
**МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ
В ПРИБОРОСТРОЕНИИ**

УДК 681.586.67

© Т. В. Осипова, А. М. Баранов, И. И. Иванов, 2022

**МЕТОД ГЛАВНЫХ КОМПОНЕНТ КАК АЛЬТЕРНАТИВНЫЙ
АЛГОРИТМ ОБРАБОТКИ ДАННЫХ
ТЕРМОКАТАЛИТИЧЕСКИХ СЕНСОРОВ**

В данной работе проведено исследование возможности и оценка использования метода главных компонент как альтернативного алгоритма обработки данных термокаталитических сенсоров. На основе полученных результатов было установлено, что влияние рабочей температуры сенсора на отклик минимально, что позволяет объединить их в единую систему. При этом однозначно определяется разная концентрация поданого газа. Определена методика исследования и представлены результаты, показавшие, что метод главных компонент позволяет различать отклики сенсоров при разной концентрации без использования дополнительных математических операций подготовки полученных данных.

Кл. сл.: термокаталитический сенсор, метан, обработка данных, метод главных компонент, распознавание газа

ВВЕДЕНИЕ

В настоящее время ввиду увеличения числа промышленных объектов, в том числе потенциально опасных, широкое распространение получили исследования в области охраны окружающей среды, которые сориентированы на снижение возможных последствий той или иной деятельности. Исследования направлены на разработку эффективных методов контроля норм допустимых воздействий на безопасность жизнедеятельности человека и природных экосистем. Среди таких исследований особое место занимают исследования по созданию и применению систем мониторинга среды, окружающей человека. Одним из направлений данного вида мониторинга является разработка датчиков для анализа газового состава воздуха, которые не только определяют концентрацию того или иного целевого газа, но способны распознать этот газ в составе смеси. Это делает совершенствование методов газового анализа и их применение в быту и промышленности актуальной задачей.

Традиционные методы анализа состава воздушной среды основываются на статистических методах обработки информации, в частности на таких фундаментальных, как количественный и качественный анализы. В данном случае качественный анализ направлен на идентификацию различных газов, которые присутствуют в воздухе, а количественный анализ — на определение концентрации анализируемого типа газа.

Для анализа состава воздуха используют газовые сенсоры различных типов [1–3]. Их всех объединяет тот факт, что все они преобразуют концентрацию анализируемого газа в электрический (реже в другой) сигнал. Такие измерения являются относительно простыми и фактически не содержат какой-либо математической обработки, за исключением простейшего пересчета концентрации в ток или напряжение. Для выполнения данного пересчета сенсоры калибруются на целевой газ.

Такой подход ограничивает возможности всех видов газовых сенсоров, в частности требует точного знания типа газа в воздухе. Это связано с тем, что, как правило, газовые сенсоры не обладают селективностью, т.е. способны "чувствовать" разные газы, в том числе и одновременно. Проблема низкой селективности присутствует в самом подходе при традиционных измерениях, когда объем данных недостаточен.

В частности, для проведения измерений термокаталитический сенсор калибруется в двух точках: при нулевой и известной концентрациях газа. Таким образом, фактически измеряется конечная температура сенсора при известной концентрации, т.е. результатом является только одно измеренное значение температуры. Это не позволяет использовать данный метод для газовых смесей неизвестного состава, поскольку сенсор будет реагировать на все горючие газы в воздухе.

Решить проблему селективности можно двумя способами: доработкой самих сенсоров (например, за счет применения фильтров, пропускающих только целевой газ, катализаторов и т.д.) [4]

или разработкой математических методов обработки получаемых данных.

В условиях быстрого развития технической и технологической базы, а также цифровой экономики возрастает важность применения математических методов анализа данных большого размера. Особое место в исследованиях занимают мульти-сенсорные системы [5], которые состоят из набора сенсоров и системы распознавания. Во многих случаях количество сенсоров превышает 3, что увеличивает количество исследуемых образцов газа. Поэтому при анализе данных, полученных при работе с большим количеством сенсоров, следует выбрать наиболее удобный метод обработки данных, чаще всего используются метод главных компонент (МГК) [6–8] и линейно дискриминантный анализ (ЛДА) [7].

В работе [7] представлены результаты анализа МГК и ЛДА для откликов массивов сенсоров на диоксид азота, аммиак и хлор. Области локализации откликов на газы различны, что позволяет распознать, какой газ воздействует на систему. Таким образом, МГК хорошо распознает отклики сенсора на разные газы.

Исследования в работе [8] показали, что полученные количественные характеристики различий образцов целесообразно использовать для автоматической классификации методами многомерной статистики. Данный результат показывает принципиальную возможность использования МГК для разделения типов исследуемых газов.

Задачей данного исследования является оценка возможности применения МГК в качестве альтернативной методики обработки данных термokatалитических сенсоров, в частности для измерения концентрации метана в воздухе для сенсоров с разной рабочей температурой. Важность данных исследований связана с оптимизацией параметров и характеристик термokatалитических сенсоров, в частности чувствительности и селективности, а также с необходимостью изменения методик измерения для получения большего числа данных для последующей математической обработки и повышения селективности сенсоров.

МЕТОДИКА ЭКСПЕРИМЕНТА

При исследовании сложных химических систем можно выделить два подхода [6]:

- построение содержательных моделей в виде систем дифференциальных или интегро-дифференциальных уравнений и применение к ним специальных методов решения;
- анализ данных, основанный на многофакторном формальном моделировании.

Формальное моделирование часто эффективно в случаях, когда непонятно, как строить содержательную модель, либо ее построение требует чрезмерных вычислений.

В основе многофакторного анализа данных лежат проекционные методы. Эти методы позволяют выделить в больших массивах данных скрытые переменные и анализировать связи, существующие в изучаемой системе.

Важное место в методах анализа многомерных данных занимает МГК. Центральная концепция метода — понятие главной компоненты. Так называют специальный тип скрытой переменной, которая описывает максимальное разнообразие данных. Она не может быть объявлена явно и непосредственно измерена. В математическом смысле главная компонента — это линейная комбинация всех объясняющих признаков. При этом к первой главной компоненте относят максимум общей дисперсии объясняющих признаков, ко второй — максимум оставшейся дисперсии без учета влияния первой и т.д.

После использования МГК новая модель имеет существенно меньшее количество переменных, в силу чего такой подход может интерпретироваться как проекционный.

В качестве сенсоров были взяты промышленные термokatалитические сенсоры, которые серийно выпускаются фирмой "НТЦ ИГД" [9]. Сенсор состоит из рабочего и сравнительного элементов, которые находятся в одном корпусе (рис. 1).

Обычно для проведения измерений сенсор нагревается до рабочей температуры, при которой имеет место горение целевого газа. Рабочая температура сенсора может варьироваться в достаточно широком диапазоне и зависит от протекающего через сенсор тока. Так как горение происходит с выделением тепла, то температура сенсора изменяется. В свою очередь это приводит к изменению сопротивления сенсора, которое регистрируется измерительной схемой. Это одно измеренное значение и является откликом сенсора.

