

Классификационный анализ параметров CuBr-лазера

И.П.Илиев, С.Г.Гочева-Илиева, Н.В.Саботинов

Впервые на основе большого количества экспериментальных данных проведен статистический классификационный анализ параметров CuBr-лазера с длинами волн излучения 510.6 и 578.2 нм. С помощью многомерного кластерного анализа изучены десять основных величин, влияющих на лазерную мощность. Приведены таблицы и дендрограммы классификации этих величин. Полученные результаты хорошо согласуются с результатами проведенного нами ранее исследования лазерных параметров с помощью многомерного факторного и регрессионного анализа и дополняют их. Решены некоторые задачи планирования эксперимента.

Ключевые слова: лазер на бромиде меди, кластерный анализ, лазерная мощность.

1. Введение

Наличие большого количества экспериментальных данных в принципе дает возможность проведения независимого феноменологического исследования основных зависимостей и соотношений для описания сложных процессов и систем. Фундаментальное значение здесь имеет классификационный анализ, поскольку, как известно, он может служить как для разработки научных теорий, так и для выявления специфических свойств отдельных компонентов систем, важных для планирования эксперимента и управления им [1]. Нужно отметить, что многомерные статистические методы классификации и кластеризации данных применяются преимущественно в экономике и социальной сфере, однако и в инженерных науках, и в физике их использование также весьма перспективно [2].

В частности, применение классификационного анализа для исследования источников лазерного излучения, в том числе лазеров на бромиде меди (CuBr), позволяет решать следующие задачи: 1) классифицировать независимые величины (переменные) лазерной системы по макрокатегориям и уменьшить их число до нескольких значимых групп; 2) определить место каждой из независимых величин в общей иерархии; 3) выявить взаимосвязи рассматриваемых величин (близкие и далекие), их взаимное влияние, а также влияние на остальные величины или на целые группы; 4) определить степень влияния (или расстояния) группы переменных или отдельной независимой величины на зависимые лазерные характеристики, такие как энергетические характеристики, лазерная эффектив-

ность и др.; 5) использовать полученные классификационные и группирующие переменные для построения различных типов моделей для прогнозирования поведения исследуемой системы, в том числе для планирования отсеивающего и экстремального экспериментов, последовательности проведения измерений, распределения ресурсов. Классификационный анализ позволяет выявить также новые свойства лазерной системы и характерные зависимости, которые нельзя получить путем других известных теоретических и экспериментальных методов.

В настоящей работе рассматривается CuBr-лазер с длинами волн излучения $\lambda_1 = 510.6$ нм и $\lambda_2 = 578.2$ нм. Исследование проведено на основе большого объема экспериментальных данных, полученных за последние десятилетия в Лаборатории лазеров на парах металлов Института физики твердого тела им. Георгия Наджакова Болгарской академии наук [3–11]. Вкратце описаны результаты работ [12–14], полученные с помощью многомерного факторного анализа. Проведен многомерный классификационный статистический анализ с применением иерархического кластерного анализа физических переменных, а также анализ расчетов, и решены некоторые из сформулированных выше задач. Результаты сравниваются с данными классификации, установленной с помощью факторного анализа.

Статистические расчеты выполнены на базе 25 %-ной выборки из всех имеющихся данных для CuBr-лазера с помощью программного пакета SPSS [15].

2. Основные результаты многомерного факторного анализа, полученные ранее

Классификация в многомерном факторном анализе основана на корреляции переменных. Целью является определение заранее неизвестного числа макропеременных (факторов), группирующих независимые входные переменные по степеням их взаимной корреляции. Полученные факторы обычно не коррелируют между собой, что позволяет использовать их для дальнейшего анализа, например для построения регрессионных моделей и предсказания поведения системы. При наличии разного

I.P.Iliev. Department of Physics, Technical University, Plovdiv, 25 Tzanko Djuvshtabanov St., 4000 Plovdiv, Bulgaria; e-mail: iliev55@abv.bg
S.G.Gocheva-Ilieva. Department of Applied Mathematics and Modelling, Faculty of Mathematics and Informatics, Paisii Hilendarski University of Plovdiv, 24 Tzar Assen St., 4000 Plovdiv, Bulgaria; e-mail: snow@uni-plovdiv.bg, snegocheva@yahoo.com
N.V.Sabotinov. Metal Vapour Lasers Department, Georgi Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences, 72 Tzarigradsko Chaussee, 1784 Sofia, Bulgaria; e-mail: nsabotinov@issp.bas.bg

Табл.1. Ротированная матрица факторных коэффициентов, полученных методом варимакс с кайзер-нормализацией [12–14].

