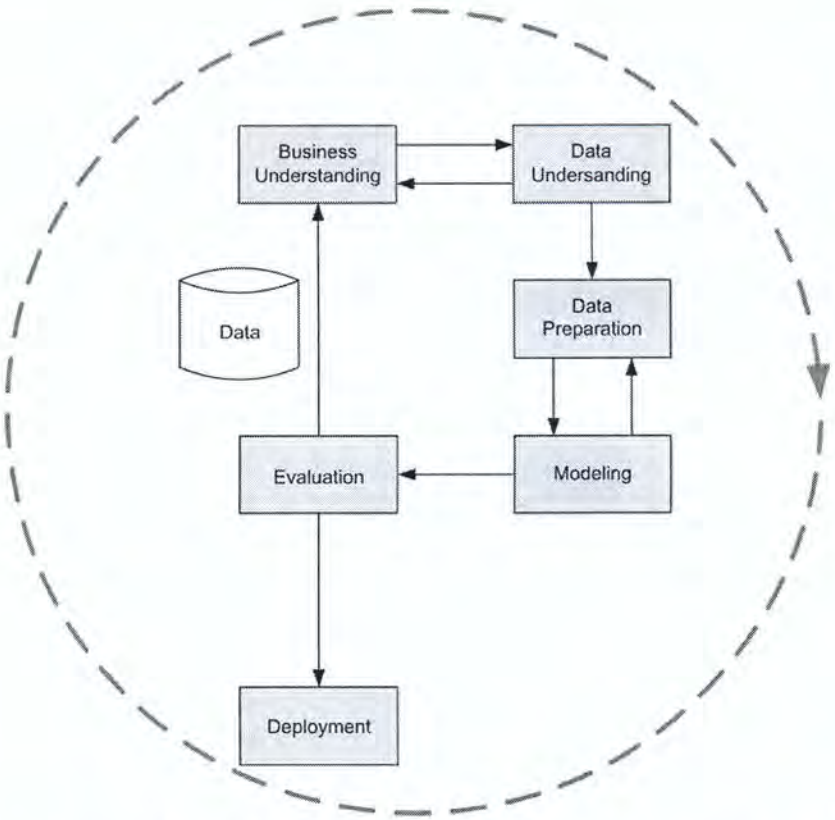

Figure 1.
Iterative process of data mining

- *Evaluation.* Evaluate the model (the patterns discovered) from the quality and effectiveness perspectives before deploying, and discover whether the model achieves the objectives set for it in the business-understanding phase.
- *Deployment.* make use of the model (patterns) created.

Note that it is very important to preprocess the data before continuing with the data mining process (Larose, 2005), otherwise the end-results will be misleading.

Recommender systems are implemented in various domains. In the commercial world, the best known example of a recommender system is Amazon.com (www.amazon.com), a web-based book seller, where personalized recommendation is made to its customers by combining three approaches: traditional collaborative filtering, cluster models, and search-based methods (Linden *et al.*, 2003). Gao *et al.* (2005) proposed a recommender approach to stock data analysis based on common user interests and a neural network. Although the experimental platform used a stock data set, their approach can also be useful in other numerical applications and in text-based digital libraries.

Qin *et al.* (2008) clustered partial preference relations as a means for agent prediction of users' preferences. New preferences for a user may be predicted by examining preferences of other users in the same cluster. They experimented on commonly used dataset for testing collaborative filtering technology: the MovieLens dataset (The MovieLens dataset is an existing dataset of movie ratings available at: http://grouplens.org). In the DL domain, Huang *et al.* (2002) used a graph-based recommender system that combines the content-based and collaborative approaches. Tsai and Chen (2008) recommend items from an e-library environment using adaptive resonance theory and data mining techniques. They cluster users by using adaptive resonance theory and then apply Apriori algorithm to generate rules for recommending appropriate literature.

3. REKOB

Our data mining system – REKOB – supports new users of the KOBSON library. In fact, KOBSON is the major provider of digital learning resources in Serbia (Consortium of Serbian Libraries) (Kosanovic, 2002), but in our research, we also use the term KOBSON to denote the corresponding DL in a technical sense. KOBSON provides Serbian students, teachers, and researchers access to foreign journals and learning resources. The DL services that KOBSON supports include Science Direct (SD, www.sciencedirect.com), Springer (SP, www.springerlink.com), Blackwell (BW, www.blackwell-synergy.com), Proquest (PQ, http://proquest.umi.com), JS (JSTORE, www.jstor.com), etc.

3.1 Overview of the data mining process

We have defined the data mining problem and made an initial plan for recommending KOBSON services to new users by analyzing the process with domain experts and identifying their need to improve KOBSON's user-oriented services. Our objective was to help new users of the KOBSON DL (as well as the ones who have a problem in finding relevant resources) in finding appropriate information by recommending them the service that similar users have found the most useful for them. The recommendation can be generated starting from the user's profile information and from the similarity of the user's behavior (when interacting with the KOBSON DL) with that of other users.

The data mining process is performed as illustrated in Figure 2. Users' search history data (from the KOBSON proxy server log) and their profiles (from the KOBSON database) are collected and then analyzed. Appropriate data from the log are parsed using regular expressions. From the log file, we extract users who have downloaded at least one paper from the DL. The parsed log data and user profiles are then loaded into the previously created database tables (Oracle 10g rel. 10.2, www.oracle.com), as shown in Figure 2. In the next step, data preparation, these data are transformed and normalized into a format suitable for clustering the users.

After the data preparation is completed, the data mining is performed. The results of the data mining are user patterns – clusters of similar users generated based on the users' profile information and search behavior. This is done using the *k*-means data-mining algorithm (Campos *et al.*, 2003), further discussed in the Appendix. The user patterns are generated in the form of clustering rules (see section 3.5 and Figure 3 for details).

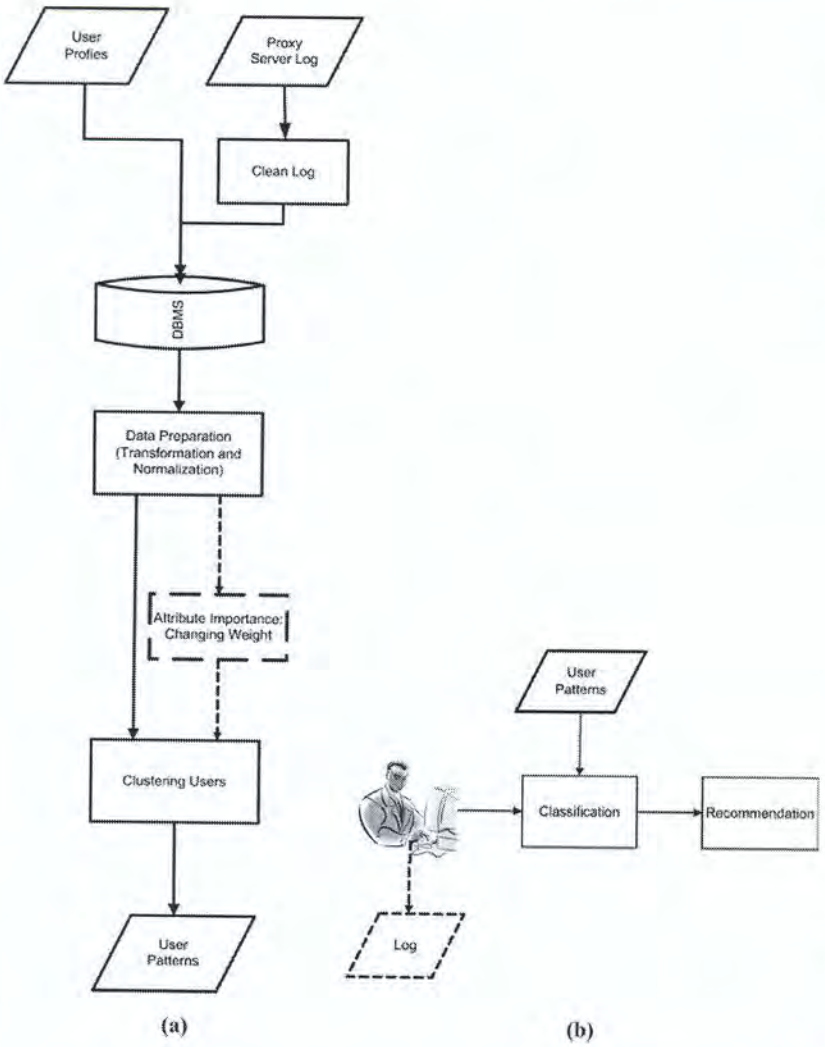

Figure 2.
The REKOB system data mining (a) processes (b) recommendation

In order to perform the classification of new users efficiently, the most significant attributes of data in each cluster should be identified. This is especially important in case, when there are a large number of attributes, to reduce them to the most significant ones. For discovering the most significant attributes, the minimum description length algorithm (MDL, 2008) is used (it is also further detailed in the Appendix). In REKOB, MDL is used to assign a higher weight to a specific, more significant attribute.

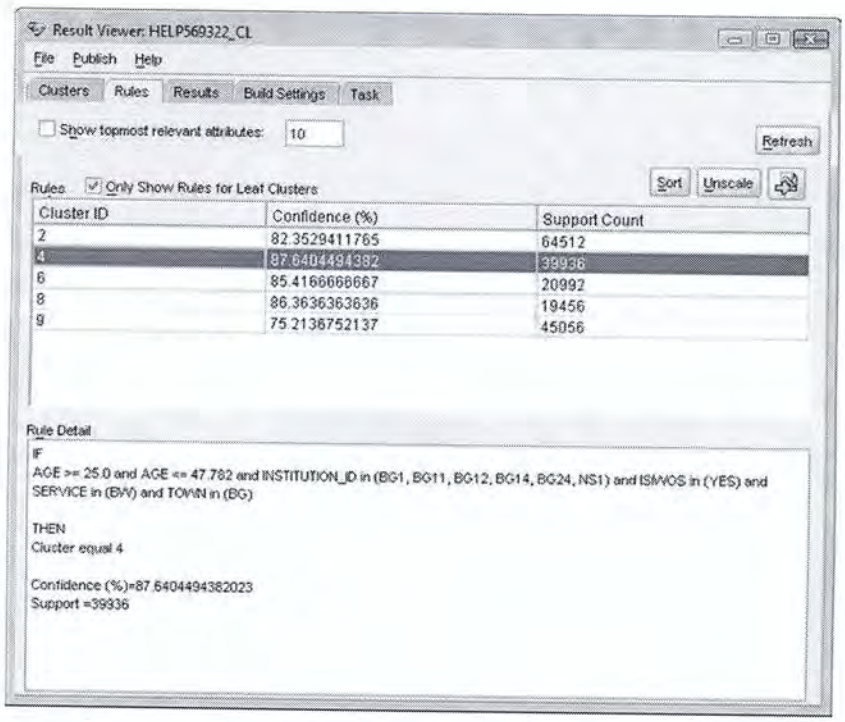

Figure 3.
An example of a clustering rule, with the corresponding confidence and support factors

3.2 Recommendations

Once the clusters (user patterns) are created, new users can be matched to them and the user category (the most similar cluster) can be determined (see Figure 2). Finally, a suitable recommendation for a new user is created based on the Naive Bayes classification algorithm (Larose, 2006). This algorithm is presented in more detail in the Appendix. Note that information about the new user's interaction is also logged into the system log. Likewise, information about all users is periodically refreshed in the log and the data mining is run again in order to update the user patterns.

3.3 Data

In the data mining process outlined in Figure 2, we used real-world data from the KOBSON DL (see Figure 2, top). To be able to use the resources and services of the KOBSON DL, a user has to register first. The user profile data collected through the registration procedure (user id, name, institution, age, gender and e-mail) is slightly modified for the data mining process. To protect the users' privacy, we replaced their names and institutions with codes. Hence, only authorized individuals will be aware of the users' identity and the institution he/she is with.

In addition to the users' demographic and administrative data collected during the registration process, we also used data collected from the proxy server (KOBSON,

2005), and related to the download of articles, using specific services from the DL (see the beginning of section 3). In the period 2004-2006, the number of downloaded articles per year was in the range 650,000-850,000.

