CART Method for Modeling the Output Power of Copper Bromide Laser

Iliycho P. Iliev, Desislava S. Voynikova, and Snezhana G. Gocheva-Ilieva

Abstract—This paper examines the available experiment data for a copper bromide vapor laser (CuBr laser), emitting at two wavelengths - 510.6 and 578.2nm. Laser output power is estimated based on 10 independent input physical parameters. A classification and regression tree (CART) model is obtained which describes 97% of data. The resulting binary CART tree specifies which input parameters influence considerably each of the classification groups. This allows for a technical assessment that indicates which of these are the most significant for the manufacture and operation of the type of laser under consideration. The predicted values of the laser output power are also obtained depending on classification. This aids the design and development processes considerably.

Keywords—Classification and regression trees (CART), Copper Bromide laser (CuBr laser), laser generation, nonparametric statistical model.

I. INTRODUCTION

ONE of the practical approaches when studying the behavior of a particular complex technical system is the use and retrieval of the available experiment data. These data contain significant information related to the working processes and the relationships between individual components of the real system - technical parameters (dimensions, construction materials), operating parameters, service life, etc. With the help of suitable statistical methods and techniques, these relationships can be expressed, the functioning of the system can be described and analyzed, and its future behavior and development predicted.

This paper examines a copper bromide vapor laser which is a type of metal vapor laser in the visible zone. It is considered that the laser in question has been studied in detail but due to its specific characteristics and wide range of applications, it continues to be the subject of scientific and commercial interest, and therefore, development [1], [2]. Alongside engineering design, the modeling (analytical, numerical, statistical, simulating or other types) of devices or technical systems is also widely applied in practice. Standard mathematical modeling includes systems of differential and integral equations, and other mathematical methods, describing the system and allowing the calculation of solutions for the processes, occurring within the system under investigation, as well as the performance of simulations. Here, the most widely used types of models are kinetic models. These describe the particles and processes occurring in the operating laser medium. There is a large number of such publications for metal vapor lasers, including for copper bromide vapor lasers, see for instance [3]–[5]. Although kinetic models describe the major processes within the laser medium and the interactions between particles using hundreds of equations, a general drawback of theirs is that they cannot provide a complex direct estimate of output characteristics such as the average output power, laser efficiency, service life, etc.

During the last few years, models were developed and applied on the basis of accumulated experiment data, providing models of statistical relationships, dependencies, and classifications of basis laser parameters, for which experiment data is available. Traditional parametric models of metal vapor lasers have been developed and analyzed in [6]–[10]. Multivariate regression with principal components analysis, hierarchical cluster analysis, factor analysis, and other statistical techniques have been used. A non-linear model of output power has been built in [11]. Nonparametric models were obtained using the Multivariate Adaptive Regression Splines (MARS) method in [6], [11]. In the recent paper [12], the models describe over 98% of experiment data with a relative accuracy comparable to that of measurements, making it possible to predict the output power of future lasers.

In this paper, another powerful nonparametric modeling method - CART (Classification and Regression Trees) - is applied to available data for a copper bromide vapor laser. This method allows the separation of all observations from the considered independent variables in non-interacting groups in the form of a binary tree according to the degree of influence on the dependent variable. In this study the dependent variable is the laser output power.

General objective of this study is to classify and determine the influence of 10 input laser characteristics (supplied power, geometric design of the tube, neon pressure, reservoir temperature, etc.) on the average output power based on available experiment data. For the first time, the powerful nonparametric technique CART, described in [13]-[15], is applied for data of metal vapor lasers. CART is one of the basic data mining algorithms and is widely used for either classification or estimation problems (like regression) [15].

I. P. Iliev is with the Technical University – Sofia, branch Plovdiv, Plovdiv, BULGARIA (phone: 00359-32-265-842; e-mail: iliev55@ abv.bg).