Переменная	Компонент (фактор)		
	1	2	3
P_{in}	0.942		
d_r	0.905		
L	0.789		
D	0.744		
P_L		-0.913	
p_{H_2}			0.943

Примечание. Факторные коэффициенты, меньшие 0.5, не приведены.

типа переменных, участвующих в физических процессах, происходящих в лазерной системе, они сначала приводятся к стандартизованному (безразмерному) виду. Данные должны также удовлетворять определенным статистическим критериям. Их нужно рандомизировать, исследовать типы их распределений, условия адекватности модели и т. д. [2, 12–16].

Для проведения классификации в настоящей работе рассматриваются следующие десять основных переменных (параметров): внутренний диаметр лазерной трубки D , внутренний диаметр диафрагмы d_r , длина активной зоны (расстояние между электродами) L , подводимая электрическая мощность P_{in} , электрическая мощность на единицу длины P_L , частота повторения электрических импульсов f , давление буферного газа неона p_{Ne} , давление добавочного газа водорода p_{H_2} , эквивалентная емкость конденсаторной батареи C , температура резервуара с $CuBr$ T_r . В качестве основной зависимой переменной будет рассматриваться выходная лазерная мощность P_{out} .

В [12, 13] было установлено, что только переменные D , d_r , L , P_{in} , P_L и p_{H_2} коррелируют с выходной мощностью P_{out} , а также между собой. Эта корреляция слабая, а для переменных f , p_{Ne} , C и T_r она отсутствует вообще. Процедура факторного анализа в данном случае требует исключения этих переменных из дальнейших расчетов. Далее рассматривались только оставшиеся шесть (1–6) переменных: D , d_r , L , P_{in} , P_L и p_{H_2} . Они были сгруппированы в три взаимно ортогональных фактора: в первый фактор входили переменные P_{in} , d_r , L и D , во второй – P_L , а в третий – p_{H_2} . Их факторные коэффициенты приведены в табл.1.

Нужно отметить, что полученная с помощью факторного анализа классификация независимых переменных является частичной, т. к. при этом не учитываются оставшиеся четыре (7–10) величины – f , p_{Ne} , C и T_r – из-за их слабых корреляций.

3. Классификация с помощью кластерного анализа

В отличие от факторного анализа, процедуры кластерного анализа основаны на классификации объектов по степеням их гомогенности и близости [1]. Формирование групп (кластеров) по заданным критериям проводится путем объединения однородных объектов, а сами кластеры должны оставаться неоднородными. В качестве количественной оценки близости объектов рассматривается некоторая выбранная метрика, чаще всего обычное евклидово расстояние $d_{ij} = [\sum_{k=1}^p (x_{ik} - x_{jk})^2]^{1/2}$ между точками X_i и X_j (объектами) p -мерного векторного пространства [1].

Существует большое число методов кластеризации. При малом числе исследуемых объектов, как и в нашем случае, самыми подходящими являются иерархические агломеративные методы. Результаты приводятся в виде таблиц и дендрограмм (древовидных диаграмм), которые выражают иерархическую структуру матрицы сходства и правила получения кластеров. Имеется множество способов объединения объектов в кластеры, а затем и самих кластеров. В случае кластеризации в виде «цепи» используются методы межгрупповой связи (between groups linkage) и ближайшего соседа (nearest neighbor). Однако конкретный выбор метрики, подходящего числа кластеров и метода их формирования является важным, а иногда и трудным этапом кластерного анализа.