We partitioned the data into a training data set and a test data set. We used the training data set to discover the user patterns, and then conducted pattern evaluations using the test data set. The results obtained with the test data are an indicator of how the user patterns will perform with the new data. We built and tested several models (i.e. several collections of clusters (user patterns)), using different techniques, and made a choice based of the performance on the test data set.

3.4 Data preparation

Preparing the data for data mining required to parse the data from the KOBSON proxy server log (see Table I) by using regular expressions and to extract only those users who had downloaded at least one paper.

After loading these selected data into the database (Oracle 10 g rel. 2), we obtained two relevant tables in the database: user (see Table II) and service (see Table III). The user table contains demographic and administrative information about the users of the KOBSON DL. The serivce table contains the data (extracted from the log, Table I) that describes instances of the DL usage patterns, which include the user, the service used, the date, and time of access, and the files downloaded. The initial version of the service table that we obtained contained about 230,000 records.

In the further data preparation, we join the user table and the service table and group by User_id and Service. The resulting table is the download table (see Table IV)

Attribute	Value
IP address	XXX.XXX.XXX.XXX
Agent	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
Userid	User1
Date:time	[22/Sep/2004:14:18:16 + 0100]
Method	GET
URL	service/doc1.pdf
Protocol	HTTP/1.1
Status	200
Size	119562

Table I.
A record from the proxy server log

Attribute	Data type and precision
User_id	VARCHAR2 (9)
Name	VARCHAR2 (30)
Surname	VARCHAR2 (30)
Institution_id	NUMBER (6)
Town	VARCHAR2 (30)
Age	NUMBER (3)
Gender	VARCHAR2 (1)
Position	VARCHAR2 (20)
E-mail	VARCHAR2 (50)

Table II.
The user table in the database

and it is used in further DM process. User_id, Institution_id, Town, Gender, Age are attributes from the user table. The service attribute denotes the service used by the concrete user. The Rec_Num attribute stands for the number of papers downloaded by the specific user and by using the specific service.

3.5 Discovering user patterns

We used Oracle Data Miner10.2 (www.oracle.com/technology/products/bi/odm/odminer.html) in the data mining process. In particular, we relied on using OracleDataMiner JavaApi, but it is also possible to generate PL/SQL code for the specific model. First, we explored the data visually and statistically in order to familiarize better with the data and acquire the knowledge necessary to guide the data mining process. To do this, we used histograms (where the data distribution can be seen more easily). For example, if we want to analyze the age attribute, and we set the number of bins to five, the histogram will show five bars (or groups), age values for each group, the number of cases in each bin, and the corresponding percentage (see Figure 4).

Normalization of numerical data has been done by using min/max techniques in which all the data is normalized to the values between 0 and 1. In the process of indentifying similar users by applying the enhanced *k*-means and other data mining techniques (as described previously and in the Appendix), we set the maximum number of clusters to five and the maximum number of the algorithm iterations to ten. Euclidean distance has been used for the distance function.

After such a clustering, we extracted rules to describe the clusters, with the corresponding confidence and support factors (see the Appendix). An example of such a rule with the corresponding confidence is shown in Figure 3. Support count shows the number of cases for which the rule is true, and confidence (%) is the probability that a case described by this rule will actually be assigned to the cluster. In plain English, the meaning of the clustering rule shown in Figure 3 is:

Table III.
The service table in the database

Attribute	Data type and precision
User_id	VARCHAR2 (9)
Service	VARCHAR2 (10)
Date	DATE
File	VARCHAR2 (250)

Table IV.
The download table in the database

Attribute	Data type and precision
User_id	VARCHAR2 (20)
Institution_id	VARCHAR2 (20)
Town	VARCHAR2 (2)
Rec_num	NUMBER (9)
Gender	VARCHAR2 (1)
Age	NUMBER (3)
Service	VARCHAR2 (20)

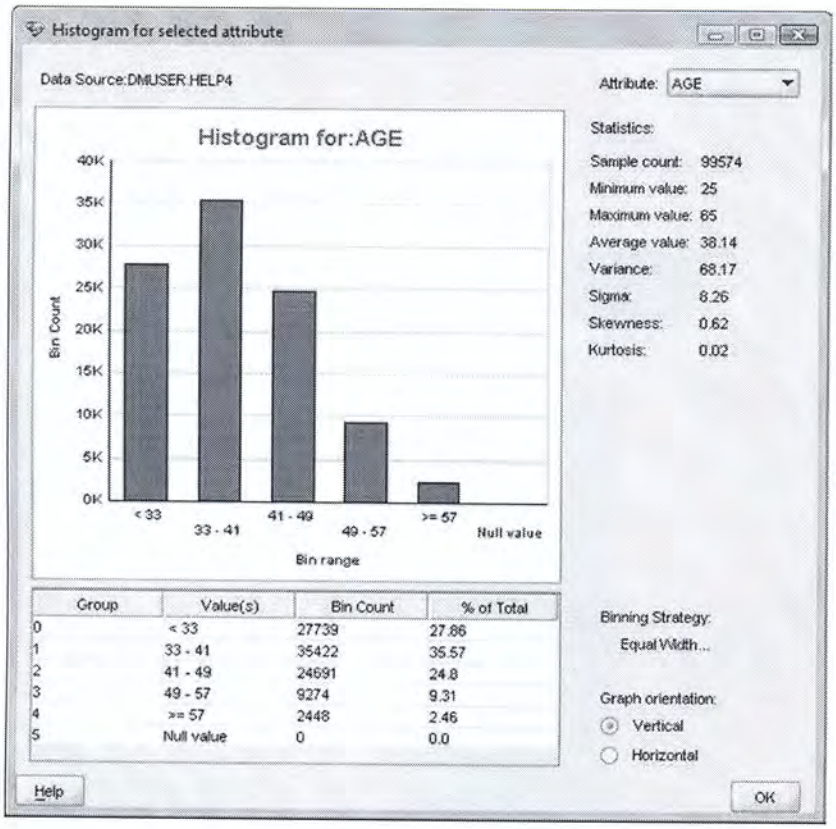

Figure 4.
A histogram for the age attribute

If:

- the user is a man, and he is between 25 and 47.74 years old;
- he is from a specific institution (e.g. BG1, BG11, BG12, BG14, NS1); and
- he is from the town "BG".

Then:

- he would use the BW service with the confidence of 80.48780 (support: 16,896 cases).

Note that this rule (as well as all other clustering rules discovered) relates users (see Table II) and KOBSON services they used (see Table III). In other words, the user patterns (clusters) actually describe information of the "which users use what KOBSON services" kind. This information is essential for recommending services to the new users, once they are matched to the user patterns (clusters).

In order to characterize the user patterns more effectively, we explored the most significant attributes of the KOBSON services. Finding the most significant attributes,

is especially important in case where there is a lot of attributes, and it is too expensive (and almost useless) to consider all of them within the data mining process. Figure 5 shows that the institution's code (institution_id) is the most important attribute for the target service, whereas gender is not so important (what may have been expected). In other words, Figure 5 reflects the importance of certain attributes of a specific service for recommending that service to the user. Accordingly, we may use this for weight factors, so institution_id will have a higher weight than the other attributes (see Figure 5) in clustering.

4. Evaluation

As can be seen in Figure 1, the crucial step before deploying the model is its evaluation.

Classification algorithms may be tested with new records, based on the classification model built. The results obtained with the test data, can be used as an indicator, of how well the model will work with new data. We used 60 per cent of our data for building the model and the remaining 40 per cent for testing it. The data are divided in two sets by random selection of cases using the split transformation.

The test indicated that the Naïve Bayes classification has 62.2 per cent better predicting confidence than the "naive" rule. The "naive" rule is usually used as the baseline for evaluating the performance of classification (Campos *et al.*, 2005), (Hamm, 2007). Using the naive rule method, a new record will be classified as a member of a

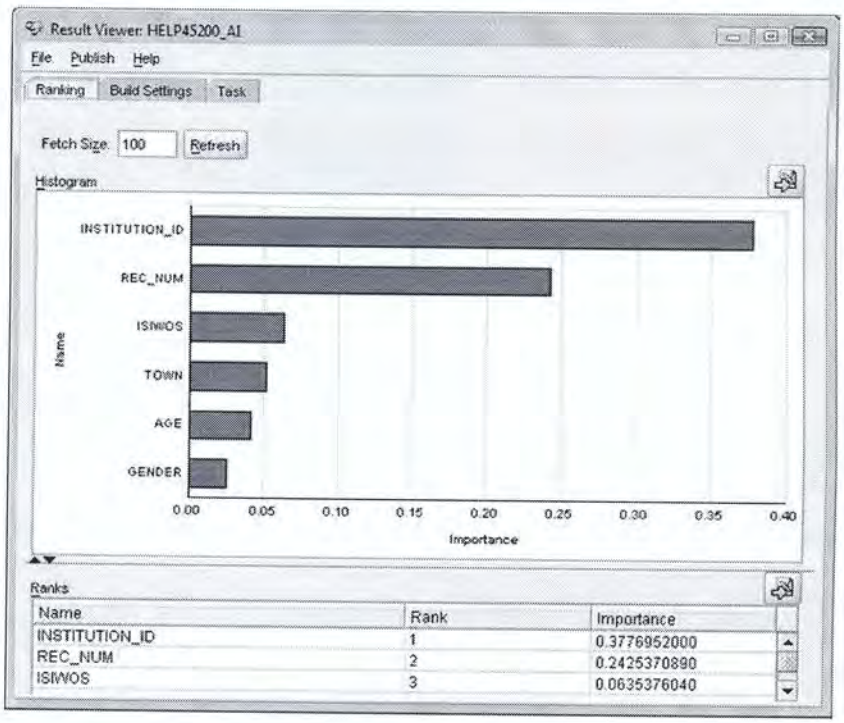

Figure 5.
Example of attribute importance

major class. So in our case, the naive rule would classify that all users would use the SD service (Science Direct), as a dominant service. The Decision Tree with clustering achieved a predictive confidence of 44 per cent better than the naive rule.

Furthermore, when we tried to test the classification using the Naïve Bayes algorithm without clustering the users first, we achieved a worse result – the predicting confidence was 43.49 per cent. Therefore, we chose Naïve Bayes with clustering for further matching. The rules for classification have been created, deployed, and added to the application.

We can further evaluate our model using the confusion matrix (see Figure 6). The confusion matrix is made on the test data, when the output value (in this case, the service the user has used) is known and is compared to the values predicted by the model. In the confusion matrix, the rows show the actual data, whereas the columns are the predictions made by the classification model. The confusion matrix measures the probability that the model predicts correct and incorrect values. Moreover, by analyzing the confusion matrix we can find out the types of errors that the model is likely to make. The sum of the values in the matrix is equal to the number of scored records in the input data table.

In the confusion matrix shown in Figure 6, it may be seen that the most frequent services, such as SD (Science Direct), PQ (ProQuest), and BW (Blackwell) are predicted

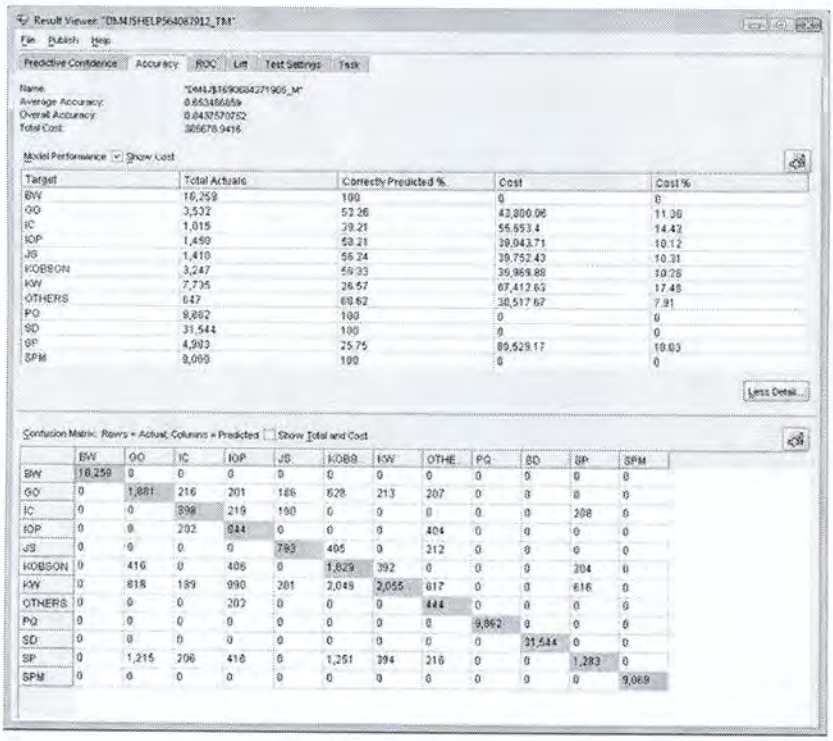

Figure 6.
Accuracy, cost and confusion matrix of the model

correctly, but there were errors in predicting less frequent services. For example, all of the 31.544 records for the SD service was predicted correctly, whereas for the JS (JStore) service 793 records were predicted correctly and 577 incorrectly. The correctness of its prediction is 793/(577 + 793) = 0.5788 (or 57.88 per cent).