В данном исследовании проводилось изучение температурной зависимости отклика сенсора. Была применена методика измерения откликов сенсоров, описанная в работе [10], которая позволила получить экспериментальные зависимости отклика сенсоров от приложенного к мостовой схеме напряжения питания для метана при концентрации 0.47 и 2.5% об. Были исследованы три группы сенсоров с рабочей температурой 409, 444 и 478 °С, которая определяется рабочим током 46, 50 и 54 мА соответственно. При постоянном рабочем напряжении (2.8 В в мостовой измерительной схеме [11]) разный рабочий ток достигается за счет разного начального сопротивления сенсора.

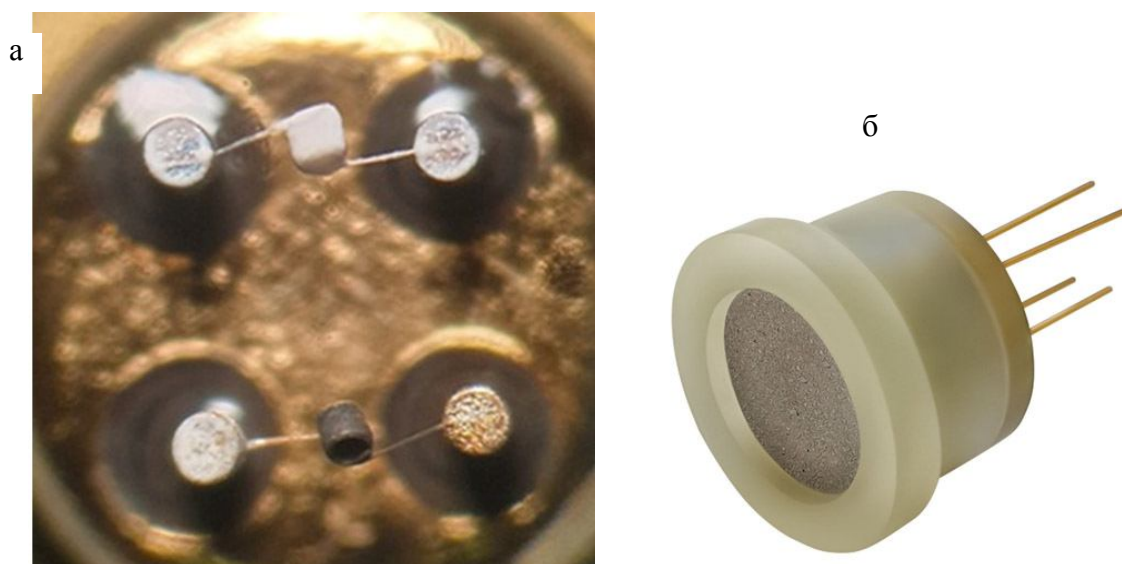


Рис. 1. Фотография термокаталитического сенсора.
а — внутренний вид сенсора (черный цилиндр — рабочий сенсор, белый цилиндр — элемент сравнения);
б — внешний вид корпуса сенсора

Табл. 1. Характеристика исследуемых групп сенсоров

Номер сенсора (образец)	Сопротивление при комнатной температуре (R_0), Ом	Рабочий ток, мА	Рабочая температура, °С
1	14.3	46	409
2	12.7	50	444
3	10.9	54	478

Далее при описании результатов будем оперировать номерами сенсора в соответствии с приведенной табл. 1.

Для исследования исходные данные, т.е. измеренные температурные зависимости отклика сенсора, были разделены на три набора значений:

- две концентрации применяемого газа при одном значении тока;
- одна концентрация применяемого газа при каждом значении тока;
- две концентрации применяемого газа при каждом значении тока.

Каждый набор обработан при помощи МГК, после чего наиболее оптимальный набор исследован в соответствии со следующей методикой.

Ввиду того, что метод чувствителен к значениям исходных данных с большим диапазоном [12], данные должны быть стандартизированы:

$$s_i = \frac{x_i - \bar{x}}{\sigma},$$

где s_i — стандартизированное значение;

x_i — исходное значение;

\bar{x} — среднее арифметическое исходного значения;

σ — стандартное отклонение исходного значения.

Стандартизированные данные обрабатываются при помощи МГК. Главные компоненты имеют три показателя информативности: стандартное отклонение главной компоненты, доля объясненной дисперсии, кумулятивная доля объясненной дисперсии [13].

По этим показателям производится поиск оптимального представления данных, т.е. определение количества главных компонент, достаточных для объяснения исходных данных. Существуют три правила поиска: правило Кайзера (значимы те

компоненты, собственное значение которых превосходит единицу), "тест осыпи" (оптимальное количество определяется как число, при достижении которого прирост объясненной дисперсии становится незначительным) и эмпирическое правило (количество главных компонент зависит от совокупного процента объясненной дисперсии, который должен быть не менее 80%).

После выбора количества главных компонент необходимо рассмотреть веса исходных данных, с какими они входят в полученные главные компоненты, и определить наиболее значительный. Это позволит узнать, какая переменная оказывает наибольшее влияние на главную компоненту.

Таким образом, исследование было проведено в следующей последовательности:

- стандартизация исходных данных;
- применение МГК к стандартизированным данным;
- определение трех показателей информативности главных компонент;

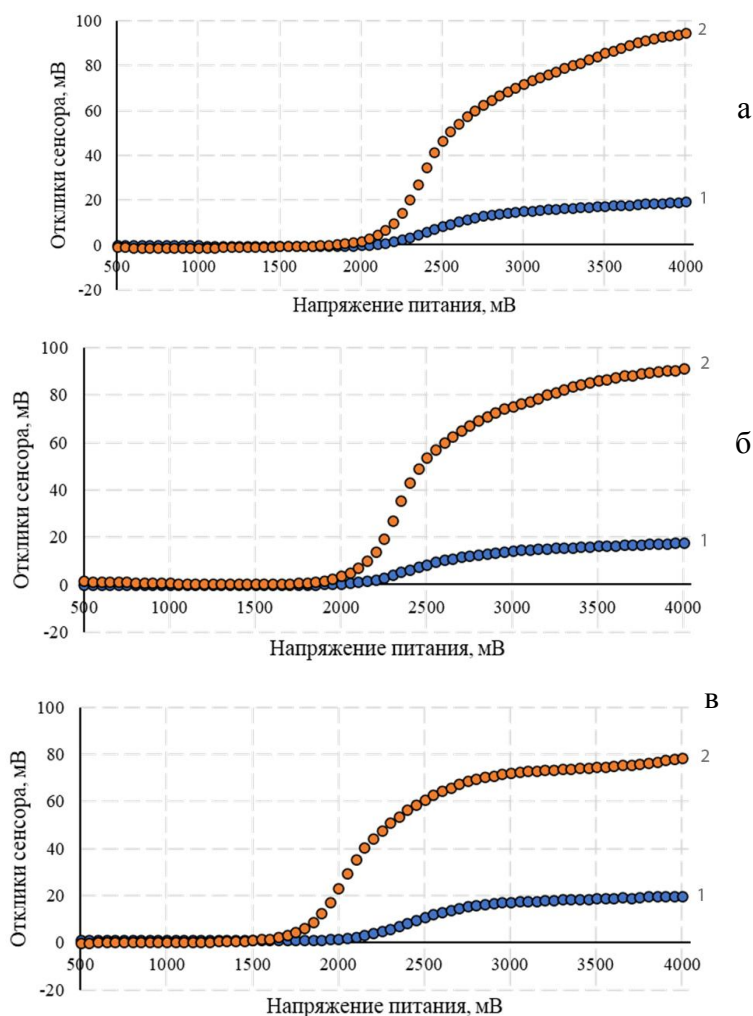
– поиск оптимального представления данных (количества главных компонент);

– определение переменной, оказывающей наибольшее влияние на выбранную главную компоненту.

