МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

Федеральное государственное бюджетное образовательное

учреждение высшего образования

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»



Факультет прикладной математики и информатики

Кафедра ТПИ

Дисциплина: «Методы и технологии анализа больших данных»

Лабораторная работа №2

**СОКРАЩЕНИЕ РАЗМЕРНОСТИ ДАННЫХ**

Факультет: ФПМИ

Группа: ПМИМ-31

Выполнили: Монгуш Н. С., Тарулин М. А., Филипенко Ю. Д.

Преподаватель: Сивак М. А.

Дата выполнения:

Отметка о защите:

**Задание**

1. Используя метод главных компонент, выполнить сокращение размерности данных.

**Ход работы**

Метод главных компонент позволяет уменьшить число переменных, выбрав самые изменчивые из них.

Алгоритм нахождения главных компонент:

1. Формулу для первой главной компоненты можно записать в виде:

где – коэффициенты, которые характеризуют разброс значений вдоль оси при условии, что ;

,

где – исходные значения переменных;

– среднее для исходных переменных.

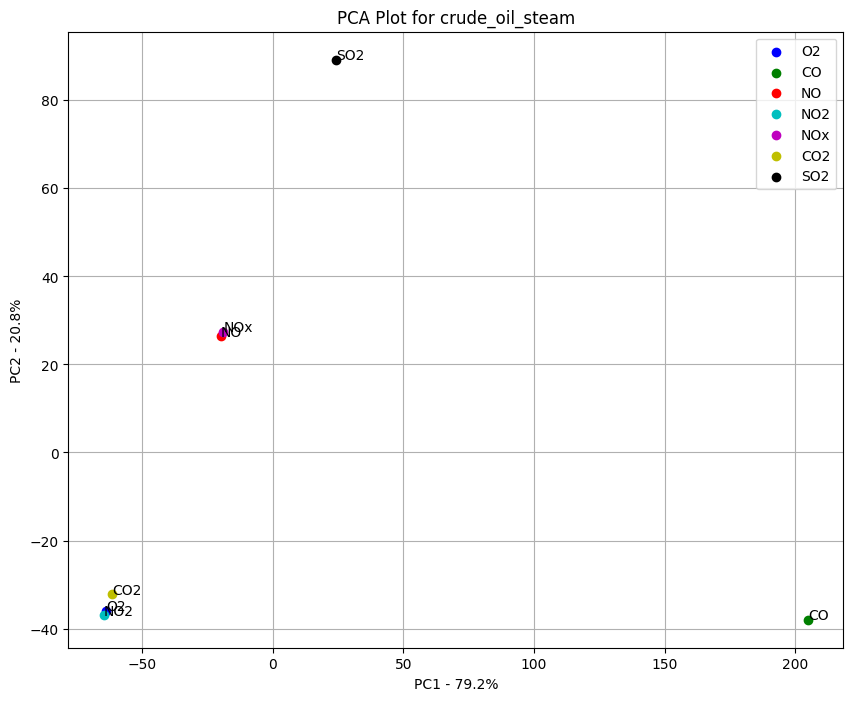


Рис. 1. График метода главных компонент для сочетания мазут-пар.

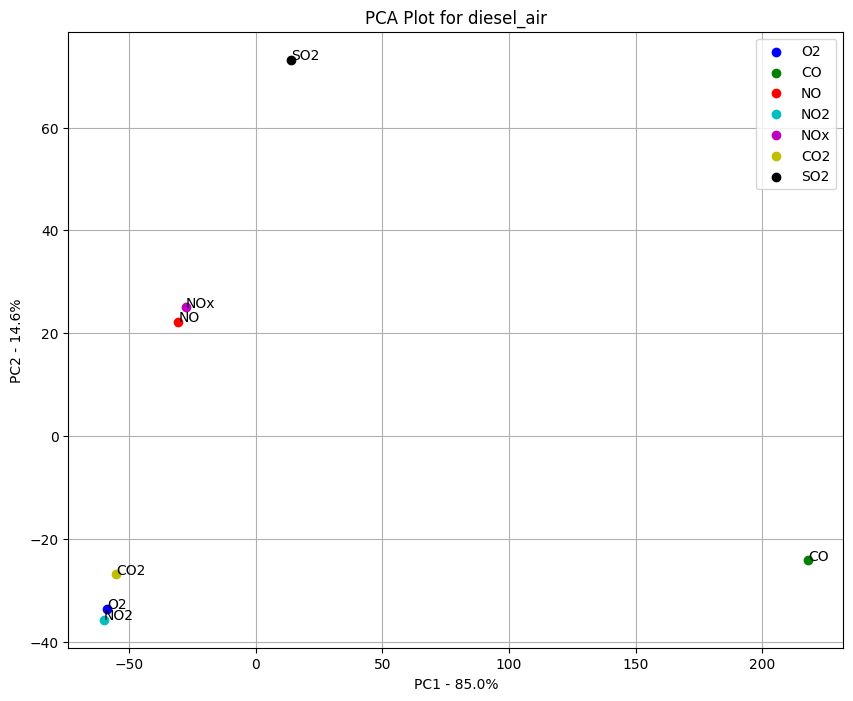


Рис. 2. График метода главных компонент для сочетания дизельное топливо-воздух.