The overall accuracy is the simple accuracy of the model, whereas the average accuracy is the average per-class accuracy achieved at a specific probability threshold that is greater than the accuracy achieved at all other possible thresholds (ODM, 2005). The overall accuracy is 84, whereas the average accuracy is lower, 65. The prediction is the most likely target value for this case, and the probability is the confidence in that prediction.

5. Conclusion and future work

In this paper we have proposed a solution for recommending digital library users a service from the library, based not only on statistical significance of service usage, but also considering the users' profiles. Our main research was focused on helping users to find relevant material more easily. We achieved it by using data mining techniques on historical data and by recommending the services that similar users would choose. We first clustered the users based on their profiles together with their search behavior. It has been shown that the users in the same cluster have high preference for using similar services. The results show that the *k*-means clustering and the Naïve Bayes classification can be used together to improve the service recommendation. Finally, we applied our model to test data in order to evaluate its accuracy. It has been shown that the overall accuracy is satisfactory, especially in frequently occurring services. In the near future, we plan to add an effective visual representation for recommending specific services to the users and to discover most significant problems that the users encounter by using text mining techniques to analyze the users' e-mails or free text interviews.

References

Campos, M., Milenova, B. and Mccracken, M. (2003), *Enhanced K-means Clustering*, United States Patent Application 20030212520, available at: www.freepatentsonline.com/y2003/0212520.html (accessed 26 August 2009).

Campos, M., Stengard, P. and Milenova, B. (2005), "Data-centric automated data mining", *Fourth International Conference on Machine Learning and Applications*, IEEE Computer Society, Los Angeles, CA, pp. 97-104.

Chang, C.C. and Chen, R.S. (2006), "Using data mining technology to solve classification problems – a case study of campus digital library", *The Electronic Library*, Vol. 24 No. 3, pp. 307-21.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinart, T., Shearer, C. and Wirth, R. (2000), *CRISP–DM Step-by-step Data Mining Guide*, available at: www.crisp-dm.org/CRISPWP-0800.pdf (accessed 26 August 2009).

Chen, H. (2003), *Towards Building Digital Library as an Institution of Knowledge*, NSF Post Digital Library Futures Workshop, Chatham, MA, available at www.sis.pitt.edu/%7Edlwkshop/paper_chen.html (accessed 26 August 2009).

Devedzic, V. (2001), "Knowledge discovery and data mining in databases", in Chang, S.K. (Ed.), *Handbook of Software Engineering and Knowledge Engineering Vol. 1 – Fundamentals*, World Scientific Publishing, Singapore, pp. 615-37.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From data mining to knowledge discovery: an overview", in Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), *Advances in Knowledge Discovery and Data Mining*, American Association for Artificial Intelligence, Menlo Park, CA, pp. 1-34.

Gao, K., Wang, Y.-C. and Wang, Z.Q. (2005), "Similar interest clustering and partial back-propagation-based recommendation in digital library", *Library Hi Tech*, Vol. 23 No. 4, pp. 587-97.

Geisler, G., McArthur, D. and Giersch, S. (2001), "Developing recommendation services for a digital library with uncertain and changing data", *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital libraries, Roanoke, VI*, ACM, New York, NY, pp. 199-200.

Hamm, C.K. (2007), *Oracle Data Mining Gold from Your Warehouse*, Rampant TechPress, Kittrell, NC.

Huang, Z., Chung, W., Ong, T.-H. and Chen, H. (2002), "A graph-based recommender system for digital library", *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, Portland, OR*, ACM, New York, NY, pp. 65-73.

Kantardzic, M. (2003), *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, Hoboken, NJ.

KOBSON (2005), internal data of the project on the evaluation of the Serbian authors publishing productivity.

Kosanovic, B. (2002), "Koordinirana nabavka inostranih izvora naučno-tehničkih informacija u Srbiji: stanje i perspektive", *Infoteka*, Vol. 3 Nos 1-2, pp. 55-64.

Larose, D. (2005), *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Hoboken, NJ.

Larose, D. (2006), *Data Mining: Methods and Models*, John Wiley & Sons, Hoboken, NJ.

Liao, S., Kao, K., Liao, I., Chen, H. and Huang, S. (2009), "PORE: a personal ontology recommender system for digital libraries", *The Electronic Library*, Vol. 27 No. 3, pp. 496-508.

Linden, G., Smith, B. and York, J. (2003), "Amazon.com recommendations: item-to-item collaborative filtering", *Internet Computing*, Vol. 7 No. 1, pp. 76-80.

MDL (2008), *Minimum Description Length, Oracle® Data Mining Concepts*, available at: http://download.oracle.com/docs/cd/B28359_01/datamine.111/b28129/algo_mdl.htm#CHDGBDAJ (accessed 26 August 2009).

ODM (2005), *Oracle Data Mining Concepts 10g Release 2*, available at: http://download.oracle.com/docs/html/B14339_01/4descriptive.htm#i1005741 (accessed 26 August 2009).

Qin, M., Buffett, S. and Fleming, M.W. (2008), "Predicting user preferences via similarity-based clustering", *Proceedings of the Twenty-first Canadian Conference on Artificial Intelligence, Windsor, Ontario*, Springer-Verlag, Heidelberg, pp. 222-33.

Rissanen, J. (1978), "Modeling by shortest data description", *Automatica*, Vol. 14, pp. 465-71.

Tsai, C.S. and Chen, M.Y. (2008), "Using adaptive resonance theory and data-mining techniques for materials recommendation based on the e-library environment", *The Electronic Library*, Vol. 26 No. 3, pp. 287-302.

Appendix

Clustering

For grouping similar users together, we employed an unsupervised learning technique-clustering. The data within a cluster are more similar to one another than to those in other clusters. We cluster users according to their personal profiles and search behavior

(extracted from the log) by using the enhanced *k*-means algorithm (Campos *et al.*, 2003). It is a hierarchical top-down clustering algorithm, which generates clusters according to a user-specified criterion. The algorithm randomly defines initial centroids, which approximate centers of gravity, and uses a distance measure to calculate the distance between centroids and data objects (Hamm, 2007). It assigns probabilities to the generated clusters, by assuming that, the data in each cluster follows an isometric Gaussian distribution.

The rules that define when a data item fits into a cluster are specified as If-Then assertions of the form: "IF A implies B, THEN the cluster is Cluster1". Typically, these clustering rules are uncertain, and there are two certainty measures associated with each rule – confidence and support. For each rule, the confidence is the conditional probability of B given A, and the support for the rule is the estimation of the number of cases for which the rule is true.

The minimum description length (MDL) was first introduced by Rissanen (1978). The idea is that the simplest, most compact representation of data is the best and most probable explanation of the data. In the MDL algorithm, each attribute is considered as a simple predictive model of the target class and attributes are ranked according to their significance in predicting a target (MDL, 2008).

Classification

Classification is the task of mapping data into predefined classes (Fayyad *et al.*, 1996). Classification task begins with training data for which the target values are known. Classification of a collection consists of dividing the items that make up the collection into categories or classes (ODM, 2005). The model is built on historical data, and the goal is to predict the target class for each record of new data. In a specific case, the most likely target class is given as a prediction, and an appropriate probability indicates the confidence level of the prediction. Confusion matrix is used to display the type of errors the model is likely to make and is used for model evaluation.

Different classification algorithms use different techniques for finding relations between the predictor attributes' values and the target attributes' values in the build data. There is no perfect algorithm for all problems; the usefulness of an algorithm depends on the size of the dataset, the types of patterns that may exist in the data, the type of data, the goal of analysis, and many other factors (Hamm, 2007).

Generally, classification may be done for a binary target or for a multiclass target. (Binary targets have only two values, whereas multiclass targets have more than two values.) In a multiclass target, we may rank top choices for each case. In REKOB, the target has been multiclass. The data may be split in two sets, one for building the classification model and the other for testing the built model, where the model accuracy in predicting the target value is measured. In REKOB, we used the Naive Bayes classification algorithm (Larose, 2006). Because it is very fast and the attributes are independent from each other. Also, it is suitable for a multiclassification problem and for small number of predictor attributes. The Naïve Bayes algorithm is based on the Bayes' Theorem that states that the probability of event A occurring given that event B has occurred (P(A|B)) is proportional to the probability of event B occurring given that event A has occurred multiplied by the probability of event A occurring ((P(B|A)P(A)) (ODM, 2005). Naive Bayes algorithm makes the assumption, that each attribute is conditionally independent of the others. Even if this assumption of independence is violated in practice, it does not significantly degrade the model's predictive accuracy. In addition to the Naïve Bayes algorithm, we also tested the Decision Tree algorithm (Larose, 2005) for our classification case, but we achieved better results using the Naïve Bayes algorithm.

About the authors

Ana Kovacevic received her BSc and Msc degrees in Electrical Engineering from the University of Belgrade, where she also received her PhD in 2010. She is currently a Lecturer at the

University of Belgrade, Faculty of Security Studies. Her research interests include data mining, text mining, digital libraries, databases, visualization, and multimedia. She is a member of the GOOD OLD AI research network. Ana Kovacevic can be contacted at: kana@rcub.bg.ac.rs

Vladan Devedzic is a Professor of Computer Science with the Department of Software Engineering, FON – School of Business Administration, University of Belgrade. His long-term professional goal is to bring together ideas from the broad fields of intelligent systems and software engineering. His current professional and research interests include knowledge modeling, ontologies, semantic web, intelligent reasoning techniques, software engineering, and application of artificial intelligence to education and healthcare.

Viktor Pocajt received his BSc degrees in mechanical engineering (1988) and chemical engineering (1992), MBA (1994), and PhD (1999), all from the University of Belgrade. His main professional orientation is development and implementation of specialized e-business solutions and web-based information systems, applications and services for the metal working industry and environmental engineering. With more than 20 years of experience in metal working industry, research, IT development, applied computing, and consulting, he managed the implementation of several ERP systems, as well as development of a number of e-business solutions, specialized Web-oriented software systems and databases for the global market. Viktor has lectured on more than 70 specialized seminars and courses, and authored and co-authored about 40 scientific papers. He is currently a Professor at the Faculty of Technology and Metallurgy, University of Belgrade, Serbia.

THE COMBINED METHOD FOR UNCERTAINTY EVALUATION IN ELECTROMAGNETIC RADIATION MEASUREMENT

by

Aleksandar M. KOVAČEVIĆ [1*], ***Ana V. KOVAČEVIĆ*** [2], ***Koviljka Dj. STANKOVIĆ*** [3], ***and Uroš D. KOVAČEVIĆ*** [3]

[1]Technical Test Center, Ministry of Defense, Belgrade, Serbia
[2]Faculty of Security Studies, University of Belgrade, Serbia
[3]Faculty of Electrical Engineering, University of Belgrade, Serbia

Scientific paper
DOI: 10.2298/NTRP1404279K

Electromagnetic radiation of all frequencies represents one of the most common and fastest growing environmental influence. All populations are now exposed to varying degrees of electromagnetic radiation and the levels will continue to increase as technology advances. An electronic or electrical product should not generate electromagnetic radiation which may impact the environment. In addition, electromagnetic radiation measurement results need to be accompanied by quantitative statements about their accuracy. This is particularly important when decisions about product specifications are taken. This paper presents an uncertainty budget for disturbance power measurements of the equipment as part of electromagnetic radiation. We propose a model which uses a mixed distribution for uncertainty evaluation. The evaluation of the probability density function for the measurand has been done using the Monte Carlo method and a modified least-squares method (combined method). For illustration, this paper presents mixed distributions of two normal distributions, normal and rectangular, respectively.