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

На рис. 2 представлены S-образные зависимости отклика термокаталитического сенсора от подаваемого напряжения питания в установившейся атмосфере, содержащей метан.

Из полученных зависимостей видно, что для каждого образца концентрации 0.47% об. и 2.5% об. имеют практически одинаковый характер изменения (температура сенсора влияет на стадию начала каталитической реакции и стадию насыщения), что позволяет однозначно определить, какая концентрация метана в воздухе.



а

б

в

Рис. 2. Экспериментальная зависимость отклика сенсора от напряжения питания на концентрации метана 0.47% об. и 2.5% об. а — образец № 1, б — образец № 2, в — образец № 3

На рис. 2, в, отклик сенсора образца № 3 на концентрации 2.5% об. на 10 мВ меньше, чем у образцов № 1 и № 2.

На рис. 3 представлены экспериментальные зависимости отклика сенсора от напряжения питания при одной концентрации. Важно отметить, что при традиционном методе измерения при напряжении 2.8 В все сенсоры для каждой концентрации метана показывают близкое значение отклика.

Как видно из полученных зависимостей, отклик сенсора отсутствует до напряжения питания 2 В на образцах № 1 и № 2 и 1.6 В — на образце № 3. В то же время характер изменения отклика сенсора трех образцов схож между собой как при концентрации 0.47% об., так и при 2.5% об. Предполагается, что влияние рабочей температуры на отклик будет минимально, что позволит не учитывать в исследованиях рабочий ток сенсора для уверенного распознавания типа газа.

Исходя из приведенных экспериментальных зависимостей, при использовании одного сенсора графически распознать и проанализировать концентрацию газа возможно, но с увеличением количества сенсоров в системе возрастает объем получаемых данных, что усложняет процесс анализа. В частности, в данном исследовании рассматривается система, состоящая из 8 сенсоров. В таких системах важно организовать исходные данные и оценить, какие признаки имеют наибольшее влияние.

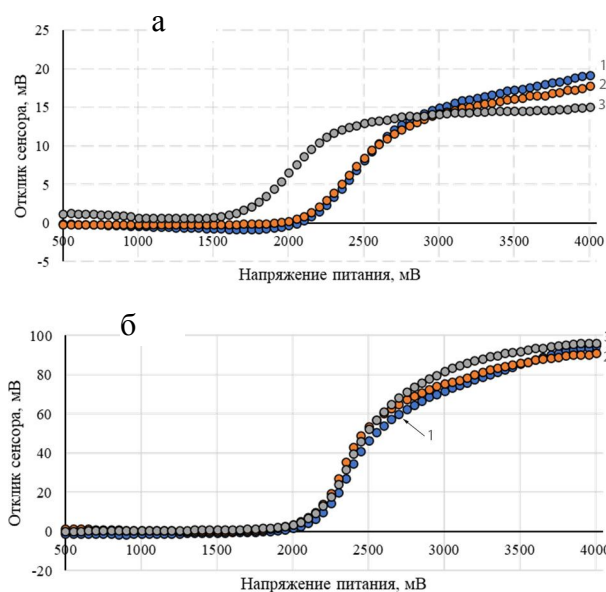


Рис. 3. Зависимость сенсорного отклика от напряжения питания для трех образцов (1 — образец № 1, 2 — № 2, 3 — № 3). а — на концентрации метана 0.47% об.; б — на концентрации метана 2.5% об.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Обработка исходных данных реализована средствами языка python 3.9.5 с использованием библиотек numpy v1.20.3, pandas 1.2.4, sklearn 0.0, а также в среде программирования Rstudio 1.4.

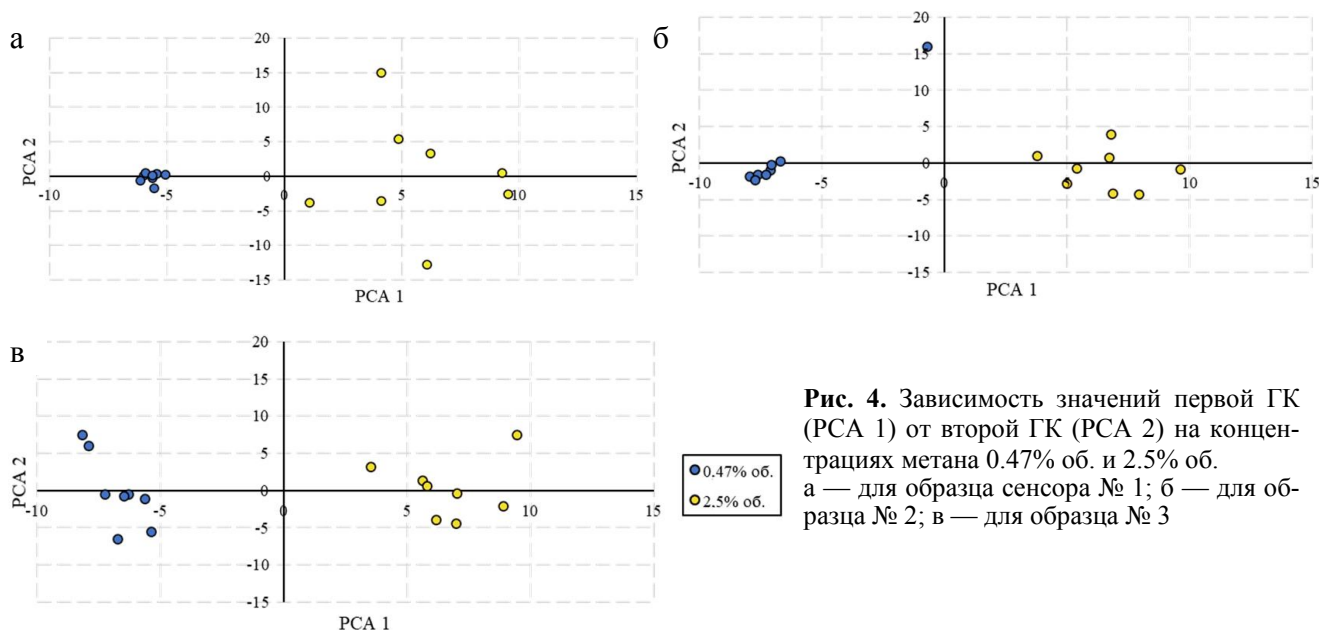


Рис. 4. Зависимость значений первой ГК (PCA 1) от второй ГК (PCA 2) на концентрациях метана 0.47% об. и 2.5% об. а — для образца сенсора № 1; б — для образца № 2; в — для образца № 3

Результаты обработки набора значений двух концентраций при одном токе при помощи МГК приведены на рис. 4.