D. S. Voynikova and S. G. Gocheva-Ilieva are with the Faculty of Mathematics and Informatics, "Paisii Hilendarski" University of Plovdiv, Plovdiv, BULGARIA (e-mails: voinikova@uni-plovdiv.bg and snow@uni-plovdiv.bg).

In this study the following basic problems are solved: (1) building an optimal solution regression tree using CART; (2) determining the adequate linear model on the basis of this tree; (3) using the model to estimate known experiments; (4) validation of model; (5) comparison of results to previous parametric and nonparametric models of the same type of laser.

The obtained model describes about 97% of the available data and demonstrates very good predictive qualities. This approach is oriented to guide the construction and design of new copper bromide vapor lasers with increased output power.

The statistical study has been carried out using the CART software package [16].

II. OBJECT OF STUDY

The copper bromide vapor laser is an improved version of a pure copper vapor laser. It is the most powerful and effective laser in the visible spectrum demonstrating high coherence and convergence of the laser beam. We are investigating variations of this laser, invented and developed at the Laboratory of Metal Vapor Lasers at the Georgi Nadjakov Institute of Solid State Physics of the Bulgarian Academy of Sciences, Sofia. The first patents related to this type of laser are [17], [18]. The copper bromide vapor laser is one of the 12 laser sources which have a wide range of applications and are commercially viable [1], [2]. The development and improvement of CuBr lasers is seen as a fundamental step in the study of copper lasers as a whole.

Copper bromide vapor lasers are sources of pulse radiation in the visible spectrum (between 400 and 720nm) emitting at two wavelengths: green 510.6nm and yellow - 578.2nm. They are considered to be high-pulse lasers. Neon is used as a buffer gas. In order to improve efficiency, small quantities of hydrogen are added. Unlike the high-temperature pure copper vapor laser, the copper bromide vapor laser is a low-temperature one, with an active zone temperature of about 500°C [1]. The laser tube is made out of quartz glass without high-temperature ceramics as a result of which it is significantly cheaper and easier to manufacture. The discharge is heated by electric current (self-heating laser).

It produces light impulses tens of nanoseconds long. Its main advantages are: short initial heating period, stable laser generation, relatively long service life, high values of output power and laser efficiency.

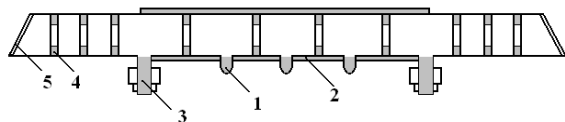A simple scheme of the laser tube is given in Fig. 1.

Fig. 1 Scheme of the laser tube of a copper bromide vapor laser: 1 – copper bromide reservoirs, 2 – heat insulation of the active volume, 3 – porous copper electrodes, 4 – inner rings, 5 – mirrors.

III. DATA

This paper takes into account the following 10 independent input variables (predictors) and one dependent variable (response) - laser output power $Pout$ (W). The independent variables are: D (mm) – inner diameter of the laser tube, DR (mm) – inner diameter of the ring (without rings, $D=DR$), L(cm) – length of the active zone (distance between the electrodes), PIN (kW) – electric power supplied to the discharge, PL (kW/cm) - electric power per unit length with 50% losses, PRF (kHz) – electric pulse repetition frequency, PNE (torr) – buffer gas pressure (neon), $PH2$ (torr) – pressure of the added gas (hydrogen), C (nF) – equivalent capacity of the condensation battery, TR (^{o}C) – temperature of copper bromide reservoirs.

The study uses the values of these variables taken from $n = 387$ experiments, published in [19]–[26]. It needs to be noted that the maximum output power achieved is $Pout=120$ W in an experiment where the following values were measured for the input parameters as given above: (58, 58, 200, 5, 12.5, 0.6, 17.5, 20, 1.3, 490) [25].

The statistical summary for the whole dataset is given in Table I.