Key words: combined method, electromagnetic radiation, disturbance power, mixed distribution, probability density function

INTRODUCTION

Electromagnetic radiation has been around since the birth of the universe. Light is its most familiar form. Electric and magnetic fields are part of the spectrum of electromagnetic radiation which extends from static electric and magnetic fields, through radiofrequency and infrared radiation, to X-rays. RF radiation is electromagnetic radiation in the frequency range of 3 kHz to 300 GHz on the electromagnetic spectrum and it is in the non-ionizing band of the spectrum. Non-ionizing just means there is not enough energy to break chemical bonds between molecules. Unlike ultraviolet light, gamma rays and X-rays are in the ionizing part of the electromagnetic spectrum. The electromagnetic spectrum encompasses both natural and human-made sources of electromagnetic fields. During the 20th century, environmental exposure to man-made electromagnetic fields has been steadily increasing as growing electricity demand, ever-advancing technologies and changes in social behaviour have created more and more artificial sources. Their presence has affected almost every aspect of living (home, work, travelling, school, *etc.*).

An electronic or electrical product should not generate electromagnetic radiation which may influence the environment (other products, persons). Directive 2004/108/EC of the European Parliament and EU Council regulates the electromagnetic compatibility of equipment. It aims to ensure the functioning of the internal market by requiring equipment to comply with an adequate level of electromagnetic compatibility (EMC) [1, 2].

EMC measurement results need to be accompanied by quantitative statements about their accuracy. This is particularly important when decisions about product specifications are taken. One of the common problems that we face when examining EMC is an inconsistent approach to adjusting various specified or standardized tests. Consequently, some of the standardized EMC measurements include precisely defined ways of evaluating uncertainty in measurement [3]. For instance, measurement instrumentation uncertainty (MIU), standards compliance uncertainty (SCU) and other types of uncertainties are presented.

* Corresponding author; e-mail: aleksandarkovacevic1962@yahoo.com

The SCU contains all uncertainties due to MSU, the set up of the equipment under test (EUT) including the lead under test (LUT), and uncertainties due to the measurement procedure and measurement space.

In practice, the uncertainty in the result of a standardized measurement may arise from many possible uncertainty sources. In a measurement standard, each uncertainty source should be specified in a quantitative way by using one or more influence quantities. Consequently, many uncertainty sources in the domain of EMC measurements were not studied enough and need further studying. EMC tests and measurements typically have large uncertainties of at least several decibels [4].

The basic document for evaluating and expressing uncertainty in measurement is the guide to the expression of uncertainty in measurement (GUM) [5]. Consequently, the GUM proposes a standard procedure which is known as GUF (GUM uncertainty framework), which is applied to linear or linearized models [6].

This paper presents a Type B uncertainty budget evaluation for the case of disturbance power measurements in the mains leads of an apparatus according to the standard SRPS EN 55014-1:2010 [7]. The uncertainty budget is limited to Measurement Instrumentation Uncertainty (MIU) [8, 9]. We propose a new model which uses mixed distribution for uncertainty evaluation [10, 11]. The evaluation of the probability density function (PDF) of the output quantity (measurand) has been done using the Monte Carlo method and a modified least-squares method (combined method). In addition, the model equation represents one purely additive linear model whose terms are independent [8]. For illustration, we present mixed distributions of two normal distributions, normal and rectangular, respectively. The results obtained by the Monte Carlo method and the modified least-squares method are compared to corresponding results when applying the standard GUM procedure [10, 11].

MEASUREMENT MODEL

The evaluation of measurement uncertainty is based on the knowledge of the measurement process and input quantities which influence the results of that measurement. The knowledge of the measurement process is expressed by the so-called model equation which reflects the interrelation between the measurand (output quantity) and the input quantities [12]. The knowledge of input quantities is represented by appropriate PDF.

The measurement of disturbance power is performed with an absorbing clamp (in addition to the measurement at the mains leads of the vacuum cleaner), according to the standard SRPS EN 55014-1:2010 [7].

For determining the measurand value, the standardized measurement method is used. Namely, the measurand is the disturbance power. The disturbance power P corresponding to the measured voltage V at each measurement frequency is calculated by using the clamp factor CF obtained from the absorbing clamp calibration procedure described in [13]

$$P = V + CF \tag{1}$$

where P is the disturbance power in dBpW, V – the measured voltage in dBμV, and CF – the clamp factor in dBpW/μV.

In addition, the clamp factor is given in the following equation

$$CF = A - 10\log_{10}(50) = A - 17 \tag{2}$$

where, A is the measured insertion loss in dB.

The possibility of variations of obtained measurand values becomes smaller, which reduces the measurement uncertainty that the standardized measurement method is being used.

The basic model of uncertainty includes the following separated sources of uncertainty in a measurement:

- setup of the equipment under test (EUT) ,
- measurement procedure,
- measurement space, and
- measurement means.

The uncertainty budget for the case of disturbance power measurements (the absorbing clamp test method – ACTM) as described in [8, 9] are not suitable for actual compliance tests in accordance with the CISPR specification given in [14]. Namely, this uncertainty budget is limited to MIU. Uncertainties due to the setup of the EUT, including the LUT, and due to the measurement procedure and measurement space (Faraday cage), are not taken into account.

The used measurement means are various for various EMC measurement methods, but for the case of disturbance power measurements in the mains leads of an apparatus the following general sources of Type B uncertainty in a measurement can be identified:

- receiver reading,
- receiver accuracy,
- frequency step error, and
- mismatch.

Receiver reading will vary for reasons which include measuring system instability, receiver noise, and meter scale interpolation errors (uncertainty determined by the least significant digit fluctuation).

The accuracy can be taken from the manufacturer's specification or calibration report of the receiver. If necessary, the uncertainty for different types of signals/responses may be considered, *i. e.*, CW accuracy, pulse amplitude response accuracy, pulse repetition response accuracy.

The frequency step error should be considered if we use an automated receiver with a programmed step

size. In this case, R&S ESVP, serial number 879691/037, was used as the test receiver and it was used for manual measurement of disturbance power, so that the frequency step error was not considered.

Apart from the given sources of uncertainty in a measurement, uncertainty sources originating from the absorbing clamp should be added to the measurement of disturbance power. It should be mentioned that our measurements were done in the frequency range of 30 MHz to 300 MHz and that an absorbing clamp (MDS 9, Robert Luthi GmbH, serial number 71170) was used. So, when the absorbing clamp (AbC) is used for the measurement of disturbance power, the following sources of uncertainty in the measurement can be identified:

- AbC receiver attenuation,
- AbC insertion loss, and
- AbC receiver mismatch.

The attenuation of the connection between the receiver and the absorbing clamp is obtained from the calibration report or manufacturer's data.

Absorbing clamp insertion loss can be taken from the calibration report. In addition, the loss of the RF cable is included.

For disturbance power measurements, the mismatch is given in the following eq.

$$M = 20\log_{10}(1 \pm \Gamma_{ac}\Gamma_{r}) \tag{3}$$

where Γ_{ac} is the voltage reflection coefficient (VRC) of the absorbing clamp and Γ_{r} – the VRC of the measurement receiver.

Consequently, Γ is related to voltage standing wave ratio (*VSWR*) by

$$\Gamma = \frac{VSWR - 1}{VSWR + 1} \tag{4}$$

In this case, the measurement receiver specification of $VSWR \leq 2.0{:}1$ (attenuation 0 dB) and the absorbing clamp specification of $VSWR \leq 3.25{:}1$ are assumed. In addition, this can be improved and the associated uncertainty reduced by applying a 6 dB attenuator on the output of the absorbing clamp [8].

The model equation for the evaluation of MIU is given in [8] by the following eq.

$$V = V_{r} + L_{c} + L_{ac} - 10\log_{10}(50) + \\ + \delta V_{sw} + \delta V_{pa} + \delta V_{pr} + \delta M \tag{5}$$

Equation (5) represents a purely additive linear model whose terms are independent. Information on the terms of the expression in the model equation is given in tab. 1. Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by Type A evaluation of measurement uncertainty, other components by Type B evaluation of measurement uncertainty. Consequently, Type A evaluation is done by calculation from a series of repeated observations using statistical methods and resulting in a probability distribution that is assumed to be normal. Type B evaluation of measurement uncertainty can also be characterized by standard deviations evaluated from probability density functions based on experience or other relevant information.

VALUES OF INPUT QUANTITIES

Table 1 presents an uncertainty budget, a Type B evaluation of the MIU for the case of disturbance power measurements. Namely, the given data are obtained from the manufacturer's specifications, calibration reports, and instrumentation manuals [15, 16], and are used for the evaluation of the MIU, according to the ISO-GUM [5].

The standard uncertainty $u(x_i)$ is calculated by dividing the value of the uncertainty associated with x_i by the coverage factor k_p, whose value depends on the

Table 1. Uncertainty budget of measurement instrumentation uncertainty (MIU) according to the GUM for disturbance power measurements

Input quantity	X_i	Estimate X_i (x_i)		Standard uncertainty $u(x_i)$ [dB]	Sensitivity coefficient c_i	Contribution to the standard uncertainty $u_i(y) = c_i u(x_i)$
		Value [dB]	Probability distribution function			
Receiver reading	V_r	±0.1	Rectangular $k_p = 1.732$	0.058	1	0.058
AbC-receiver attenuation	L_c	±0.1	Normal $k_p = 2.000$	0.050	1	0.050
AbC insertion loss	L_{ac}	+3.0 −0.2	Normal $k_p = 2.000$	0.800	1	0.800
Receiver sine wave voltage	δV_{sw}	±1.0	Normal $k_p = 2.000$	0.500	1	0.500
Receiver pulse amplitude response	δV_{pa}	±2.0	Rectangular $k_p = 1.732$	1.155	1	1.155
Receiver pulse repetition rate response	δV_{pr}	±2.0	Rectangular $k_p = 1.732$	1.155	1	1.155
AbC-receiver mismatch	δM	+1.40 −1.67	*U*-shaped $k_p = 1.414$	1.086	1	1.086

choice of the probability density function (PDF) and confidence level associated with the given value.

For rectangular or U-shaped probability distribution, where X_i is estimated to lie between $(x_i - a^-)$ and $(x_i + a^+)$, with a level of confidence of 100%, $u(x_i)$ is taken as $a/3^{1/2}$ or $a/2^{1/2}$, respectively [8]. In addition, $a = (a^+ - a^-)/2$ is the half-width (semi-width) of the probability distribution. For a normal probability distribution, the divisor is 2 if the value of the uncertainty associated with x_i has a level of confidence of 95,45% (the value is twice the experimental standard deviation). The estimated value x_i is determined with $x_i = (a^+ + a^-)/2$ [5].

EVALUATION OF MEASUREMENT INSTRUMENTATION UNCERTAINTY

Previous sections present the uncertainty budget of the MIU according to GUM for the case of disturbance power measurement standard SRPS EN 55014-1:2010 [7].

This section presents the application of the combined method to the uncertainty budget of the MIU for disturbance power measurement. Consequently, the model equation for the evaluation of MIU is given with eq. (5). Namely, the Monte Carlo method and the modified least-squares method (combined method) are applied in two cases for two independent input quantities from the given expression. The combined method is used for the evaluation of the probability density function for the output quantity (mixed distribution) according to the probability density function from two independent input quantities, *i. e.*, two independent input quantities assigned by normal distributions, two independent input quantities where the first quantity is assigned a normal distribution and the second is assigned a rectangular distribution. Convolution can be used for the general additive model $Y = X_1 \pm X_2 \pm \cdots \pm X_N$ by combining X_1 and X_2, and then combining this result with X_3, and so forth [17]. The instance of many input variables has not been discussed in this paper.

Monte Carlo simulations for obtaining mixed distributions are done by the procedure described in [10, 11, 18].