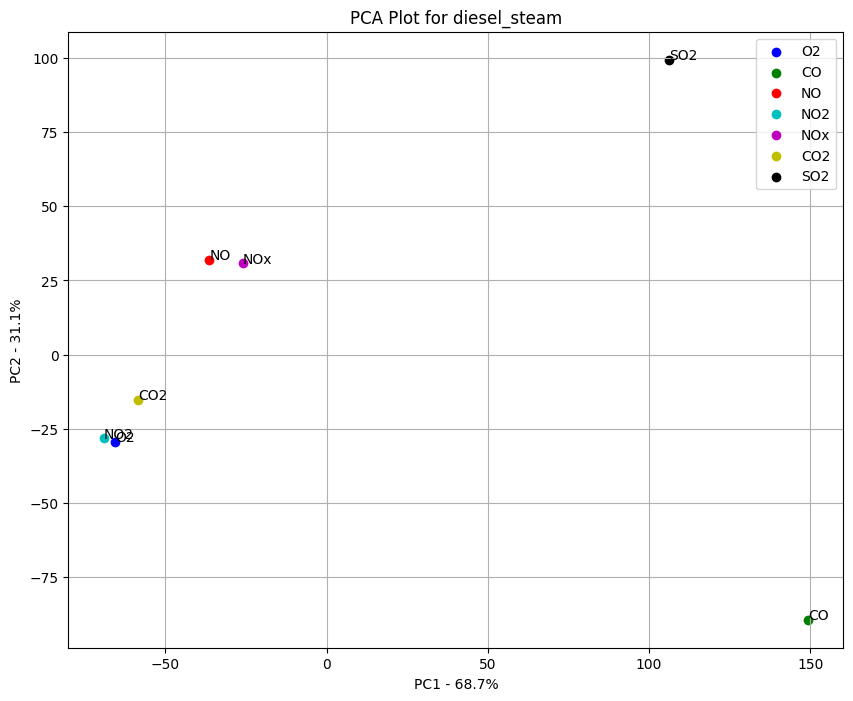


Рис. 3. График метода главных компонент для сочетания дизельное топливо-пар.

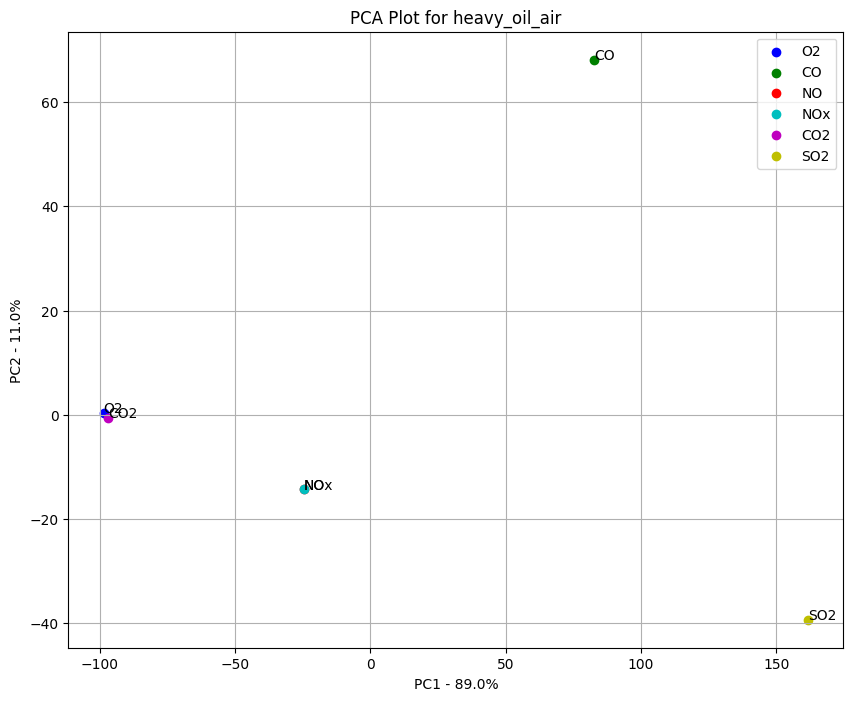


Рис. 4. График метода главных компонент для сочетания нефть-воздух.