It should be noted that all the variables are not normally distributed, which follows from the values of asymmetry (skewness) and kurtosis. Especially, from Table I it can be seen that the absolute values of the coefficients of skewness for 5 variables (PIN, PRF, PNE, C, TR) are very high. Similar is observed for the coefficients of kurtosis, where 8 of the variables, namely $D, DR, L, PH2, PRF, PNE, C,$ and TR are too high. As a rule, following [27] the inequality

$$\frac{|Skewness|}{Std.Deviation\ of\ Skewness} > 2.5$$

can be used as a somewhat arbitrary guideline that distribution is markedly skewed and it would be prudent to use nonparametric statistics. This is the case for all variables of experiments described the investigated CuBr laser. However, different transformations of our data also showed that their distributions are quite different from the normal.

For this reason, nonparametric methods, including data mining methods (CART, MARS, etc.) which have no requirements towards the type of data distribution, both as a whole and for subsets, are more suitable [15], [28].

IV. SHORT DESCRIPTION OF THE CART METHOD

The CART method algorithm, as indicated by the name, solves the classification and regression problem. It was developed between 1974-1984 by Leo Breiman, Jerry Friedman, Charles Stone and Richard Olshen [13].

CART is a nonparametric solution tree technique which builds classification or regression trees depending on whether the dependent variable is categorical or numerical. In our case, this is a classification and regression tree.

TABLE I
DESCRIPTIVE STATISTICS OF THE DATASET OF CuBr LASER

	Min	Max	Mean	Std. D.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
D	15.00	58.00	46.59	10.072	-0.809	0.12	1.451	0.25
DR	4.50	58.00	34.83	18.31	0.265	0.12	-1.602	0.25
PIN	1.00	5.00	2.10	1.27	1.065	0.12	-0.321	0.25
L	30.00	200.00	106.59	70.70	0.478	0.12	-1.670	0.25
PL	5.00	16.67	10.92	2.51	-0.467	0.12	0.183	0.25
PH2	0.00	0.80	0.36	0.25	-0.416	0.12	-1.430	0.25
PRF	3.20	125.50	23.24	25.69	3.589	0.12	11.530	0.25
PNE	8.00	250.00	22.56	24.17	6.389	0.12	46.454	0.25
C	0.33	4.00	1.33	0.61	2.313	0.12	6.233	0.25
TR	350.00	590.00	478.22	23.25	-1.673	0.12	7.332	0.25
Pout	0.25	120.00	34.024	35.57	0.808	0.12	-0.862	0.25
Valid N	387							

The algorithm is intended for the building of a binary solutions tree. The output set of observations is divided into groups at the end nodes (leaves) of the tree. The goal is to find a tree which allows for a good distribution of the data with the lowest possible relative error of prediction. Each branch of the tree ends with a terminal node and each observation falls into exactly one end node, defined by a unique set of rules.

More specifically, the objective of the regression tree approach is to distribute the data in relatively homogeneous (small standard deviation) end nodes and to obtain a mean observed value at each node in the form of a predicted value. The building of a tree starts from a parent node, containing all observations. At each step (at each running node) a rule is applied to divide the set of observations within the node into two subsets according to some condition preset for the current independent variable (predictor) X_k of the type

$$X_k \leq \theta_j \quad \text{or} \quad X_k > \theta_j \qquad (1)$$

where θ_j is the threshold value. If a given observation from the current node meets this condition, it is transferred to a group in the left hereditary node, and if not - it goes to the right hereditary node. In this way, the separation by nodes is repeated multiple times until an end node is reached. The general criterion for the selection of a predictor variable at each node and its threshold value is the minimum deviation from all possible predictors and threshold values. Defining a given node as a terminal one depends on the minimum error achieved as per a preset criterion for the minimum number of observations or some other type of restriction [15], [28]. The observations which find their way to a given tree node are defined by a series of restrictions of the type (1), starting at the root of the tree.