The number of classes of histogram k is determined according to eq. (6)

$$k = \sqrt{n} \tag{6}$$

For determining the value of k, other formulas exist in statistics [18]. When determining the empirical formula for k, the basic criterion is that at least one value of the random variable x fits in each class of histograms, providing histogram continuity. On the other hand, k must be greater than 3 in order for the histogram form to indicate the law of distribution of the random variable [19]. It is implied that k is taken as an integer value.

The values for the Monte Carlo simulations are taken from tab. 1. Consequently, the value of N, the total number of trials was 10^6. The number of data used for the simulation is $n = 10000$. Risk conformity was $\alpha = 0.05$, that is the confidence level $(1 - \alpha)$ was 0.95. One more data that is important for our simulation was the mixed coefficient ε which was 0.5. The results obtained by the combined method are compared to the corresponding results when applying the GUM.

The first example given refers to two independent input quantities, L_{ac} and δV_{sw}, assigned by two normal PDF. Table 1 shows that for the two given input quantities L_{ac} and δV_{sw} estimates are given with a^-/a^+, *i.e.*, –0.2/+3.0 dB and ±1 dB, respectively. Then the estimated values are $x_1 = 1.40$ and $x_2 = 0$, and standard uncertainties $u(x_1) = 0.8$ and $u(x_2) = 0.5$, respectively (see tab. 1). After entering the given data into a computer program created in Visual Basic 6.0, the generation of pseudorandom numbers is done (in our case, 10000). For pseudorandom numbers generated in this manner, a histogram is drawn (fig. 1) that represents the empirical curve of a mixed normal-normal distribution. In fig. 1 the empirical curve (histogram) is shown in a white line.

After that, determining point estimates parameters of a mixed normal-normal distribution is done by the combined method [10, 11]. With parameters that are determined like this, the estimated density function of a mixed normal-normal distribution is obtained. The values of these parameters are pseudorandom and with them the fitting of a mixed normal-normal distribution (estimated curve) is tried in a histogram. It should be mentioned that the coming of the curve through the mid-point of each class histogram is considered to be the best fitting.

In fig. 1, the estimated curve is shown in a grey line. Consequently, the theoretical curve is shown by a black line and represents the results obtained according to the GUM (fig. 1). It is noticeable that the fitting of the estimated curve (grey line) in a histogram (empirical curve) is very good, which indicates that the unknown parameters of this distribution are estimated well. Also, it is noticeable that the estimated curve

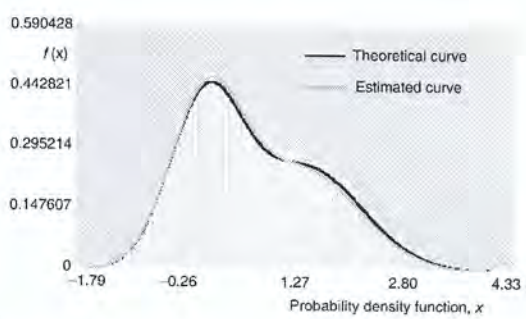

Figure 1. Mixed normal-normal distribution obtained by the combined method and GUM, respectively

(grey line) differs slightly from the theoretical curve (black line). This difference was the result of the evaluation of parameters of the mixed distribution (whose values are pseudorandom) and the number N of iterations (the total number of trials).

The second example which was used refers to two independent input quantities δV_{sw} and δV_{pa}, assigned by normal and rectangular PDF, respectively. Table 1 shows that for the two given input quantities, δV_{sw} and δV_{pa}, estimates are given with ±1 dB and ±2 dB, respectively. Then the estimated values are $x_1 = x_2 = 0$, and standard uncertainties $u(x_1) = 0.5$ and $u(x_2) = 1.155$, respectively (see tab. 1). As in the previous example, the given data are recorded into the computer program, and then the pseudorandom numbers (n = 10000) generating is done. For pseudorandom numbers generated in this manner a histogram is drawn (fig. 2) which represents the empirical curve of a mixed normal-rectangular distribution. In fig. 2, as in the previous example, the empirical curve (histogram), estimated curve, and the theoretical curve are shown by the white line, the grey line and the black line, respectively.

It is noticeable that the fitting of the estimated curve (grey line) in the histogram (the empirical curve)

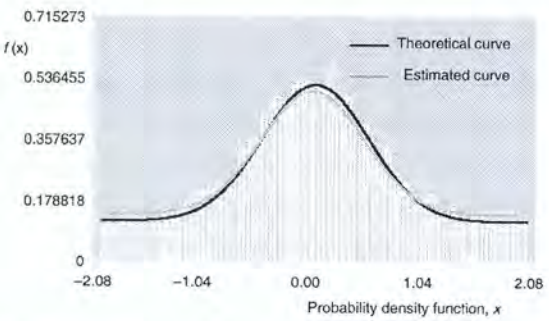

Figure 2. Mixed normal-rectangular distribution obtained by the combined method and GUM, respectively

is very good and does not deviate a lot from the theoretical curve (black line).

This difference was the result of the evaluation of parameters of the mixed distribution (whose values are pseudorandom) and the number N of iterations.

CONCLUSIONS

Electromagnetic radiation has affected almost every aspect of living. In our homes and at work, we are exposed to radiation emanating from electronic or electrical equipment. If you have met the requirements of electromagnetic compatibility for electronic or electrical equipment, you have satisfied the needs of the people. EMC measurement results need to be accompanied by quantitative statements about their accuracy. This is particularly important when decisions about product specifications are taken. This paper presents a Type B uncertainty budget evaluation for disturbance power measurements. The uncertainty budget is limited to MIU. We propose a new model which uses mixed distribution for uncertainty evaluation. The evaluation of the PDF for the measurand (output quantity) has been done using the Monte Carlo method and a modified least-squares method (combined method). The combined method is applied in two cases for two independent input quantities which were associated to appropriate PDF. It was shown that the combined method can be applied alternatively for the determination of the probability density functions of the output quantity to a satisfactory degree of accuracy. Namely, applying the combined method produces a mixed distribution, *i. e.*, PDF for the output quantity, which fits well (the estimated curve) in histograms and differs slightly from the produced results according to the GUM (theoretical curve).

ACKNOWLEDGEMENT

The Ministry of Education, Science and Technological Development of the Republic of Serbia supported this work under Contracts III 43009 and 171007.

AUTHOR CONTRIBUTIONS

Theoretical analysis and experiments were carried out by A. M. Kovačević. Literature research was carried out by A. M. Kovačević, K. Dj. Stanković, and A. V. Kovačević. The manuscript was written by A. M. Kovačević, with the suggestions of all authors. The figures were prepared by A. M. Kovačević and A. V. Kovačević. All authors analysed and discussed the results.

REFERENCES

[1] ***, CENELEC, Directive 2004/108/EC of the European Parliament and of the Council, 2004
[2] ***, Official Gazette of RS, No. 13, Book of Regulation on Electromagnetic Compatibility (in Serbian), 2010
[3] ***, IEC/CISPR, Geneva, Standard CISPR 16-4-1: Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – Part 4-1: Uncertainties, Statistics and Limit Modeling – Uncertainties in Standardized EMC tests, 2005
[4] 888, NIS 81, The treatment of uncertainty in EMC measurements, NAMAS Executive: National Physical Laboratory, UK, 1994
[5] ***, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, Geneva. Guide to the Expression of Uncertainty in Measurement, 1995 (ISBN 92-67-10188-9)

[6] ***, JCGM, Evaluation of Measurement Data-Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'-Propagation of distributions Using a Monte Carlo Method, 2008
[7] ***, ISS, Belgrade, Standard SRPS EN 55014-1: Electromagnetic Compatibility – Requirements for Household Appliances, Electric Tools and Similar Apparatus – Part 1: Emission (in Serbian), 2010
[8] ***, IEC/CISPR, Geneva. Standard CISPR 16-4-2: Specification for Radio Disturbance and immunity Measuring Apparatus and Methods – Part 4-2: Uncertainties, Statistics and Limit Modeling – Uncertainty in EMC Measurements, 2003
[9] ***, LAB 34, The Expression of Uncertainty in EMC Testing, NAMAS Report: UKAS, 2002
[10] Kovačević, A., Brkić, D., Osmokrović, P., Evaluation of Measurement Uncertainty Using Mixed Distribution for Conducted Emission Measurements, *Measurement, 44* (2011), 4, pp. 692-701
[11] Kovačević, A., Evaluation of Measurement Uncertainty in EMC Testing, Ph. D. dissertation (in Serbian), Faculty of Electrical Engineering, University of Belgrade, 2011
[12] Sommer, K. D., Siebert, B. R. L., Systematic Approach to the Modeling of Measurements for Uncertainty Evaluation, *Metrologia, 43* (2006), 4, pp. S200-S210
[13] ***, IEC/CISPR, Geneva, Standard CISPR 16-1-3: Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – Part 1-3: Radio Disturbance and Immunity Measuring Apparatus – Ancillary Equipment – Disturbance Power, 2004
[14] ***, IEC/CISPR, Geneva, Standard CISPR 16-2-2, Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – Part 2-2: Methods of Measurement of Disturbances and Immunity – Measurement of Disturbance Power, 2010
[15] ***, Rohde&Schwarz, Munchen, Germany. Operating manual, TEST RECEIVER ESVP, 1997
[16] ***, Robert Luthi GmbH, Biel, Switzerland, Operating manual, Absorbing clamp MDS 9, 1972
[17] Cox, M. G., Siebert, B. R. L., The Use of a Monte Carlo Method for Evaluating Uncertainty and Expanded Uncertainty, *Metrologia, 43* (2006), pp. 178-188
[18] Kovačević, A., *et al.*, Uncertainty Evaluation of the Conducted Emission Measurements, *Nucl Technol Radiat, 28* (2013), 2, pp. 182-190
[19] Chapouille, P., De Pazzis, R., Reliability Systems (in French), Masson et Cie, Paris, 1968

Received on May 7, 2014
Accepted on December 13, 2014

**Александар М. КОВАЧЕВИЋ, Ана В. КОВАЧЕВИЋ,
Ковиљка Ђ. СТАНКОВИЋ, Урош Д. КОВАЧЕВИЋ**

КОМБИНОВАНА МЕТОДА ЗА ПРОЦЕНУ МЕРНЕ НЕСИГУРНОСТИ ПРИ МЕРЕЊУ ЕЛЕКТРОМАГНЕТСКОГ ЗРАЧЕЊА

Електромагнетско зрачење на свим фреквенцијама представља један од најчешћих и најбрже растућих утицаја на животну средину. Све популације изложене су различитим степенима, а нивои ће наставити да расту како технологија буде напредовала. Електронски или електрични производи не смеју генерисати електромагнетско зрачење које може утицати на животну средину, при чему резултати мерења треба да буду праћени квантитативним изјавама о њиховој тачности. Ово је нарочито важно када се доносе одлуке о карактеристикама производа. У овом раду представљен је буџет мерне несигурности при мерењу снаге сметњи код производа као део електромагнетског зрачења. За процену мерне несигурности предложен је један модел који користи мешовиту расподелу. Метода Монте Карло и модификована метода најмањих квадрата (комбинована метода) коришћене су за процену функције густине расподеле мерене величине. Као илустрација, у овом раду је представљена мешовита расподела од две нормалне расподеле, нормалне и правоугаоне расподеле, респективно.

Кључне речи: комбинована метода, електромагнетско зрачење, снага сметњи, мешовита расподела, функција густине расподеле

UNCERTAINTY EVALUATION OF THE CONDUCTED EMISSION MEASUREMENTS

by

Aleksandar M. KOVAČEVIĆ [1*], ***Dejan D. DESPOTOVIĆ*** [2], ***Zoran A. RAJOVIĆ*** [3], ***Koviljka Dj. STANKOVIĆ*** [2], ***Ana V. KOVAČEVIĆ*** [4], ***and Uroš D. KOVAČEVIĆ*** [2]

[1]Technical Test Center, Ministry of Defense, Belgrade, Serbia
[2] Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia
[3] Electric Power Industry of Serbia, Belgrade, Serbia
[4] Faculty of Security Studies, University of Belgrade, Belgrade, Serbia

Scientific paper
DOI: 10.2298/NTRP1302182K

For the evaluation of measurement uncertainty in measuring the conduction emission, in this paper we propose a new model which uses mixed distribution. Evaluation of probability density function for the measurand has been done using Monte Carlo method and a modified least-squares method (combined method). In addition, the number of data n and the number of classes of histogram k which were used for simulation, were varied.