Как видно из полученных графиков, области распознавания двух концентраций возможно однозначно выделить. Области локализации распределены слева направо, т.е. вдоль первой ГК (на графике PCA 1). Рассмотрев веса каждой ГК, полученные в ходе обработки, можно отметить, что

первая ГК связана с откликом сенсора при напряжении питания [2500, 2700] мВ, т.к. значения в данном диапазоне имеют наибольшее отклонение для двух концентраций, а вторая ГК — в диапазоне [1103, 1453] мВ.

Согласно рис. 4, разброс значений распределен вдоль второй ГК (на графике PCA 2). В табл. 2 представлены значения дисперсии исходных данных и второй ГК.

Табл. 2. Дисперсия исходных данных и ГК

Концентрация метана	Образец № 1		Образец № 2		Образец № 3	
	Исходные данные	Вторая ГК	Исходные данные	Вторая ГК	Исходные данные	Вторая ГК
0.47% об.	0.093	0.491	0.571	37.381	0.208	23.961
2.5% об.	11.722	65.886	0.111	8.021	0.132	15.212

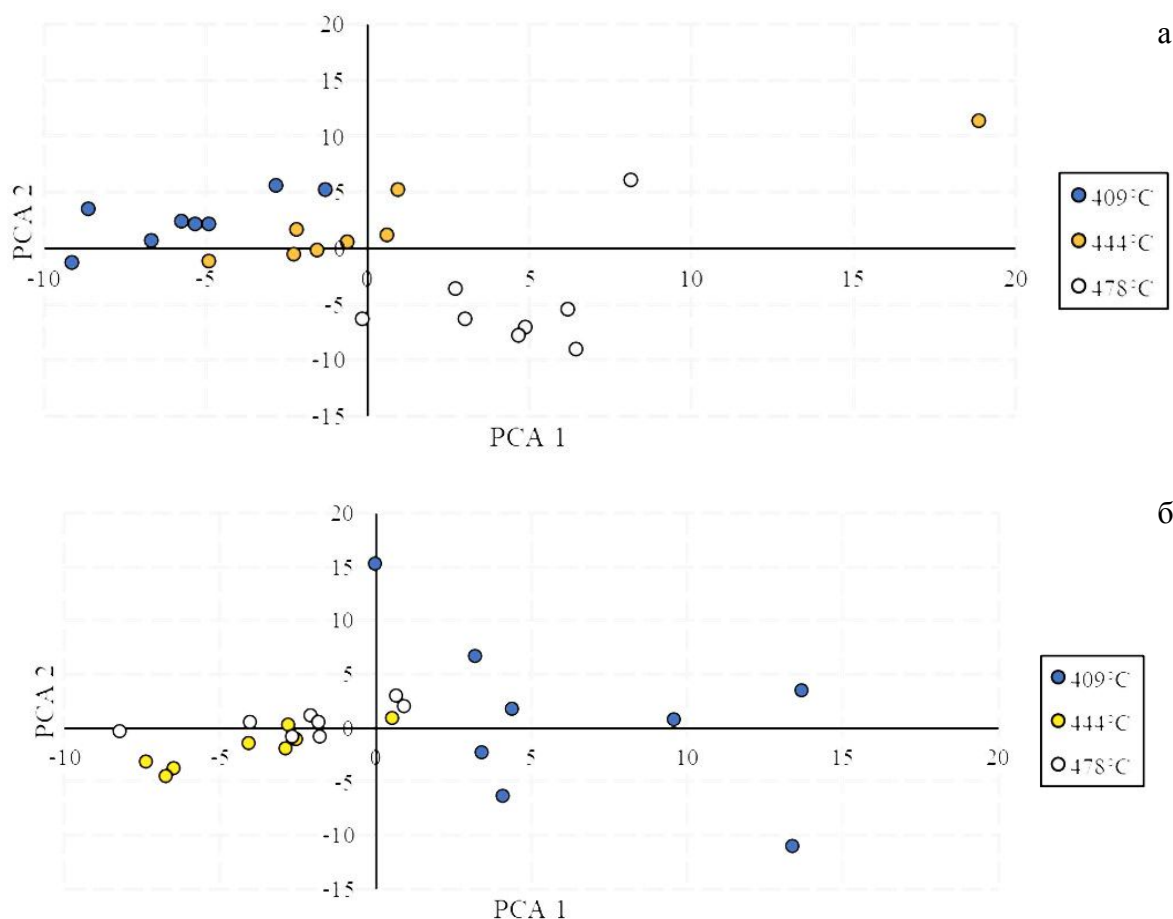


Рис. 5. Зависимость значений первой ГК (PCA 1) от второй ГК (PCA 2) для трех образцов сенсора. а — на концентрации метана 0.47% об.; б — на концентрации метана 2.5% об.

Полученный результат показывает, что разброс напрямую связан со значениями исходных данных. В образце № 2 на концентрации 0.47% об. присутствует единичный выброс, вследствие чего дисперсия сильно увеличена. Аналогично представлено в образце № 3.

В итоге, несмотря на разброс значений каждой из ГК, можно сделать вывод, что распознавание концентраций при одном токе возможно.

Результаты обработки набора значений одной концентрации при трех токах с помощью МГК приведены на рис. 5.

Из полученного графика видно, что области распознавания при одной концентрации, но разном значении тока выделить возможно, но они имеют пересечения. В данном случае возможен вариант использования таких данных для распознавания одной концентрации, но не для разделения на значения тока.

Результат обработки набора значений двух концентраций при трех токах при помощи МГК приведен на рис. 6.

Как видно из графика, рассчитанные данные при концентрации 0.47% об. образовали единую область распознавания. Аналогично можно выделить область для концентрации 2.5% об., хотя и имеется большой разброс между данными каждого сенсора при температуре 409 °С, отклонение в среднем составило ~5 мВ. Из этого можно сделать вывод, что при обработке исходных данных зависимость от тока потребления минимальна, следовательно, значения на одной концентрации при разных токах можно совмещать в один набор данных.

Рассмотрим информативность этого набора значений с применением трех правил поиска оптимального представления данных. Собственные значения главных компонент по правилу Кайзера (значимы те компоненты, собственное значение которых превосходит единицу) представлены в табл. 3.