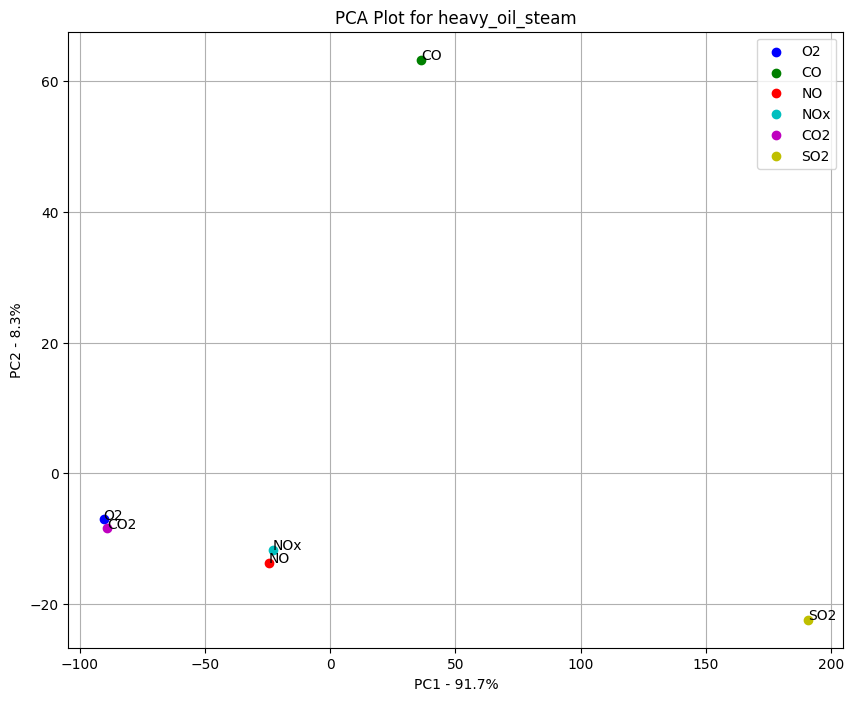


Рис. 5. График метода главных компонент для сочетания нефть-пар.

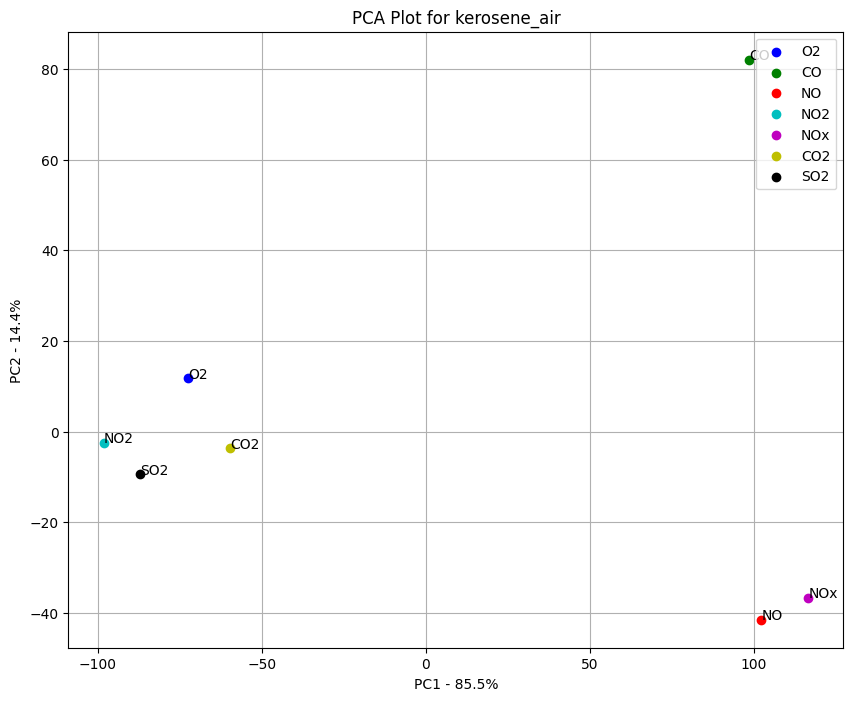


Рис. 6. График метода главных компонент для сочетания керосин-воздух.