3.1. Результаты кластерного анализа

Сначала проведем частичный кластерный анализ только для первых шести независимых переменных (D , d_r , L , P_{in} , P_L и p_{H_2}), используемых в предыдущей классификации. Наша цель – сравнить полученные результаты с результатами факторного анализа.

Первый этап – получение матрицы, содержащей результаты сравнения объектов (табл.2). В нашем случае в качестве показателя, характеризующего степень сходства (различия) объектов выбран квадрат евклидова расстояния. Нужно отметить, что в табл.2 представлены только результаты сравнения на первом шаге, когда каждый объект рассматривается как кластер. Методом средних межгрупповых связей независимые переменные группируются в три кластера, как это видно из табл.3. В первый кластер входят переменные D , d_r , L и P_{in} , во второй – P_L , а в третий – p_{H_2} . Таким образом получаем полное соответствие с результатами факторного анализа (см. табл.1).

Из табл.2 видно, что минимальное значение коэффициента, характеризующего степень гомогенности формируемых кластеров (в нашем случае квадрат евклидова расстояния), равно 12.1 и связывает переменные d_r и L . Соответственно это первая связь, характеризующая степень сходства, наблюдаемому на рис.1. Следующий по величине коэффициент 20.1 связывает d_r и P_{in} , поэтому далее переменная P_{in} группируется с уже сформирован-

Табл.2. Матрица сходства (близости) шести переменных.

Переменная	D	d_r	L	P_{in}	P_L	p_{H_2}
D	0	23.9	36.7	44.7	227.5	103.0
d_r	23.9	0	12.1	20.1	223.8	86.0
L	36.7	12.1	0	21.2	247.8	67.3
P_{in}	44.7	20.1	21.2	0	189.3	91.7
P_L	227.5	223.8	247.8	189.3	0	217.9
p_{H_2}	103.0	86.0	67.3	91.7	217.9	0

Табл.3. Принадлежность шести переменных к трем кластерам в группе из трех кластеров.

Переменная	Номер кластера
D	1
d_r	1
L	1
P_{in}	1
P_L	2
p_{H_2}	3

Табл.4. Матрица сходства всех десяти переменных.

Переменная	D	d_r	L	P_{in}	P_L	p_{H_2}	f	p_{Ne}	C	T_r
D	0	19.5	33.0	39.4	202.4	95.9	145.0	158.4	101.5	115.9
d_r	19.5	0	9.5	16.9	201.1	79.7	155.3	161.0	105.8	101.6
L	33.0	9.5	0	19.4	223.3	58.9	159.0	149.5	120.2	115.0
P_{in}	39.4	16.9	19.4	0	169.3	83.1	155.1	145.0	107.6	116.3
P_L	202.4	201.1	223.3	169.3	0	196.0	113.4	88.3	141.2	132.6
p_{H_2}	95.9	79.7	58.8	83.1	196.0	0	164.8	140.7	158.2	174.3
f	144.7	155.3	159.0	155.1	113.4	164.9	0	97.4	139.8	120.5
p_{Ne}	158.4	161.0	149.5	145.0	88.3	140.7	97.4	0	172.6	126.5
C	101.5	105.8	120.2	107.6	141.2	158.2	139.8	172.6	0	93.4
T_r	115.9	101.6	115.0	116.3	132.6	174.3	120.5	126.5	93.4	0

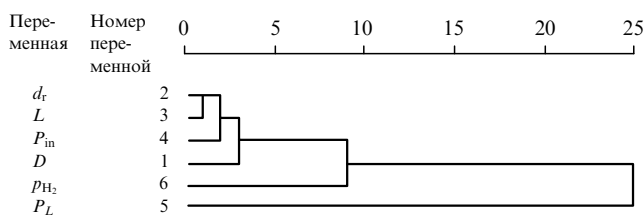


Рис.1. Дендрограмма шести независимых переменных, полученная методом межгрупповых связей. Здесь и на рис.2 и 3 на горизонтальной оси указаны нормированные квадраты евклидовых расстояний; значению 25 соответствует максимальная метрика 247.8 из табл.2.