Key words: measurement uncertainty, conducted emission, probability density function, mixed distribution, modified least-squares method, Monte Carlo method

INTRODUCTION

The Guide to the Expression of Uncertainty in Measurement (GUM) [1] is the internationally accepted master document for the evaluation of uncertainty. In addition, for measurement uncertainty assigning for linear or linearized models, the GUM suggests one standard procedure which is known as the GUM uncertainty framework (GUF) [2]. The GUM Supplement 1 [2] is based on one general concept of propagating the probability density functions (PDF), where in order to obtain PDF for the measurand using of the Monte Carlo method (MCM) was suggested. Consequently, the law of propagation of uncertainties is based on a construction of a linear approximation of the model function [3].

One of the common problems that we face when examining electromagnetic compatibility (EMC) is an inconsistent approach to adjusting various specified or standardized tests. Consequently, some of the standardized EMC measurements are included within the precisely defined ways for evaluating uncertainty in measurement [4]. EMC tests and measurements typically have large uncertainties of at least several decibels [5]. Today, electronic equipment is considered to be the critical project element of armament and military equipment means and systems. So, for example, modern telecommunication devices are characterized, on one hand, by the great power of ultra broadband transmitters, and on the other, by sensitivity of receivers [6]. Many uncertainty sources in the domain of EMC measurements were not studied well and need further studying.

This paper presents a new model which uses mixed distribution for uncertainty evaluation of conducted emission measurement [7, 8]. Namely, evaluation of PDF for the measurand (output quantity) has been done using a MCM and a modified least-squares method. Consequently, the MCM required numerical calculation of approximate PDF values [7-10].

MODEL AND METHODS

Mixed distribution

As a model for determining the density function of a mixed distribution, two independent input quantities and one output quantity were taken (see fig. 1).

Consequently, each one of the input quantities is determined by expectation which is equal to the given estimate x_i, as well as to corresponding standard deviation which is equal to the given standard uncertainty $u(x_i)$. The output quantity (measurand) Y is determined by the best evaluation y, which is assigned by the standard uncertainty $u(y)$.

* Corresponding author; e-mail: aleksandarkovacevic1962@yahoo.com

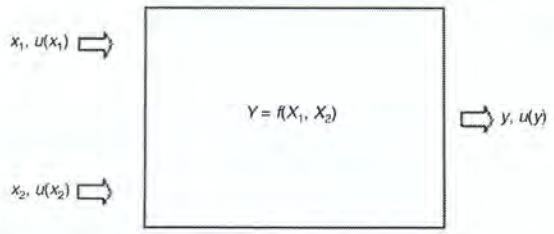

Figure 1. Illustration of the law of propagation of uncertainty for a linear model

Probability density function of the output quantity $f(x)$ (density function of a mixed normal-normal distribution) of two independent input quantities that are determined by two normal PDF is given with eq. (1)

$$f(x)=\varepsilon f_1(x)+(1-\varepsilon)f_2(x), \quad 0\le\varepsilon\le 1 \tag{1}$$

Consequently, probability density functions for the input quantities, $f_1(x)$ and $f_2(x)$, are given by eqs. (2) and (3), respectively

$$f_1(x)=\frac{1}{s_1\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-m_1}{s_1}\right)^2\right] \quad -\infty<x<+\infty,\ s_1>0 \tag{2}$$

$$f_2(x)=\frac{1}{s_2\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-m_2}{s_2}\right)^2\right] \quad -\infty<x<+\infty,\ s_2>0 \tag{3}$$

where m_1 and s_1 are the parameters that represent the mean and standard deviation of the first normal distribution, respectively, and m_2 and s_2 – the parameters that represent the mean and standard deviation of the second normal distribution. ε is the mixed coefficient of these distributions which indicates the proportion of the first distribution in the mixed distribution. The value of this coefficient is an interval from 0 to 1, $\varepsilon\in(0, 1)$, so that the proportion of the second distribution in the mixed distribution equals $1-\varepsilon$.

In the fig. 2 the propagation of distributions for two independent input quantities assigned by the normal distributions are illustrated.

The mixed normal-normal distribution function for the output quantity, $F(x)$, is given with eq. (4)

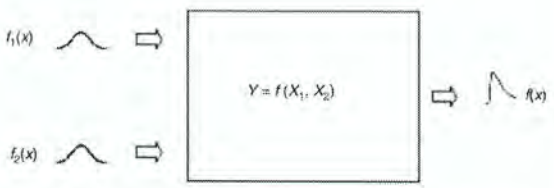

Figure 2. Illustration of the propagation of distributions

$$F(x)=\varepsilon F_1(x)+(1-\varepsilon)F_2(x), \quad 0\le\varepsilon\le 1 \tag{4}$$

Consequently, distribution functions for the input quantities, $F_1(x)$ and $F_2(x)$, are given by eqs. (5) and (6), respectively

$$F_1(x)=\frac{1}{s_1\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{x-m_1}{s_1}\right)^2\right] \quad -\infty<x<+\infty,\ s_1>0 \tag{5}$$

$$F_2(x)=\frac{1}{s_2\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{x-m_2}{s_2}\right)^2\right] \quad -\infty<x<+\infty,\ s_2>0 \tag{6}$$

Pseudorandom numbers that belong to the first and the second normal distribution are determined by the eqs. (7) and (8), respectively

$$x_i'=m_1+s_1u_i', \quad i=1,2,\ldots,N_1 \tag{7}$$

$$x_j'=m_2+s_2u_j', \quad j=1,2,\ldots,N_2 \tag{8}$$

where N_1 and N_2 are the pseudorandom numbers total of the first and the second normal distribution, respectively, r_i, r_j – the pseudorandom numbers, $r_i, r_j\in(0, 1)$, and u_i', u_j' – the lower quantiles of standardized normal distribution, $N(0, 1)$.

The lower quantile of standardized normal distribution, u_i', that has the mean equal to 0 and the standard deviation equal to 1, it is determined by generating the pseudorandom number $r_i\in(0, 1)$, which represents the lower quantum $r_i=F_1(u_i')$, and then the appropriate value for u_i' is determined. For determining the values for u_i' special subprogram was used here, and they can be found in statistic tables, *i. e.* the inverse function of standardized normal distribution [11].

The lower quantile of standardized normal distribution, u_j', is obtained in an identical way using the pseudorandom number $r_j\in(0, 1)$, which represents the lower quantum $r_j=F_2(u_j')$.

When the obtained values of pseudorandom numbers x_i' and x_j' are mixed, *c. f.* eqs. (7) and (8), a mixed normal-normal distribution which has n values (n – pseudorandom numbers total of the mixed distribution, $n=N_1+N_2$) is obtained.

Density function of a mixed normal-rectangular distribution is given with the eq. (1), and the appropriate probability density functions for the input quantities, $f_1(x)$ and $f_2(x)$, are given by eqs. (9) and (10), respectively

$$f_1(x)=\frac{1}{s\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right] \quad -\infty<x<+\infty,\ s>0 \tag{9}$$

$$f_2(x)=\frac{1}{b-a}, \quad a\le x\le b,\ b>a \tag{10}$$

where m and s is the parameters that represent the mean and standard deviation of the normal distribution, respectively, and a and b are the parameters that represent lower and upper limits of a rectangular distribution.

As in the previous case, mixed normal-rectangular distribution function, $F(x)$, is given by eq. (4). Consequently, distribution functions for the input quantities, $F_1(x)$ and $F_2(x)$, are given by eqs. (11) and (12), respectively

$$F_1(x)=\frac{1}{s\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right] \quad -\infty<x<+\infty,\quad s>0 \tag{11}$$

$$F_2(x)=\frac{x}{b-a}-\frac{a}{b-a},\quad a\le x\le b,\ b>a \tag{12}$$

Pseudorandom numbers that belong to the first normal distribution and the second rectangular distribution are determined by eqs. (13) and (14), respectively

$$x_i'=m+su_i',\quad i=1,2,\ldots,N_1 \tag{13}$$

$$x_j'=a+(b-a)r_j,\quad j=1,2,\ldots,N_2 \tag{14}$$

When the obtained values of pseudorandom numbers x_i' and x_j' are mixed, *c. f.* eqs. (13) and (14), a mixed normal-rectangular distribution which has n values (n – pseudorandom numbers total of the mixed distribution, $n = N_1 + N_2$) is obtained.

Density function of a mixed normal-triangular distribution is given by eq. (1), and the corresponding probability density functions for the input quantities, $f_1(x)$ and $f_2(x)$, are given by eqs. (15) and (16), respectively

$$f_1(x)=\frac{1}{s\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right] \quad -\infty<x<+\infty, s>0 \tag{15}$$

$$f_2(x)=\begin{cases}\dfrac{4(x-a)}{(b-a)^2}, & a\le x\le c,\\ \dfrac{4(b-x)}{(b-a)^2}, & c<x\le b,\end{cases} \tag{16}$$

where m and s are the parameters which represent the mean and standard deviation of the normal distribution, respectively, a and b – the parameters which represent the lower and upper limits of a symmetric triangular distribution, and $c = (a + b)/2$ is the parameter mode of the symmetric triangular distribution.

Mixed normal-triangular distribution function, $F(x)$, is given by eq. (4). Consequently, the distribution functions for the input quantities, $F_1(x)$ and $F_2(x)$, are given by eqs. (17) and (18), respectively,

$$F_1(x)=\frac{1}{s\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right] \quad -\infty<x<+\infty,\quad s>0 \tag{17}$$

$$F_2(x)=\begin{cases}\dfrac{2(x-a)^2}{(b-a)^2}, & a\le x\le c,\\ 1-\dfrac{2(b-x)^2}{(b-a)^2}, & c<x\le b,\end{cases} \tag{18}$$

Pseudorandom numbers that belong to the first normal distribution and the second triangular distribution are determined in the eqs. (19) and (20), respectively

$$x_i'=m+su_i',\quad i=1,2,\ldots,N_1 \tag{19}$$

$$x_j'=\begin{cases}a+(b-1)\sqrt{\dfrac{r_j}{2}}, & 0<r_j\le 0.5,\\ b-(b-a)\sqrt{\dfrac{1-r_j}{2}}, & 0.5<r_j<1,\end{cases} \quad j=1,2,\ldots,N_2 \tag{20}$$

When the obtained values of pseudorandom numbers x_i' and x_j' are mixed (cf eqs. (19) and (20)), a mixed normal- triangular distribution which has n values (n – pseudorandom numbers total of the mixed distribution, $n = N_1 + N_2$) is obtained.

Point estimates parameters of a mixed distribution

Point estimates parameters of the probability density function for the output quantity (density function of a mixed distribution) are obtained by the combined method which consists of the MCM and the modified least-squares method [7, 8]. Consequently, the MCM can be used to approximate the PDF for the output quantity [12-16].

Also, point estimates parameters of density function of a mixed distribution can be obtained in the analytical procedure, but it is more complicated, and sometimes difficult to execute.

Point estimates parameters of density function of a mixed normal-normal distribution are determined using eqs. (21)-(24)

$$\hat{m}_{1i}=1.05x_{\min}+0.95(x_{\max}-x_{\min})r_{m1;i},\quad i=1,2,\ldots,N \tag{21}$$

$$\hat{s}_{1i}=0.1s+1.25sr_{s_1;i}\quad i=1,2,\ldots,N \tag{22}$$

$$\hat{m}_{2j}=1.05x_{\min}+0.95(x_{\max}-x_{\min})r_{m2;j},\quad j=1,2,\ldots,N \tag{23}$$

$$\hat{s}_{2j}=0.1s+1.25sr_{s_2;j}\quad j=1,2,\ldots,N \tag{24}$$

where x_{min} and x_{max} are the minimum and maximum values which were taken by the random variable x of a mixed distribution, respectively, s is the empirical standard deviation of a mixed distribution, N – the total number of trials (iterations), $r_{m_1;i}, r_{s_1;i}, r_{m_2;j}, r_{s_2;j} \in (0, 1)$ – the pseudorandom numbers, respectively.