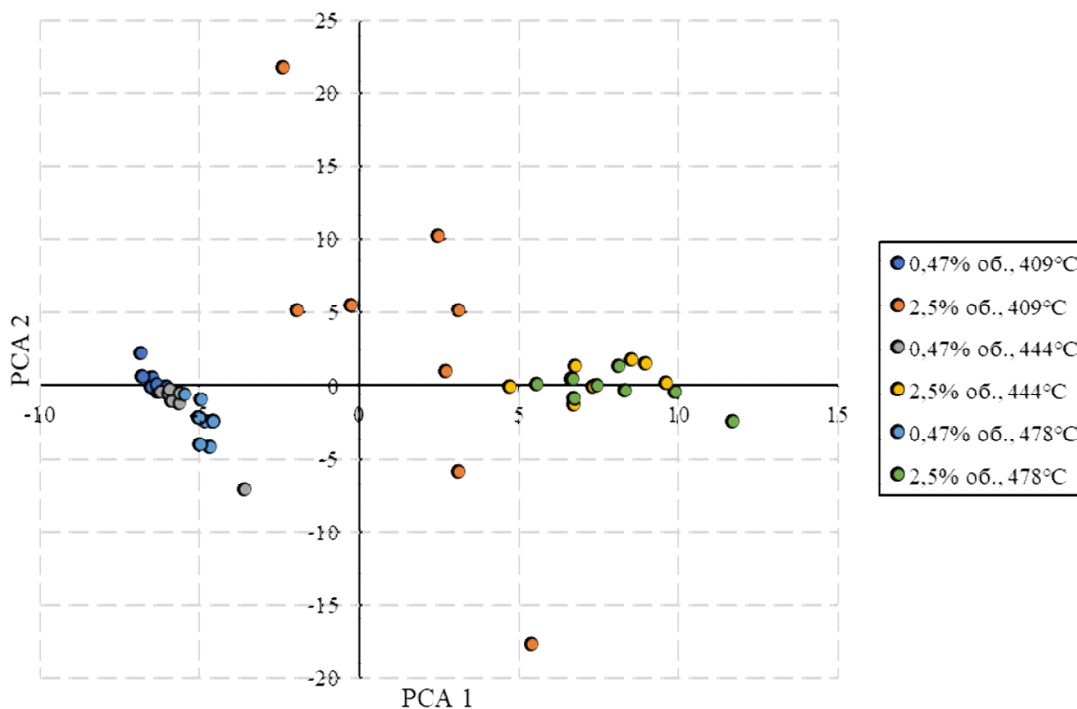


Рис. 6. Зависимость значений первой ГК (PCA 1) от второй ГК (PCA 2) на концентрациях метана 0.47% об. и 2.5% об.

Табл. 3. Собственные значения главных компонент

№ п/п	Номер главной компоненты	Собственное значение
1	ГК 1	38.55168
2	ГК 2	23.85346
3	ГК 3	7.008886
4	ГК 4	0.836804
5	ГК 5	0.566286

Табл. 4. Совокупный процент объясненной дисперсии

№ п/п	Номер главной компоненты	Совокупный процент объясненной дисперсии, %
1	ГК 1	54.3
2	ГК 2	87.9
3	ГК 3	97.776
4	ГК 4	98.954
5	ГК 5	99.752

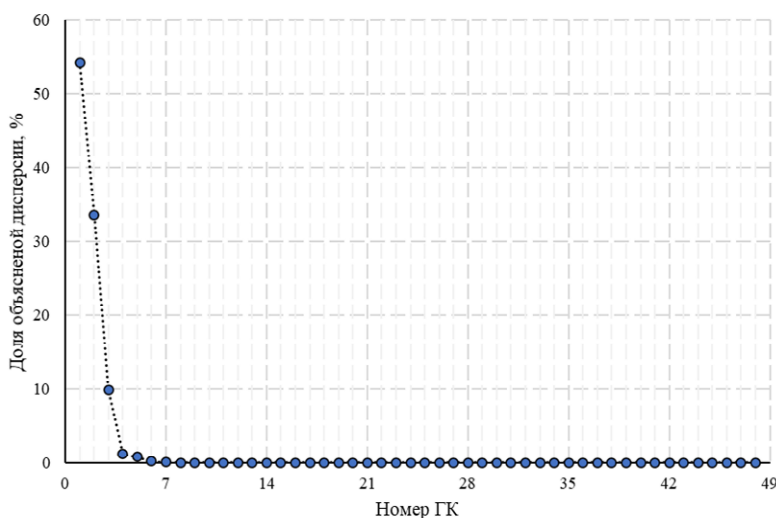
Как видно из полученных данных, по правилу Кайзера можно утверждать, что для представления исходных данных достаточно трех главных компонент.

Значения совокупного процента объясненной дисперсии представлены в табл. 4.

По эмпирическому правилу, совокупный процент объясненной дисперсии должен быть не менее 80%. Исходя из данных табл. 4, для представ-

ления исходных данных достаточно двух главных компонент для трех наборов данных.

По "тесту осыпи", оптимальное количество определяется как число, при достижении которого прирост Δ объясненной дисперсии становится незначительным. Зависимость доли Δ объясненной дисперсии от количества главных компонент представлена на рис. 7.

**Рис. 7.** Зависимость доли Δ объясненной дисперсии от количества главных компонент

Как видно из полученных зависимостей, прирост Δ объясненной дисперсии становится незначительным при достижении третьей главной компоненты, следовательно, можно предположить, что оптимальное количество главных компонент равно 3.

Полученные результаты по трем правилам показали, что для объяснения исходных данных в каждом из трех наборов оптимальное и необходимое количество главных компонент равняется 3.

ЗАКЛЮЧЕНИЕ

В работе было проведено исследование альтернативной методики обработки данных термokatалитических сенсоров на основе метода главных компонент для измерения концентрации метана в воздухе (0.47% об. и 2.5% об.) сенсорами с разной рабочей температурой (409, 444 и 478 °C). Представленная методика исследования позволяет найти оптимальный набор исходных данных, определить количество главных компонент и оценить влияние данных на главные компоненты.

Показано, что при обработке исходных данных при помощи МГК влияние рабочей температуры минимально, что позволяет совмещать значения на одной концентрации в один набор данных. Методика позволяет однозначно различать отклики образцов разной концентрации метана без использования дополнительных операций подготовки данных.

Полученные результаты имеют важное значение для разработки и оптимизации методов анализа состава воздуха, в частности позволяют перейти от пересчетов концентрации "внутри" сенсора к математической обработке получаемых с сенсора данных и на ее основе проводить распознавание поданного газа. Так как МГК относят к методам машинного обучения без учителя, то при объединении нескольких сенсоров в единый комплекс возможно создание системы распознавания типа газа с возможностью обучения на получаемых в процессе работы данных. Также остается актуальным вопрос о возможности применения МГК при анализе смеси, состоящей из двух и более типов газа.