Validation is usually applied when building regression trees, since they may be sensitive to random errors in the data. This helps correct - shrink - the initial tree, maintaining its regression characteristics and accuracy. In the case of fewer observations and variables, the use of the statistical method of cross-validation with V-fold is recommended. This validation technique in CART allows the construction of very reliable models superior to standard regression models. In this study,

we have used the standard 10% V-fold cross-validation. The data have been randomly divided into 10 equal non-intersecting subgroups, each containing 10% of the dataset. The tree has been built using 9/10 of the data (study sample) and the remaining 1/10 (test sample) have been used for prediction and to determine the level of the error. The tree construction process is repeated 10 times and the average error of the 10 series is taken as a general estimate. This procedure ensures accurate estimation of the dependent variable and allows for the tree to be used for the classification or regression of another dataset.

The estimate $\hat{y}_{[s]}$ for the value of the prediction in a node with the number s is the mean value of all measurements for the dependent variable y, which fall within this node:

$$\hat{y}_{[s]} = \overline{y_k}, \quad y_k \in s \qquad (2)$$

V. CART MODEL OF THE OUTPUT POWER *POUT*

A CART model has been built in order to determine the relationship between laser output power and the 10 basis input laser variables. The minimum number of observations has been set at 20 for parent nodes and 10 for end nodes. This guarantees that terminal nodes will not be too small.

One more concrete objective of our investigation is to build a tree which classifies well experiments with high values of output power. For this reason, further on we will especially observe the node which contains the highest values of output power *Pout*.

In order to specify the tree and its reverse prune so as to find a tree with an optimal relative error for the data, we apply the standard cross-validation procedure described in Section IV above.

The general topological structure of the resulting regression tree with 13 nodes is given in Fig. 2.
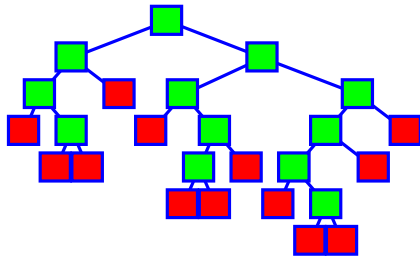
Fig. 2 Regression tree topology with 13 terminal nodes for a model with 10 predictors

Fig. 3 Diagram of the relative errors of CART models for a linear model with 10 predictors

The tree has been selected with a relative error of 3.8%. The obtained model accounts for R^2=96.8% of the sample following cross-validation with 10%V-fold. The general curve of the relative error of the models is shown in Fig. 3. It shows that the largest tree with 17 nodes has a minimal relative error of 3.7% which is insignificantly better than the chosen solution with 13 terminal nodes, providing 3.8% relative error.

Another criterion for the selection of a tree is the specific information in the nodes. Some of this information is shown in Fig. 4. Terminal node s=11 contains the highest power values with a standard deviation 7.842.
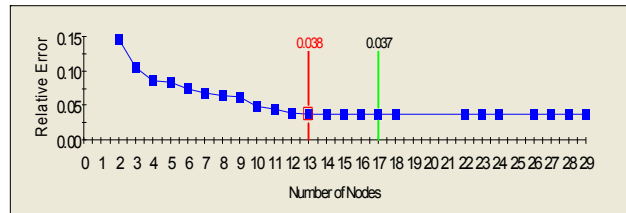
The values of the output laser power *Pout* predicted by the regression using formula (2) are the mean values of all cases, classified in the corresponding terminal nodes. In the case of s=11 it is (see also Fig. 4, terminal node 11)

$$\hat{Pout}_{[11]} = 109.286\,W \qquad (3)$$

This approximation is within a relative error of 3.8%, which is completely satisfactory, since it is comparable to the unavoidable experiment error, considered to be within 5%.