Mixed coefficient is determined using eq. (25)

$$\varepsilon_l = r_{\varepsilon;l}, \quad l = 1, 2, \ldots, N \tag{25}$$

where $r_{\varepsilon;l} \in (0, 1)$ is the pseudorandom number.

Point estimates parameters expressions of mixed distributions, eqs. (21)-(25), are obtained by experimental and numerical calculations, in order to obtain more realistic intervals for the parameters of distribution, and corresponding to the real situation.

The procedure of determining point estimates parameters of density function of a mixed normal-normal distribution consists of several steps.

First, one determines minimum and maximum values, x_{min} and x_{max}, which were taken by the random variable x of a mixed normal-normal distribution *c. f.* eqs. (7) and (8), and then, according to the generated pseudorandom numbers $r_{m_1;i}, r_{s_1;i}, r_{m_2;j}, r_{s_2;j}$, and $r_{\varepsilon;l}$, respectively, which belong to the interval (0, 1) and using eqs. (21)-(25), the i-th, j-th, and l-th values are determined for these parameters. These values are put in eqs. (1)-(3), and function values of a mixed normal-normal distribution, $f(x_j)$ – estimated density function of mixed normal-normal distribution, are determined for every value of random variable x which belongs to midpoint of each class histogram. If histogram has k classes, then there are k values of this density function of a mixed distribution. Also, the empirical values of this function are determined for each one of these classes using eq. (26)

$$f_j = \frac{n_j}{nh}, \quad j = 1, 2, \ldots, k \tag{26}$$

where n_j is the number of values which fall under j-th class of histogram, n – the total number of values which were taken by the random variable x, and h – the class width of histogram. The function f_j in eq. (26) is the empirical density function of mixed normal-normal distribution that describes the histogram of random variable x. Class width of histogram h is determined according to eq. (27)

$$h = \frac{x_{max} - x_{min}}{k} \tag{27}$$

The number of classes of histogram k are determined according to eqs. (28)-(30), respectively,

$$k_1 = \sqrt{n} \tag{28}$$

$$k_2 = 1 + \frac{1}{2}\sqrt{n} \tag{29}$$

$$k_3 = [1 + 3.3 \log_{10} n] \tag{30}$$

Consequently, that k_i, $i = 1, 2, 3$, is taken as an integer value [11].

When the midpoint of each class histogram, x_j, values f_j and $f(x_j)$ are determined, then one determines the sum of the squared deviations of these functions according to classes, using the eq. (31)

$$S_i = \sum_{j=1}^{k} [f(x_j) - f_j]^2, \; i = 1, 2, \ldots, N \tag{31}$$

where N is the total number of trials (iterations).

When the first value of the S_1 sum is determined, then for the second generated pseudorandom numbers $r_{m_1;i}, r_{s_1;i}, r_{m_2;j}, r_{s_2;j}$ and $r_{\varepsilon;l}$, respectively, the same procedure determines the parameters of the density function of mixed normal-normal distribution and the sum S_2, and then, the values of the sums are compared. The parameters of density function of a mixed normal-normal distribution, are determined in this way randomly, the ones retained are those in which the minor sum of the squared deviations of these functions is obtained. The procedure continues and the new sum S is determined, and the parameters with which this sum is the least, are retained. The procedure must be repeated from 10^5 to 10^6 times (a value of N), and even more.

The procedure for determining point estimates parameters of density function of a mixed normal-rectangular distribution and a mixed normal-triangular distribution is described in the previous procedure which refers to density function of a mixed normal-normal distribution [7, 8].

EVALUATION OF MEASUREMENT UNCERTAINTY

Measurement model

This paper observes conducted emissions measurements in power leads of military telecommunication devices in Faraday cage according to the method CE102 from the standard MIL-STD-461E [17].

For determining of the measurand value, the standardized measurement method is used [17]. Consequently, the possibility of variations of obtained measurand values becomes smaller, which influences the reducing of measurement uncertainty.

Equation model for the evaluation of the Measurement Instrumentation Uncertainty – MIU is given in [18], in the eq. (32)

$$V = V_r + L_c + L_{LISN} + \delta V_{sw} + \delta V_{pa} + \delta V_{pr} + \\ + \delta F_{step} + \delta Z + \delta M \tag{32}$$

Equation (32) represents one purely additive linear model, whose terms are independent. Information about terms of an expression in the model equation is given in tab. 1. Measurement uncertainty comprises, in general, many components. Some of these may be

Table 1. Uncertainty budget according to the GUM for the conducted emission measurements

Input quantity	X_i	Estimate X_i (x_i)		Standard uncertainty $u(x_i)$ [dB]	Sensitivity coefficient c_i	Contribution to the standard uncertainty $u_i(y) = c_i u(x_i)$
		Value [dB]	Probability distribution function			
Receive reading	V_r	±0.1	Rectangular $k_p = 1.732$	0.058	1	0.058
LISN-receiver attenuation	L_c	±0.1	Normal $k_p = 2.000$	0.050	1	0.050
LISN voltage division factor	L_{LISN}	±0.2	Normal $k_p = 2.000$	0.100	1	0.100
Receiver sine wave voltage	δV_{sw}	±1.0	Normal $k_p = 2.000$	0.500	1	0.500
Receiver pulse amplitude response	δV_{pa}	±2.0	Rectangular $k_p = 1.732$	1.155	1	1.155
Receiver pulse repetition rate response	δV_{pr}	±2.0	Rectangular $k_p = 1.732$	1.155	1	1.155
Frequency step error	δF_{step}	±0.0	Rectangular $k_p = 1.732$	0.000	1	0.000
LISN impedance	δZ	+2.60 −2.70	Triangular $k_p = 2.449$	1.082	1	1.082
LISN-receiver mismatch	δM	+0.676 −0.734	U-shaped $k_p = 1.414$	−0.519	1	−0.519

evaluated by Type A evaluation of measurement uncertainty from the statistical distribution of the quantity values from series of measurements and can be characterized by standard deviations. The other components, which may be evaluated by Type B evaluation of measurement uncertainty, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information.

Values of input quantities

Table 1 presents uncertainty budget, a Type B evaluation for the case of conducted emission measurements. The given data are obtained from the manufacturer's specifications and calibration certificates, and they are used for the evaluation of the Measurement Instrumentation Uncertainty, according to the ISO-GUM [1].

This also constitutes Type A evaluation, but this paper will not consider the contribution of type A evaluation.

The standard uncertainty $u(x_i)$ is calculated by dividing the value of the uncertainty associated with x_i by the coverage factor k_p, whose value depends on the choice of PDF and confidence level which is associated to the given value.

Evaluation of the measurement instrumentation uncertainty

The previous section presents the uncertainty budget according to the GUM for the case of conducted emission measurement according to the method CE102 from the standard MIL-STD-461E [17]. Consequently, the model equation for the evaluation of the MIU is given with eq. (32). The MCM and the modified least-squares method (combined method) are applied in three cases for two independent input quantities from the given expression. Namely, the combined method is used for the evaluation of probability density function for the output quantity (mixed distribution) according to probability density function from two independent input quantities, and *i. e.*: two independent input quantities assigned by the normal distributions, two independent input quantities where the first quantity is assigned a normal distribution and the second is assigned a rectangular distribution, two independent input quantities where the first quantity is assigned a normal distribution and the second is assigned a triangular distribution [7-10].

The Monte Carlo simulations for obtaining mixed distributions are done by the procedure which was described in the previous sections. In addition, the number of classes of histogram k_i, $i = 1, 2, 3$, which was used for simulation is varied and the values are used from tab. 1. A value of N, the total number of trials was 10^6. The number of data which was used for simulation is $n = 5000$. Risk conformity was $\alpha = 0.05$, that is the confidence level $(1 - \alpha)$ was 0.95. One more data that is important for our simulation was the mixed coefficient ε which was 0.5. The results obtained by the combined method are compared to the corresponding results when applying the GUM.

In figs. 3-5 a mixed normal-normal distribution (the first example) is shown. In addition, the estimated values are $x_1 = x_2 = 0$ dB, and standard uncertainties $u(x_1) = 0.05$ dB and $u(x_2) = 0.1$ dB, respectively, (see tab. 1).

It is noticeable that the estimated curve fitting (line 2) in histogram (line 3) is very good, regardless of the choice of k_i (the number of histogram classes), which indicates that the unknown parameters of this distribution are estimated well. It should be men-

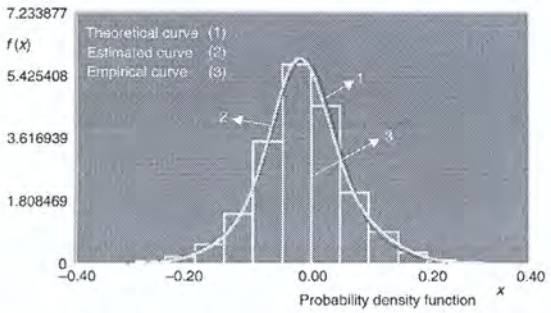

Figure 3. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_3 and the number of data which was used for simulation is $n = 5000$

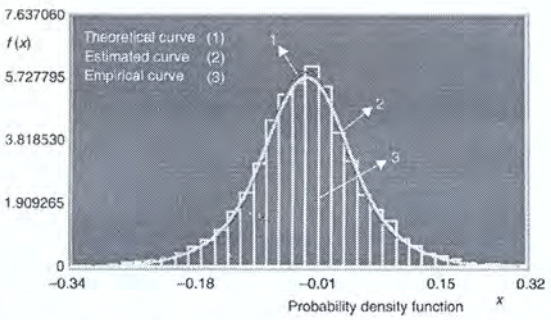

Figure 4. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_2 and the number of data which was used for simulation is $n = 5000$

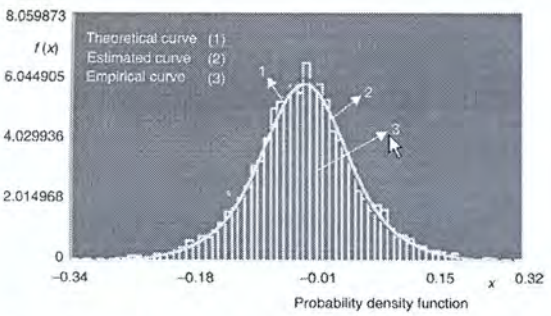

Figure 5. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_1 and the number of data which was used for simulation is $n = 5000$

tioned, that coming of the curve through the mid-point of each class histogram is considered to be the best fitting. Also, it is evident that the estimated curve (line 2) differs slightly from the theoretical curve (line 1). Consequently, the theoretical curve represents the results obtained according to the GUM. This difference was the evaluation result of the mixed distribution parameters (whose values are pseudorandom) and the number N of iterations (the total number of trials).

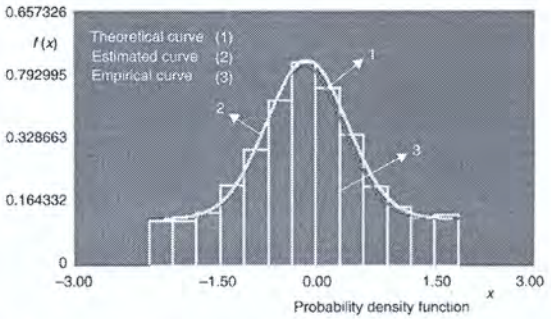

Figure 6. Mixed normal-rectangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_3 and the number of data which was used for simulation is $n = 5000$

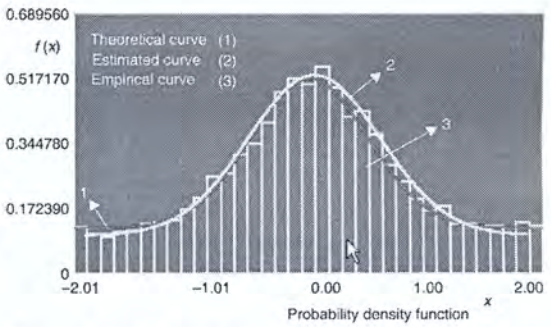

Figure 7. Mixed normal-rectangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_2 and the number of data which was used for simulation is $n = 5000$

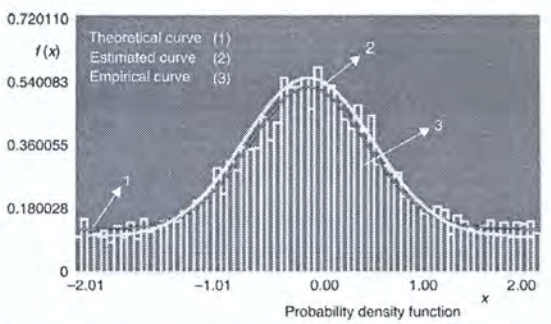

Figure 8. Mixed normal-rectangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_1 and the number of data which was used for simulation is $n = 5000$

In figs. 6-8 a mixed normal-rectangular distribution (the second example) is shown. In addition, the estimated values are $x_1 = x_2 = 0$ dB, and standard uncertainties $u(x_1) = 0.5$ dB and $u(x_2) = 1.155$ dB, respectively, (see tab. 1).