СПИСОК ЛИТЕРАТУРЫ

1. Баранов А.М., Осипова Т.В. Современные тенденции в развитии сенсоров взрывоопасных концентраций горючих газов и паров горючих жидкостей (краткий обзор) // Научное приборостроение. 2021. Т. 31, № 4. С. 3–29. URL: <http://iairas.ru/mag/2021/abst4.php#abst1>
2. Williams D.E. Electrochemical sensors for environmental gas analysis // Current Opinion in Electrochemistry. 2020. Vol. 22. P. 145–153. DOI: 10.1016/j.coelec.2020.06.006
3. Singh D., Dahiya M., Kumar R., Nanda C. Sensors and systems for air quality assessment monitoring and management: A review // Journal of environmental management. 2021. Vol. 289. Art. 112510. DOI: 10.1016/j.jenvman.2021.112510
4. Ma Y., Kaczynski J., Ranacher C., Roshanghias A., Zauner M., Abasahl B. Nano-porous aluminum oxide membrane as filtration interface for optical gas sensor packaging // Microelectronic Engineering. 2018. Vol. 198. P. 29–34. DOI: 10.1016/j.mee.2018.06.013
5. Сысоев В.В., Зюрюкин Ю.А. Мультисенсорные системы распознавания газов типа "Электронный нос": краткий обзор литературы // Вестник СГТУ. 2007. Т. 2, № 1. С. 111–119. URL: <https://elibrary.ru/item.asp?id=12793666>
6. Сулейманов А.М. Анализ экспериментальных данных методом главных компонент // Известия КГАСУ. 2005. № 1 (3). URL: [https://izvestija.kgasu.ru/files/N1\(3\)2005/Sulejmanov_81-83.pdf](https://izvestija.kgasu.ru/files/N1(3)2005/Sulejmanov_81-83.pdf)
7. Кравченко Е.И., Петров В.В., Варезников А.С. Разработка методики распознавания образцов газовых смесей с помощью мультисенсорной системы мониторинга // Инженерный вестник Дона (сетевое издание). 2012. Т. 23, № 4, ч. 2. URL: <http://www.ivdon.ru/ru/magazine/archive/n4p2y2012/1346>
8. Белозерцев А.И., Черемисина О.В., Эль-Салим С.З., Манойлов В.В., Заруцкий И.В. Алгоритмы обработки данных в газоаналитических комплексах с полупроводниковыми сенсорами для обнаружения паров ядовитых веществ в окружающей среде // Научное приборостроение. 2018. Т. 28, № 1. С. 18–29. URL: <http://iairas.ru/mag/2018/abst1.php#abst3>
9. Karпов-sensors. Производство термokatалитических сенсоров горючих газов [Электронный ресурс]. URL: <http://karpov-sensor.com/> (дата обращения: 10.01.2022).
10. Иванов И.И., Баранов А.М., Талипов В.А., Миронов С.М., Колесник И.В., Напольский К.С. Разработка эффективных сенсоров обнаружения дозврывоопасных концентраций H₂ // Научное приборостроение. 2021. Т. 31, № 3. С. 25–36. URL: <http://iairas.ru/mag/2021/abst3.php#abst4>
11. Somov A., Baranov A., Suchkov A., Karelin A., Mironov S., Karpova E. Improving interoperability of catalytic sensors // Sensors and Actuators B. 2015. Vol. 221. P. 1156–1161. DOI: 10.1016/j.snb.2015.07.082
12. Лекция: Метод главных компонент [Электронный ресурс]. URL: <http://math-info.hse.ru/f/2015-16/ling-mag-quant/lecture-pca.html> (дата обращения: 04.12.2021).
13. Баймуратов И.П. Методы автоматизации машинного обучения. СПб.: Университет ИТМО, 2020. 40 с.

*Московский авиационный институт (национальный
исследовательский университет), Москва, Россия*

Контакты: *Осипова Татьяна Владиславовна,*
t.osipova.95@mail.ru

Материал поступил в редакцию 02.02.2022

PRINCIPAL COMPONENT ANALYSIS AS AN ALTERNATIVE ALGORITHM FOR PROCESSING DATA OF THERMOCATALYTIC SENSOR

T. V. Osipova, A. M. Baranov, I. I. Ivanov

Moscow Aviation Institute (National Research University), Moscow, Russia

In this article the study of the possibility and evaluation of the use of the principal component analysis as an alternative algorithm for processing data of thermocatalytic sensors is carried out. Based on the obtained results it was found that the effect of the sensor's operating temperature on the response is minimal, which makes it possible to combine them into a single system. At the same time the different concentration of the supplied gas is uniquely determined. The research methodology is defined and the results are presented, showing that the principal component analysis makes it possible to distinguish sensor responses at different concentrations without using additional mathematical operations to prepare the obtained data.

Keywords: thermocatalytic sensor, methane, data processing, principal component analysis, gas detection

INTRODUCTION

At present, due to the increase in the number of industrial facilities, including potentially hazardous ones, research in the field of environmental protection, focused on reducing the possible consequences, has become widespread. The research is aimed at developing effective methods for monitoring the standards of permissible impacts on the safety of human life and natural ecosystems. Among such studies, a special place is taken by research on the creation and application of systems for monitoring the environment around a human being. One of the directions of this type of monitoring is the development of sensors for the analysis of the gas composition of air, which not only determine the concentration of a particular target gas, but are able to recognize this gas in the composition of the mixture. This makes the improvement of gas analysis methods and their application in everyday life and industry a relevant task.

Traditional methods for analyzing the composition of the air environment are based on statistical methods of information processing, in particular, such fundamental ones as quantitative and qualitative analyses. In this case, qualitative analysis is aimed at identifying the various gases that are present in the air, and quantitative analysis is aimed at determining the concentration of the analyzed type of gas.

Various types of gas sensors are used to analyze the composition of the air [1–3]. They are all united by the fact that they all convert the concentration of the analyzed gas into an electrical (rarely into another) signal. Such measurements are relatively simple and,

in fact, do not contain any mathematical processing, except for the simple conversion of concentration into current or voltage. To perform this conversion, the sensors are calibrated to the target gas.

This approach limits the capabilities of all types of gas sensors, in particular, it requires accurate knowledge of the type of gas in the air. This is due to the fact that, as a rule, gas sensors do not have selectivity; able to detect various gases, including at the same time. The problem of low selectivity is related to the approach itself for traditional measurements, when the amount of data is insufficient.

In particular, the thermal catalytic sensor is calibrated at two points for measurements: at zero and at a known gas concentration. Thus, in fact, the final temperature of the sensor at a known concentration is measured, i.e. the result is only one measured temperature value. This does not allow using this method for gas mixtures of unknown composition, since the sensor will respond to all combustible gases in the air.

There are two ways to solve the problem of selectivity: by improving the sensors (for example, by using filters that pass only the target gas, catalysts, etc.) [4] or by developing mathematical methods for processing the obtained data.

The importance of using mathematical methods for analyzing big data is growing in the context of the rapid development of the technical and technological bases, as well as the digital economy. A special place in research is taken by multisensor systems [5], which consist of a set of sensors and a recognition system. In many cases, the number of sensors exceeds 3, which increases the number of gas samples under investiga-

tion. When analyzing data obtained from a large number of sensors, one should choose the most convenient data processing method, most often the principal component method (PCA) [6–8] and linear discriminant analysis (LDA) [7] are used.

In [7], the results of the PCA and LDA analyses for the responses of arrays of sensors to nitrogen dioxide, ammonia, and chlorine are presented. The areas of localization of responses to gases are different, which makes it possible to recognize which gas affects the system. Thus, the PCA well recognizes the responses of the sensor to different gases.

Studies in [8] demonstrated that using the obtained quantitative characteristics of sample differences for automatic classification using multivariate statistics methods is practical. This result shows the fundamental possibility of using the PCA to separate the types of gases under study.

The objective of this study is to assess the possibility of using PCA as an alternative method for processing data from thermal catalytic sensors, in particular, to measure the concentration of methane in air for sensors operating at different temperatures. The importance of these studies is related to the optimization of the parameters and characteristics of thermal catalytic sensors, in particular, sensitivity and selectivity, as well as the need to change measurement methods to obtain more data for subsequent mathematical processing and increase the selectivity of sensors.

EXPERIMENTAL PROCEDURE

In the study of complex chemical systems, two approaches can be distinguished [6]:

- construction of models in the form of systems of differential or integro-differential equations and application of special solution methods to them;
- data analysis based on multivariate formal modeling.

Formal modeling is often effective in cases when it is not clear how to build a meaningful model, or its construction requires excessive calculations.

The principal component analysis is based on projection methods. These methods allow for the identification of hidden variables in big data sets and analysis of the relationships that exist in the system.