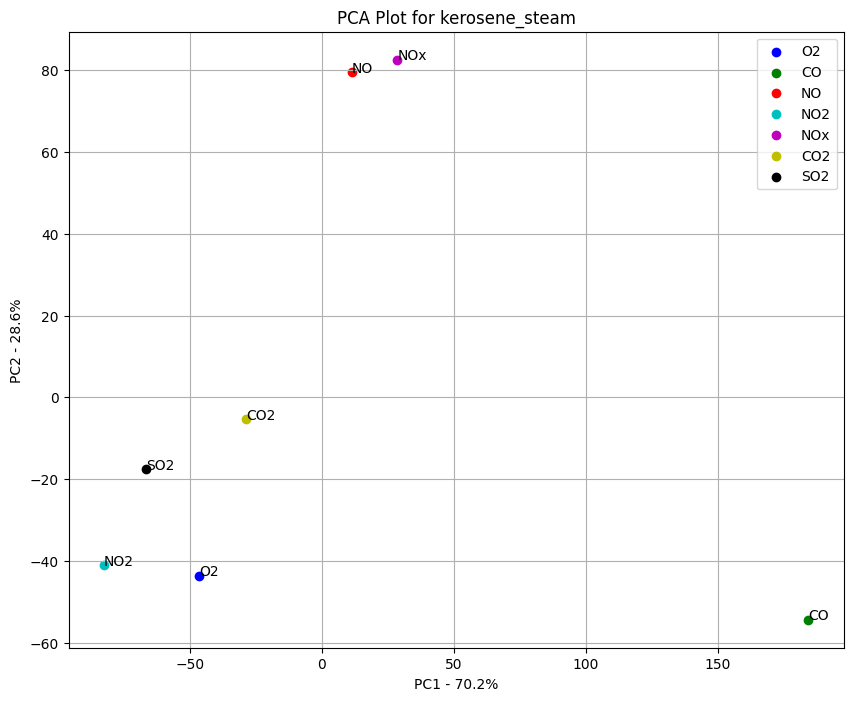


Рис. 7. График метода главных компонент для сочетания керосин-пар.

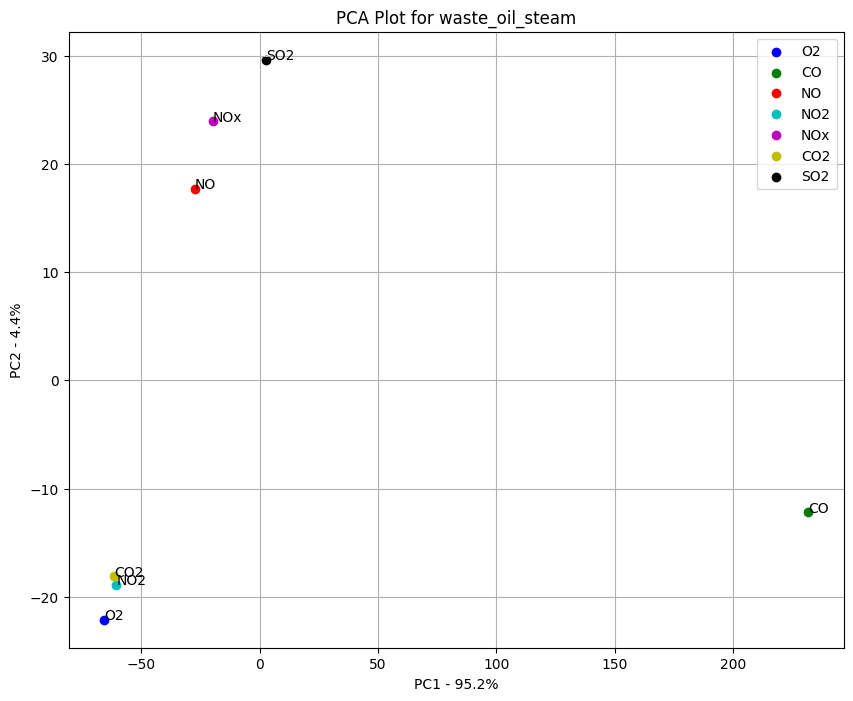


Рис. 8. График метода главных компонент для сочетания отработанное масло-пар.

Из представленных графиков можно выявить какие компоненты положительно коррелируют между собой, такие компоненты будут сгруппированы. Из особенностей можно выделить следующее:

1. для всех графиков характерно высокое положительное коррелирование компонентов O2, NO­2 и CO2, кроме сочетаний керосина, где группировка не такая плотная в сравнении с другими сочетаниями;
2. высокая плотность для компонентов NO, NO­2­ и NOX, что так же говорит о положительной корреляции.

Объясняется положительная корреляция для компонентов O2, NO­2 и CO2 тем, что NO­2 и CO2 является продуктом O2, высокая плотность для компонентов NO, NO­2­ и NOX объясняется тем, что NOx является составной частью двух компонентов: NO и NO­2.