Табл.5. Принадлежность всех десяти переменных к пяти кластерам в группах из N кластеров.

Переменная	Номер кластера			
	$N = 5$	$N = 4$	$N = 3$	$N = 2$
D	1	1	1	1
d_r	1	1	1	1
L	1	1	1	1
P_{in}	1	1	1	1
P_L	2	2	2	2
p_{H_2}	1	1	1	1
f	3	3	2	2
p_{Ne}	2	2	2	2
C	4	4	3	1
T_r	5	4	3	1

ным первым кластером и т. д. С помощью этой процедуры получаем полную структуру, приведенную на рис.1.

На втором этапе будем классифицировать все десять исходных переменных. В табл.4 представлена их матрица сходства. В основной табл.5 приведена классификация переменных по группам из двух, трех, четырех и пяти кластеров. Необходимо определить оптимальное число кластеров. Эту задачу можно решить с помощью дендрограммы, приведенной на рис.2. Она получена тем же методом, что и дендрограмма на рис.1.

Внимательный анализ последовательности процедуры кластеризации показывает, что все десять независимых переменных образуют три кластера. Первый кластер включает в себя переменные D , d_r , L , P_{in} и p_{H_2} , второй – P_L , p_{Ne} и f , а третий – C и T_r . Эта группировка соответствует колонке для трех кластеров в табл.5. Окончательно получаем три кластера для классификации всех десяти исходных независимых переменных.

Третий этап – определение места зависимой переменной P_{out} (номер 11 на рис.3) среди независимых величин. Близость между ними видна из рис.3. Как и следовало

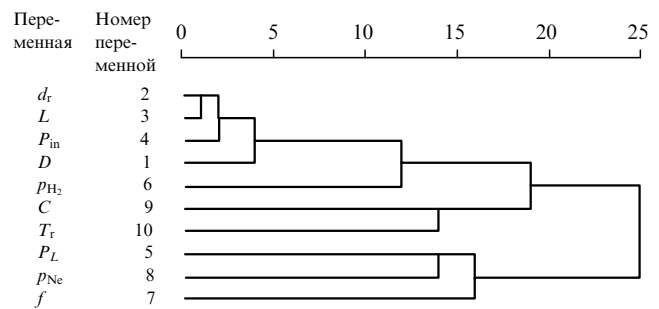


Рис.2. Дендрограмма десяти независимых переменных, полученная методом межгрупповых связей.

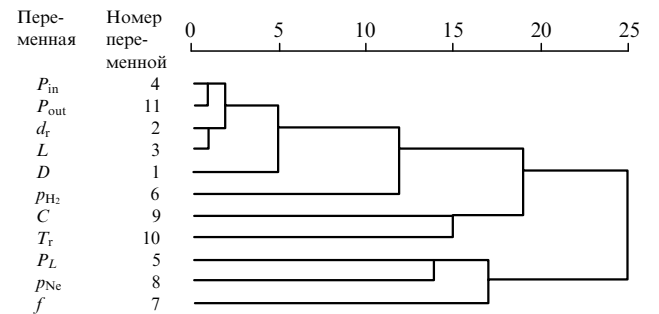


Рис.3. Дендрограмма десяти независимых переменных и P_{out} , полученная методом межгрупповых связей.

ожидать, P_{out} ближе к переменным D , d_r , L , P_{in} и p_{H_2} и образует с ними первый кластер. Последнее подтверждает существенное влияние этих пяти переменных на P_{out} .

3.2. Анализ результатов кластерного анализа

Из результатов анализа (табл.5 и рис.2) следует, что пять из исследуемых физических величин (D , d_r , L , P_{in} и p_{H_2}) имеют сильно выраженную взаимную гомогенность. При этом из табл.5 видно, что они сгруппированы так, что их нельзя разделить при увеличении числа кластеров. Тот же вывод следует из рис.2. Остается постоянным также взаимное расположение величин. Таким образом, можно заключить, что эти величины являются наиболее существенными при определении характеристик источника лазерного излучения.