Fig. 5 shows all splitters used to build the tree (compared to a part of Fig. 4). For terminal node 11, which is of interest, through the cross-section of local rules, we find the region:

$$Node11: PIN > 4.25\,kW, \quad C \leq 1.75\,nF,$$
$$15.5\,kHz < PRF \leq 19.25\,kHz \qquad (4)$$

The overall quality of approximation by the regression tree is shown in Fig. 6, showing the experiment values of output power *Pout* against those predicted by the model.
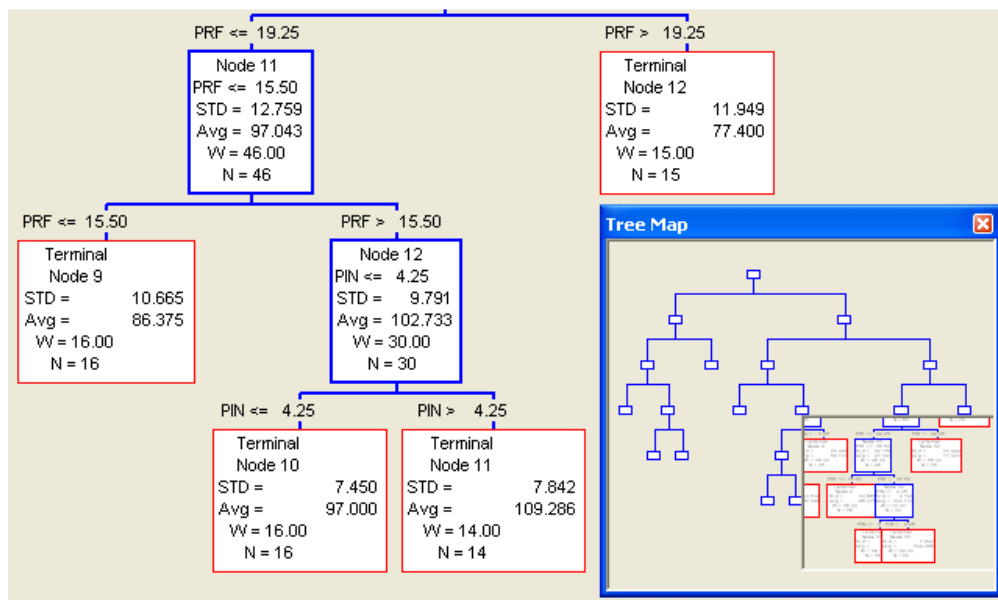

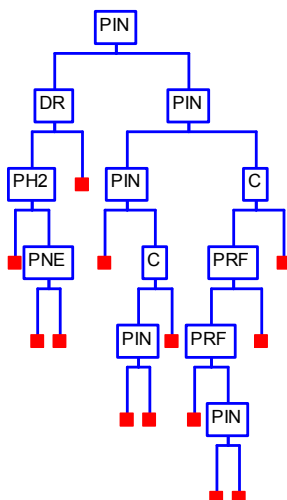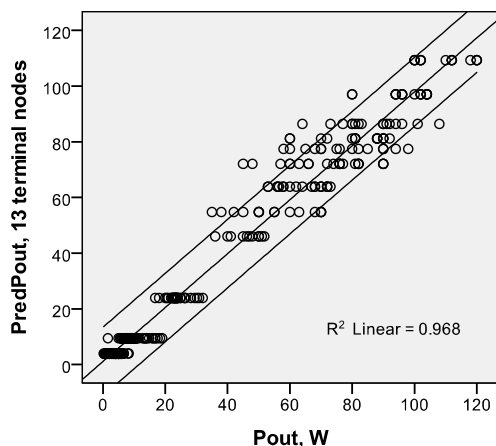Fig. 4 Specific characteristics of the nodes with maximum values of output power *Pout* in a linear model

Fig. 5 Distribution of splitters for each node of the tree

Fig. 6 Experiment values of laser output power *Pout* against predicted *PredPout* by using the obtained model with a 5% confidence interval

VI. DISCUSSION

In the obtained CART model of the 10 independent physical parameters, only 6 participate in the constructed classification tree. These defining parameters are:

$$PIN, DR, C, PH2, PRF, PNE \tag{5}$$

As it observed in Fig. 5, when the experiments (cases) are classified, three main third-level branches form, corresponding to a large degree to the three types of physical classification of copper lasers - small, medium and large bore lasers [1] (considered from left to right of the tree in Fig. 3). Of the parameters (5), *PIN* is the most important quantity. It is the root of the tree and subsequently participates in 4 more nodes related to the classification of medium and high laser power values *Pout*. For lower power values (along the left end branch in Fig. 5), the defining parameters are *PIN, DR, PH2* and

PNE. For medium output power values are - *PIN* and *C*. For high output power, these are *PIN, C* and *PRF*, respectively.