An important place in the methods of multivariate data analysis is taken by the PCA. The central concept of the method is the main component. This is a special type of hidden variable that describes the maximum variety of data. It cannot be declared explicitly and directly measured. In a mathematical sense, the main component is a linear combination of all the explanatory features. In this case, the first main component includes the maximum of the total variance of explanatory features, the second — the maximum of the

remaining variance without taking into account the influence of the first, etc.

Applying the PCA, the new model has a significantly smaller number of variables, due to which this approach can be interpreted as a projection one.

For the study of industrial thermal catalytic sensors, which are commercially produced by NTC IGD company, were used [9]. The sensor consists of working and comparative elements, that are located in the same housing (Fig. 1).

Fig. 1. Photo of a thermal catalytic sensor.
 а — the inner view of the sensor (black cylinder — working sensor, white cylinder — comparison element);
 б — the appearance of the sensor housing

Usually, for measurements, the sensor is heated to the operating temperature at which combustion of the target gas occurs. The operating temperature of the sensor can vary over a fairly wide range and depends on the current flowing through the sensor. Since combustion is accompanied by the release of heat, the temperature of the sensor changes. In turn, this leads to a change in the sensor resistance, which is recorded by the measuring circuit. This single measured value is the response of the sensor.

In this study, the temperature dependence of the sensor response was studied. The technique for measuring sensor responses described in [10] was applied to obtain experimental dependences of the sensor response on the supply voltage applied to the bridge circuit for methane with a concentration of 0.47% vol. and 2.5% vol. Three groups of sensors were studied at an operating temperature of 409 °C, 444 °C and 478 °C, caused by the operating current of 46 mA, 50 mA and 54 mA, respectively. At a constant operating voltage (2.8 V in the bridge measuring circuit [11]) a varied operating current is achieved due to the different initial resistance of the sensor. Further, to describe the results, we will operate sensor numbers in accordance with Tab. 1.

Tab. 1. Characteristics of the studied groups of sensors

The initial data, i.e. the measured temperature dependences of the sensor response were divided into three sets of values for the study:

- two applied gas concentrations vs one current value;
- one applied gas concentration vs each current value;

– two applied gas concentrations vs each current value.

After PCA was used to process each set, the most optimal set was examined using the following procedure.

Due to the fact that the method is sensitive to the values of the original data with a large range [12], the data must be standardized (1):

$$s_i = \frac{x_i - \bar{x}}{\sigma},$$

where s_i is the standardized value;

x_i is initial value;

\bar{x} is the arithmetic mean of the initial value;

σ is the standard deviation of the original value.

Standardized data is processed using the PCA. Principal components have three informative indicators: the standard deviation of the principal component, the fraction of explained variance, and the cumulative fraction of explained variance [13].

Based on these indicators, a search is made for the optimal presentation of data, i.e. determination of the number of principal components sufficient to explain the initial data. There are three search rules: the Kaiser rule (those components are significant which eigenvalue is greater than 1), "the scree" plot (the optimal amount is defined as the number at which the increase in explained variance becomes negligible), and the rule of thumb (the number of principal components depends on the cumulative percentage of explained variance, which must be at least 80%).

After choosing the number of principal components, it is necessary to consider the initial data weights, with which the initial data enters the obtained principal components, and determine the most significant one. This allows us to determine which variable has the most impact on the principal component.

Thus, the study was carried out in the following sequence:

- initial data standardization;
- application of the PCA to standardized data;
- determination of three indicators of the informativity of the principal components;
- search for the best data representation (number of principal components);
- determination of the variable that has the most impact on the selected principal component.

EXPERIMENTAL RESULTS

Fig. 2 shows S-shaped dependences of the response of the thermal catalytic sensor on the applied supply voltage in a steady atmosphere containing methane.

Fig. 2. Experimental dependence of the sensor response on the supply voltage for methane concentration of 0.47% vol. and 2.5% vol.

a — sample No. 1,

б — sample No. 2,

в — sample No. 3

The obtained dependences reveal that each sample with concentration of 0.47% vol. and 2.5% vol. changes in almost the same way (the sensor temperature affects the stage of the beginning of the catalytic reaction and the stage of saturation), which makes it possible to unambiguously determine what concentration of methane is in the air. Fig. 2, в, shows that the sample No. 3 sensor response with a concentration of 2.5% vol. is 10 mV less than for samples No. 1 and No. 2.

Fig. 3 shows the experimental dependences of the sensor response on the supply voltage at one concentration. It is important to note that all sensors with each methane concentration show a close response value in the case of the traditional measurement method use at a voltage of 2.8 V.

Fig. 3. Dependence of the sensor response on the supply voltage for three samples.

a — with a methane concentration of 0.47% vol.,

б — with a methane concentration of 2.5% vol

As can be seen from the dependences obtained, there is no sensor response up to a supply voltage of 2 V for samples No. 1 and No. 2 and 1.6 V for sample No. 3. At the same time, the nature of the change in the response of the sensor of three samples is similar to each other both at a concentration of 0.47% vol. and at 2.5% vol. It is assumed that the influence of the operating temperature on the response will be minimal, this will allow for ignoring the operating current of the sensor in studies for reliable recognition of the type of gas.

Based on the above experimental dependences, the use of one sensor makes it possible to graphically recognize and analyze the gas concentration, but with an increase in the number of sensors in the system, the amount of data obtained increases, and this complicates the analysis process. In particular, this study considers a system consisting of 8 sensors. In such systems, it is important to organize the input data and evaluate which features have the most impact.

DISCUSSION OF RESULTS

The processing of the initial data is carried out with the Python 3.9.5 using the Numpy v1.20.3, Pandas 1.2.4, Sklearn 0.0 libraries, as well as in the Rstudio 1.4 programming environment.

The results of processing a set of values of two concentrations for the same current using the PCA are shown in Fig. 4.

Fig. 4. Dependence of the values of the PCA 1 on the PCA 2 at methane concentrations of 0.47% vol. and 2.5% vol.
а — for sample No. 1; б — for sample No. 2; в — for sample No. 3

As can be seen from the obtained graphs, the areas of recognition of two concentrations can be unambiguously identified. The areas of localization are distributed from left to right, i.e. along the PCA 1. Having considered the weights of each PCA obtained during processing, it can be noted that the PCA 1 is associated with the sensor response at the supply voltage [2500, 2700] mV, since the values in this range have the largest deviation for two concentrations, and the PCA 2 — at the range [1103, 1453] mV.

As shown in Fig. 4, the dispersion of values is distributed along the PCA 2. Tab. 2 shows the values of the dispersion of the original data and the PCA 2.

Tab. 2. Dispersion of initial data and PCA

The obtained result shows that the dispersion is directly related to the values of the original data. There is a single outlier in sample No. 2 at a concentration of 0.47% vol., as a result of which the dispersion is greatly increased. Similarly, presented in sample No. 3.

As a result, despite the dispersion of the values of each of the PCAs, one can conclude that recognition of concentrations is possible for one current.

The results of processing a set of values of one concentration for three currents using the PCA are shown in Fig. 5.

Fig. 5. Dependence of the values of the PCA 1 on the PCA 2 for three sensor samples.
а — at methane concentration of 0.47% vol.; б — at methane concentration of 2.5% vol.