Полученную классификацию можно использовать при планировании отсеивающего эксперимента, где из всей совокупности величин нужно выделить основную группу переменных, которые подвергаются дальнейшему детальному изучению [2, 17]. При проведении экстремального эксперимента с целью оптимизации объекта исследования приходится также начинать с варьирова-

ния важнейших переменных по порядку их гомогенности, т. е. в последовательности d_r , L , P_{in} , D и ρ_{H_2} .

4. Заключение

Таким образом, в данной работе впервые для лазеров на парах металлов применяется классификационный анализ переменных. Рассмотрены десять независимых переменных. На основе проведенного ранее факторного анализа с помощью представительной выборки из общей совокупности всех имеющихся результатов экспериментов и принципа корреляции условно выделены две группы переменных – существенные и несущественные. Существенные переменные классифицированы в три группы (фактора). Далее классификация переменных проведена с использованием статистической техники кластерного анализа и принципа гомогенности. Получена иерархическая зависимость и установлена взаимная связь переменных. Определены три кластера и порядок классификации.

Решены некоторые задачи, связанные с применением полученных результатов для планировании отсеивающего и экстремального экспериментов.

Настоящая работа проведена при поддержке ФНИ Болгарского Министерства образования и науки (проект ВУ-МИ-205/2006) и НПД Пловдивского университета им. Паисия Хилендарского (проект 07M07).

1. Aldenderfer M.S., Blanshfield R.K. *Cluster Analysis* (Iowa: Sage Publ. Inc., 1984).
2. Джонсон Н., Лион Ф. *Статистика и планирование эксперимента*

в технике и науке: методы планирования эксперимента (М.: Мир, 1981).

3. Astadjov D.N., Vuchkov N.K., Sabotinov N.V. *IEEE J. Quantum Electron.*, **24**, 1927 (1988).
4. Vuchkov N.K., Astadjov D.N., Sabotinov N.V. *Opt. Quantum Electron.*, **23**, S549 (1991).
5. Astadjov D.N., Dimitrov K.D., Little C.E., Sabotinov N.V. *IEEE J. Quantum Electron.*, **30**, 1358 (1994).
6. Dimitrov K.D., Sabotinov N.V. *Proc. SPIE Int. Soc. Opt. Eng.*, **3052**, 126 (1996).
7. Astadjov D.N., Dimitrov K.D., Jones D.R., Kirkov V.K., Little L., Little C.E., Sabotinov N.V., Vuchkov N.K. *Opt. Commun.*, **135**, 289 (1997).
8. Astadjov D.N., Dimitrov K.D., Jones D.R., Kirkov V.K., Little C.E., Sabotinov N.V., Vuchkov N.K. *IEEE J. Quantum Electron.*, **33**, 705 (1997).
9. Endo M., Walter R.F. (Eds) *Gas Lasers* (Boca Raton: CRC, Taylor & Francis, 2007, p. 449).
10. Stoilov V.M., Astadjov D.N., Vuchkov N.K., Sabotinov N.V. *Opt. Quantum Electron.*, **32**, 1209 (2000).
11. Denev N.P., Astadjov D.N., Sabotinov N.V. *Proc. IV Int. Symp. on Laser Technologies and Lasers 2005* (Plovdiv, 2006, p. 153).
12. Илиев И.П., Гочева-Илиева С.Г. *Proc. Amer. Inst. Phys.*, **CP936**, 267 (2007).
13. Илиев И.П., Гочева-Илиева С.Г., Астаджов Д.Н., Денев Н.П., Саботинов Н.В. *Квантовая электроника*, **38** (5), 436 (2008).
14. Илиев И.П., Гочева-Илиева С.Г., Астаджов Д.Н., Денев Н.П., Саботинов Н.В. *Opt. Laser Technol.*, **40** (4), 641 (2008).
15. <http://www.realcoding.net/teach/SPSS>.
16. Kim J.-O., Mueller Ch.W. *Factor Analysis: Statistical Methods and Practical Issues* (Iowa: Sage Publ. Inc., 1986).
17. Montgomery D.C. *Design and Analysis of Experiments* (New York: Wiley, 2004).