These results also correspond to the physical processes determining the laser generation (output power). For *PIN* it can be explained by the fact, that when the supplied electric power *PIN* is increased, the energy of the electrons rises. This leads to a higher probability of the upper laser level being populated and laser generation *Pout* increases. Following the obtained classification one can conclude that the supplied electric power has decisive role in the overall CuBr laser performance.

The results indicate that the obtained linear CART model on the base of the investigated data sample describes quite well the various groups of classified cases, predicting values for the nodes, and in particular predicting the measured maximum output power within a relative error of 5%.

An important comparison can be made with the models obtained using another powerful non-parametric technique - MARS. For the same data, second degree MARS models concur with 98-99% of the data, but best the predicted values obtained in this paper [12]. However, this corresponds to the well known common conclusions when compare basic and advanced data mining algorithms [15].

On the other hand, the advantage of the presented CART model is that it provides more accurate criteria for the classification of individual experiment groups which are of practical use and cannot be obtained by other type of statistical techniques.

VII. CONCLUSION

A CART model which classifies groups of similar experiments with respect to the values of the average output laser power has been built for a copper bromide vapor laser. The variables which play the main role in increasing laser output power have been identified, as well as the intervals these should be within when conducting future experiment and developing laser sources of the same type.

REFERENCES

[1] N. V. Sabotinov, "Metal vapor lasers," in: *Gas Lasers*, M. Endo R.and F. Walter, Eds., Boca Raton: CRC Press, 2006, pp. 449-494.

[2] P. G. Foster, *Industrial applications of copper bromide laser technology*, Ph.D. Dissertation, University of Adelaide, School of Chemistry and Physics, Dept. of Physics and Mathematical Physics, Australia, 2005.

[3] M. J. Kushner and B. E. Warner, "Large bore copper vapor lasers: Kinetics and scaling issues," *Journal of Applied Physics*, vol. 54, pp. 2970-2982, 1983.

[4] *Encyclopedia of Low-temperature Plasma*, Series B, vol. 7: Numerical modeling of low-temperature plasmas, M. Ianus, Ed., Moscow, 2004 (in Russian).

[5] A. M. Boichenko, G. S. Evtushenko, and S. N. Torgaev, "Simulation of a CuBr laser," *Laser Physics*, Springer, vol. 18, pp. 1522-1525, 2008.

[6] S. G. Gocheva-Ilieva and I. P. Iliev, "Statistical models of characteristics of metal vapor lasers," New York: Nova Science Publishers, Inc., 2011.

[7] I. P. Iliev, S. G. Gocheva-Ilieva, D. N. Astadjov, N. P. Denev, and N. V. Sabotinov, "Statistical analysis of the CuBr laser efficiency improvement," *Optics and Laser Technology*, Elsevier, vol. 40, no. 4, pp. 641-646, 2008.

[8] I. P., Iliev S. G. Gocheva-Ilieva, D. N. Astadjov, N. P. Denev, and N. V. Sabotinov, "Statistical approach in planning experiments with a copper bromide vapor laser," *Quantum Electronics*, vol. 38, no. 5, pp. 436-440, 2008.

[9]   I. P. Iliev, S.G. Gocheva-Ilieva, and N.V. Sabotinov, "Classification analysis of CuBr laser parameters," *Quantum Electron*, vol. 39, pp. 143-146, 2009.

[10] S. G. Gocheva-Ilieva and I. P. Iliev, "Parametric and nonparametric empirical regression models: case study of copper bromide laser generation," Mathematical Problems in Engineering, Hindawi Publishing Corporation, vol. 2010, Article ID 697687, 15 pages, 2010.