It can be seen from the resulting graph that recognition areas in the case of the same concentration but different current values can be identified, but they have intersections. In this case, it is possible to use such data for recognizing one concentration, but not for separation into current values.

The result of processing a set of values of two concentrations for three currents using the PCA is shown in Fig. 6.

Fig. 6. The dependence of the PCA 1 values on the PCA 2 at methane concentrations of 0.47% vol. and 2.5% vol.

As can be seen from the graph, the calculated data at a concentration of 0.47% vol. formed a single region of recognition. Similarly, a region of 2.5% vol. concentration can be identified, although there is a large dispersion between the data of each sensor at a temperature of 409 °C, the deviation averaged ~5 mV. From this, we can conclude that the dependence on the current is minimal during the procession of the initial data, therefore, the values at the same concentration and for different currents can be combined into one data set.

Consider the informativity of this set of values using three rules for finding the optimal data representation. The eigenvalues of the principal components according to the Kaiser rule (significant are those components which eigenvalue exceeds 1) are presented in Tab. 3.

Tab. 3. The eigenvalues of principal components

As can be seen from the data obtained, according to the Kaiser rule, three principal components are sufficient to represent the initial data.

The values of the cumulative percentage of explained variance are presented in Tab. 4.

Tab. 4. The cumulative percentage of explained variance

As a rule of thumb, the cumulative percentage of explained variance should be at least 80%. Based on the data in Tab. 4, two principal components for three data sets are sufficient to represent the initial data.

According to "the scree" plot, the optimal number is defined as the number at which the increase in the explained variance becomes insignificant. The dependence of the fraction of the explained variance on the number of principal components is shown in Fig. 7.

Fig. 7. Dependence of the fraction of explained variance on the number of principal components

As can be seen from the dependences obtained, the increase in the explained variance becomes insignificant when the third principal component is reached, therefore, it can be assumed that the optimal number of principal components is 3.

The obtained results according to three rules showed that in order to explain the initial data in each of the three sets, the optimal and required number of principal components is 3.

CONCLUSION

In this work, an alternative method for processing data from thermal catalytic sensors based on the principal component method for measuring the concentration of methane in air (0.47% vol. and 2.5% vol.) was studied using sensors with different operating temperatures (409 °C, 444 °C and 478 °C). The presented research methodology makes it possible to find the optimal set of initial data, determine the number of principal components, and evaluate the effect of data on the principal components.

It is shown that when processing the initial data using the PCA, the influence of the operating temperature is minimal, which makes it possible to combine the values of the same concentration into one data set. The technique makes it possible to unambiguously distinguish the responses of samples with different methane concentrations without the use of additional data preparation operations.

The results obtained are of substantial importance for the development and optimization of methods for analyzing the air composition, in particular, they make it possible to switch from calculating the concentration “inside” the sensor to mathematical processing of the data received from the sensor and, on its basis, to recognize the supplied gas. Since PCA is referred to as an unsupervised machine learning method, combining several sensors into a single complex leads to the creation of a gas type recognition system capable of learning from the data obtained in the process. The issue of the possibility of using PCA in the analysis of a mixture consisting of two or more types of gas also remains topical.

REFERENCES

1. Baranov A.M., Osipova T.V. [Recent trends in the development of sensors for pre-explosive concentrations of flammable gases and vapors of flammable liquids (review)]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2021, vol. 31, no. 4, pp. 3–29. DOI: 10.18358/np-31-4-i329 (In Russ.).
2. Williams D.E. Electrochemical sensors for environmental gas analysis. *Current Opinion in Electrochemistry*, 2020, vol. 22, pp. 145–153. DOI: 10.1016/j.coelec.2020.06.006
3. Singh D., Dahiya M., Kumar R., Nanda C. Sensors and systems for air quality assessment monitoring and management: A review. *Journal of environmental management*, 2021, vol. 289, art. 112510. DOI: 10.1016/j.jenvman.2021.112510
4. Ma Y., Kaczynski J., Ranacher C., Roshanghias A., Zauner M., Abasahl B. Nano-porous aluminum oxide membrane as filtration interface for optical gas sensor packaging. *Microelectronic Engineering*, 2018, vol. 198, pp. 29–34. DOI: 10.1016/j.mee.2018.06.013
5. Sysoev V.V., Zyuryukin Yu.A. [Electronic Nose Multi-Sensor Gas Recognition Systems: review]. *Vestnik SGTU* [Bulletin of Saratov state technical university], 2007, vol. 2, no. 1, pp. 111–119. (In Russ.). URL: <https://elibrary.ru/item.asp?id=12793666>
6. Suleimanov A.M. [Analysis of experimental data by principal component method]. *Izvestiya KGASU* [News of the Kazan State University of Architecture and Engineering], 2005, no. 1 (3), pp. 81–83. URL: [https://izvestiya.kgasu.ru/files/N1\(3\)2005/Sulejmanov_81-83.pdf](https://izvestiya.kgasu.ru/files/N1(3)2005/Sulejmanov_81-83.pdf) (In Russ.).
7. Kravchenko E.I., Petrov V.V., Varezchnikov A.S. [The development of gas recognition technique using multisensor system for air monitoring]. *Inzhenernyi vestnik Dona (setevoe izdanie)* [Electronic scientific journal "Engineering Journal of Don"], 2012, vol. 23, no. 4, part 2. URL: <http://www.ivdon.ru/ru/magazine/archive/n4p2y2012/1346> (In Russ.).
8. Belozertsev A.I., Cheremisina O.V., El-Salim S.Z., Manoylov V.V., Zarutskiy I.V. [Algorithms for data processing in gas analytical complexes with semiconductor sensors for detecting contaminants in rocket fuel components]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2018, vol. 28, no. 1, pp. 18–29. DOI: 10.18358/np-28-1-i1829 (In Russ.).
9. Karpov-sensors. *Proizvodstvo termokataliticheskikh sensorov goryuchikh gazov* [Production of thermocatalytic flammable gas sensors]. URL: <http://karpov-sensor.com/> (accessed 10.01.2022). (In Russ.).
10. Ivanov I.I., Baranov A.M., Talipov V.A., Mironov S.M., Kolesnik I.V., Napolskii K.S. [Development of effective sensors for detecting pre-explosive H₂ concentrations]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2021, vol. 31, no. 3, pp. 25–36. DOI: 10.1016/j.snb.2015.07.082 (In Russ.).
11. Somov A., Baranov A., Suchkov A., Karelin A., Mironov S., Karpova E. Improving interoperability of catalytic sensors // *Sensors and Actuators B*. 2015. Vol. 221. P. 1156–1161. DOI: 10.1016/j.snb.2015.07.082
12. Moroz G., Lyashevskaya O., Shchurov I. *Lektsiya: Metod glavnykh komponent* [Lecture: Principal Component Method]. URL: <http://math-info.hse.ru/f/2015-16/ling-mag-quant/lecture-pca.html> (accessed 04.12.2021). (In Russ.).
13. Baimuratov I.R. *Metody avtomatizatsii mashinnogo obucheniya* [Methods of machine learning automation]. Saint-Petersburg: Universitet ITMO, 2020. 40 p. (In Russ.).

Contacts: *Osipova Tatiana Vladislavovna*,
t.osipova.95@mail.ru

Article received by the editorial office on 02.02.2022