It is noticeable that fitting of the estimated curve (line 2) in histogram (line 3) is very well and does not deviate a lot from the theoretical curve (line 1), regardless of the choice of k_i. This difference was the result of evaluation of parameters of the mixed distribution

(whose values are pseudorandom) and the number N of iterations.

In figs. 9-11 a mixed normal-triangular distribution (the third example) is shown. In addition, the estimated values are $x_1 = 0$ dB and $x_2 = -0.05$ dB, and the standard uncertainties $u(x_1) = 0.5$ dB and $u(x_2) = 1.082$ dB, respectively (see tab. 1).

It is noticeable that fitting of the estimated curve in a histogram is very well and does not deviate a lot from the theoretical curve, regardless of the choice k_i.

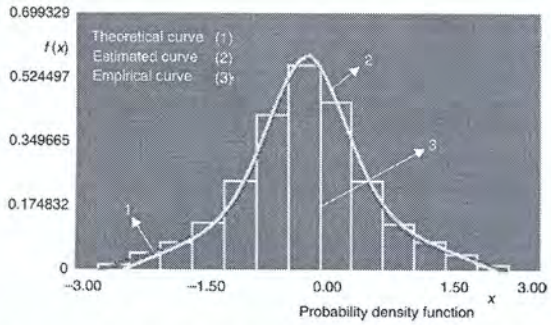

Figure 9. Mixed normal-triangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_3 and the number of data which was used for simulation is $n = 5000$

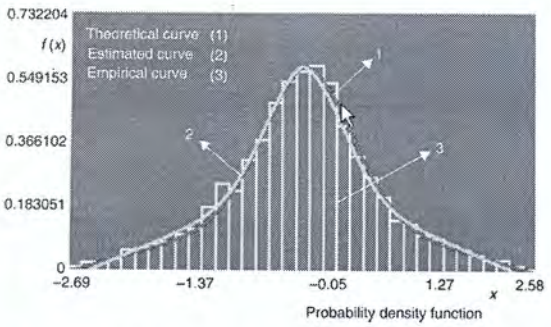

Figure 10. Mixed normal-triangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_2 and the number of data which was used for simulation is $n = 5000$

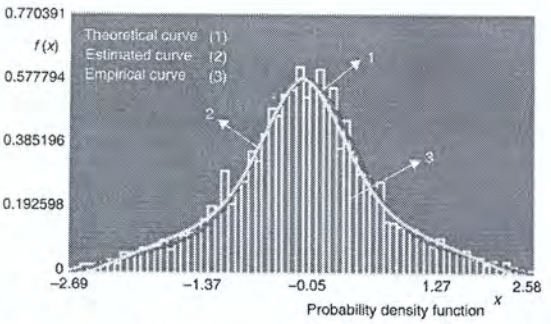

Figure 11. Mixed normal-triangular distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_1 and the number of data which was used for simulation is $n = 5000$

This difference was the result of evaluation of parameters of the mixed distribution (whose values are pseudorandom) and the number N of iterations.

In the figs. 12-14 a mixed normal-normal distribution where the number of data which was used for simulation is $n = 10000$ is shown. In addition, the estimated values are $x_1 = x_2 = 0$ dB, and standard uncertainties $u(x_1) = 0.05$ dB and $u(x_2) = 0.1$ dB, respectively (see the first example).

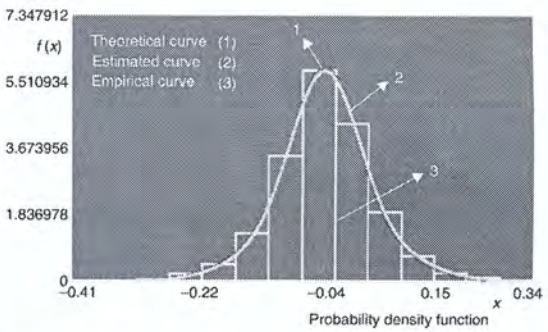

Figure 12. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_3 and the number of data which was used for simulation is $n = 10000$

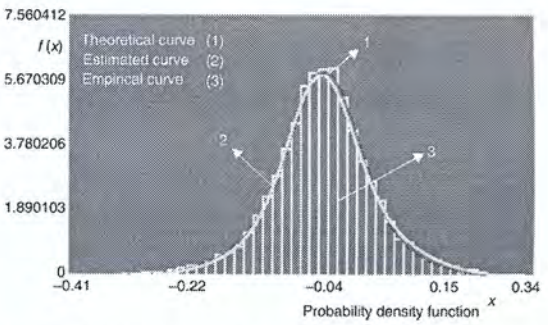

Figure 13. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_2 and the number of data which was used for simulation is $n = 10000$

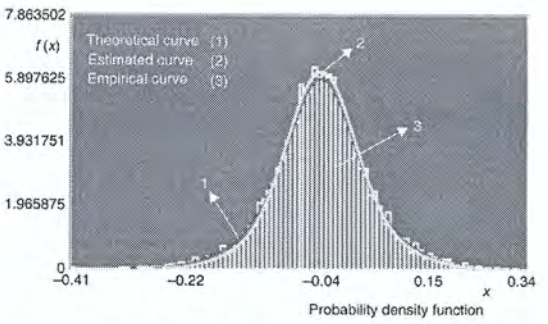

Figure 14. Mixed normal-normal distribution obtained by the combined method and GUM, respectively. The number of classes of histogram is k_1 and the number of data which was used for simulation is $n = 10000$

As in the first example, it is shown that fitting of the estimated curve in histogram (the empirical curve) is very good and differs slightly from the theoretical curve, regardless of the choice k_i and the increasing number of data n. This difference was the result of evaluation of parameters of the mixed distribution (whose values are pseudorandom) and the number N of iterations.

CONCLUSIONS

In this paper the Monte Carlo method (MCM) and the modified least-squares method, are presented for the estimate of the output quantity (measurand), which is interrelating with two input quantities. As a representative equation model for the evaluation, of the Measurement Instrumentation Uncertainty is used, and it is also used for the conducted emission measurement according to the method CE102 from the standard MIL-STD-461E. The MCM and the modified least-squares method (combined method) are applied in three cases, for two independent input quantities, which were associated with PDF. In addition, the number of data n and the number of classes of histogram k, which were used for simulation, are varied. It was shown that the combined method gives valid results with regard to physical accuracy of the presented model, which refers to the input and output quantities. Namely, applying the combined method produces a mixed distribution, *i. e.* PDF for the output quantity, which fits well (the estimated curve) in histograms and differs slightly from the produced results according to the GUM (the theoretical curve), regardless to the choice of k and n.

ACKNOWLEDGEMENT

The Ministry of Education, Science and Technological Development of the Republic of Serbia supported this work under Contracts III 43009 and 171007.

AUTHOR CONTRIBUTIONS

Theoretical analysis and experiments were carried out by A. M. Kovačević. Literature research was carried out by A. M. Kovačević, K. Dj. Stanković, and A. V. Kovačević. The manuscript was written by A. M. Kovačević, with the suggestion of all authors. The figures were prepared by A. M. Kovačević. All authors analysed and discussed the results.

REFERENCES

[1] ***, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, Geneva. Guide to the Expression of Uncertainty in Measurement, 1995 (ISBN 92-67-10188-9)
[2] ***, JCGM, Evaluation of Measurement Data-Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'-Propagation of Distributions Using a Monte Carlo Method, 2008
[3] Wubbeler, G., Krystek, M., Elster, C., Evaluation of Measurement Uncertainty and its Numerical Calculation by a Monte Carlo Method, *Meas. Sci. Technol., 19* (2008), 8, 084009
[4] ***, IEC/CISPR, Geneva, Standard CISPR 16-4-1: Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – Part 4-1: Uncertainties, Statistics and Limit Modeling – Uncertainties in Standardized EMC Tests, 2005
[5] ***, NIS 81, The Treatment of Uncertainty in EMC Measurements, NAMAS Executive: National Physical Laboratory, UK, 1994
[6] Paul, C. R., Introduction to Electromagnetic Compatibility, 2nd ed., John Wiley, New York, USA, 2006
[7] Kovačević, A., Brkić, D., Osmokrović, P., Evaluation of Measurement Uncertainty Using Mixed Distribution for Conducted Emission Measurements, *Measurement, 44* (2011), 4, pp. 692-701
[8] Kovačević, A., Evaluation of Measurement Uncertainty in EMC Testing, Ph. D. dissertation (in Serbian), Faculty of Electrical Engineering, University of Belgrade, Belgrade, 2011
[9] Vujisić, M., Stanković, K., Osmokrović, P., A Statistical Analysis of Measurement Results Obtained from Nonlinear Physical Laws, *Applied Mathematical Modeling, 35* (2011), 7, pp. 3128-3135
[10] Stanković, K., *et al.*, Statistical Analysis of the Characteristics of Some Basic Mass-Produced Passive Electrical Circuits Used in Measurements, *Measurement, 44* (2011), 9, pp. 1713-1722
[11] Chapouille, P., De Pazzis, R., Reliability Systems (in French), Masson et Cie, Paris, 1968
[12] Stanković, K., Influence of the Plain-Parallel Electrode Surface Dimensions on the Type A Measurement Uncertainty of GM Counter, *Nucl Technol Radiat, 26* (2011), 1, pp. 39-44
[13] Dolićanin, Č., *et al.*, Statistisal Treatment of Nuclear Counting Results, *Nucl Technol Radiat, 26* (2011), 2, pp. 164-170
[14] Vujisić, M., *et al.*, Simulated Effects of Proton and Ion Beam Irradiation on Titanium Dioxide Memristors, *IEEE Transactions on Nuclear Science, 57* (2010), 4, pp. 1798-1804
[15] Lazarević, Dj., *et al.*, Radiation Hardness of Indium Oxide Films in the Cooper-Pair Insulator State, *Nucl Technol Radiat, 27* (2012), 1, pp. 40-43
[16] Stanković, S., *et al.*, MSV Signal Processing System for Neutron-Gamma Discrimination in a Mixed Field, *Nucl Technol Radiat, 27* (2012), 2, pp. 165-170
[17] ***, U. S. Department of Defense. Standard MIL-STD-461E: Requirements for the Control of Electromagnetic Interference Characteristics of Subsystems and Equipment, 1999
[18] ***, IEC/CISPR, Geneva, Standard CISPR 16-4-2: Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – part 4-2: Uncertainties, Statistics and Limit Modeling – Uncertainty in EMC Measurements, 2003

Received on January 15, 2013
Accepted on June 10, 2013

Александар М. КОВАЧЕВИЋ, Дејан Д. ДЕСПОТОВИЋ, Зоран А. РАЈОВИЋ
Ковиљка Ђ. СТАНКОВИЋ, Ана В. КОВАЧЕВИЋ, Урош Д. КОВАЧЕВИЋ

ПРОЦЕНА МЕРНЕ НЕСИГУРНОСТИ ПРИ МЕРЕЊУ КОНДУКЦИОНЕ ЕМИСИЈЕ

У раду је предложен један нови модел за процену мерне несигурности при мерењу кондукционе емисије, који користи мешовиту расподелу. Монте Карло метода и модификована метода најмањих квадрата (комбинована метода) коришћене су за процену функције густине расподеле мерне величине. При томе, вариран је број података n и број класа хистограма κ који су коришћени за симулацију.

Кључне речи: мерна несигурност, кондукциона емисија, функција густине расподеле, мешовита расподела, модификована метода најмањих квадрата, Монте Карло метода