[11] S. G. Gocheva-Ilieva and I. P. Iliev, "Nonlinear regression model of copper bromide laser generation," in *Proc. COMPSTAT'2010, Y. Lechevallier, G. Saporta, Eds., 19th Int. Conf. Comput. Statistics, Paris - France, August 22-27*, Physica-Verlag, Springer_ebook, pp. 1063-1070, 2010.

[12] I. P. Iliev, D. S. Voynikova, and S. G. Gocheva-Ilieva, "Simulation of the output power of copper bromide lasers by the MARS method," *Quantum Electronics*, vol. 42, No 4, pp. 298-303, 2012.

[13] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Belmont: Wadsworth International, 1984.

[14] D. Steinberg, Dan and Phillip Colla, *CART: Tree-Structured Non-Parametric Data Analysis*, San Diego: Salford Systems, 1995.

[15] R. Nisbet, J. Elder, and G. Miner, *Handbook of statistical analysis and data mining applications*, Elsevier Academic Press, Burlington, 2009. Ch. 11.

[16] *CART® Classification and Regression Trees*, 2012. http://www.salford-systems.com/en/products/cart, Accessed 10 Jan 2013.

[17] N. V. Sabotinov, P. K. Telbizov, and S. D. Kalchev, *Bulgarian patent* N 28674, 1975.

[18] N. V. Sabotinov, N. K. Vuchkov, and D. N. Astadjov, "Gas laser discharge tube with copper halide vapors," *United States Patent* 4635271, 1987.

[19] D. N. Astadjov, N. V. Sabotinov, and N. K. Vuchkov, "Effect of hydrogen on CuBr laser power and efficiency," *Opt. Commun*. vol. 56 pp. 279-282, 1985.

[20] D. N. Astadjov, K. D. Dimitrov, C. E. Little, and N. V. Sabotinov, "A CuBr laser with 1.4 W/cm^3 average output power," *IEEE J. Quant. Electronics*, vol. 30, pp.1358-1360, 1994.

[21] V. M. Stoilov, D. N. Astadjov, N. K. Vuchkov, and N. V. Sabotinov, "High spatial intensity 10 W- CuBr laser with hydrogen additives," *Opt. and Quant. Electron*. vol. 32, pp. 1209-1217, 2000.

[22] NATO contract SfP, 97 2685, 50W Copper Bromide laser, 2000.

[23] D. N. Astadjov, K. D. Dimitrov, D. R. Jones, V. L. Kirkov, C. E. Little, N. Little, et al., "Influence on operating characteristics of scaling sealed-off CuBr lasers in active length," *Opt. Commun*. vol. 135, pp. 289-294, 1997.

[24] K. D. Dimitrov, N. V. Sabotinov, "High-power and high-efficiency copper bromide vapor laser," *SPIE*, vol. 3052, pp. 126-130, 1996.

[25] D. N. Astadjov, K. D. Dimitrov, D. R. Jones, V. K. Kirkov, C. E. Little, N. V. Sabotinov et al., "Copper bromide laser of 120-W average output power," *IEEE J. Quant. Electron*. vol. 33, pp. 705-709, 1997.

[26] N. P. Denev, D. N. Astadjov, and N. V. Sabotinov, "Analysis of the copper bromide laser efficiency," in *Proc. of Fourth Intern. Symp. on Laser Technologies and Lasers'2005, Plovdiv, Bulgaria*, pp. 153-156, 2006.

[27] L. Leech, K. C. Barrett, G. A. Morgan, SPSS for Intermediate Statistics: Use and Interpretation, 2nd ed., Lawrence Erlbaum Associates Publishers, New Jersey, 2005, ch. 2.

[28] A. J. Izenman, *Modern Multivariate Statistical Techniques Regression, Classification, and Manifold Learning*, New York: Springer, 2008, Ch. 9.

[29]