

19-я Всероссийская конференция с международным участием
«Математические методы распознавания образов»

ДИАГНОСТИКА ВОДНО-ЭТАНОЛЬНЫХ РАСТВОРОВ ПО СПЕКТРАМ КОМБИНАЦИОННОГО РАССЕЯНИЯ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ: МЕТОДЫ ПОВЫШЕНИЯ УСТОЙЧИВОСТИ РЕШЕНИЯ К ИСКАЖЕНИЯМ СПЕКТРОВ*

И.В. Исаев^{1,2}, С.А. Буриков^{1,2}, Т.А. Доленко^{1,2},
К.А. Лаптинский^{1,2}, С.А. Доленко¹

¹ НИИ ядерной физики им. Д.В. Скобельцына,
МГУ им. М.В. Ломоносова, г. Москва.

² физический факультет МГУ им. М.В. Ломоносова, г. Москва.

** Исследование выполнено за счёт гранта Российского Научного Фонда,
проект №19-11-00333.*

Введение

Обратная задача спектроскопии

Обратная задача (ОЗ) спектроскопии - определение характеристик объекта по его спектру.

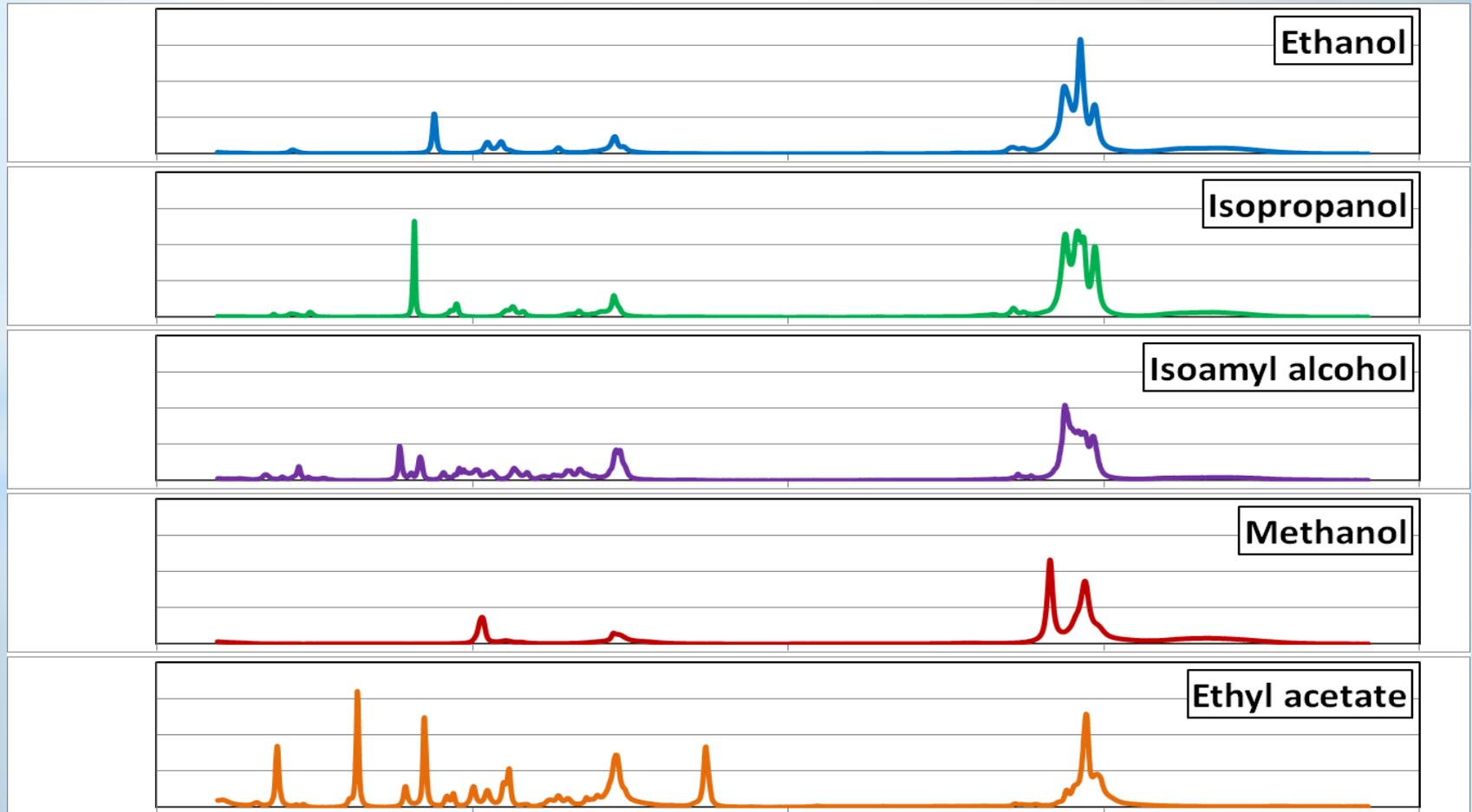
В настоящей работе рассматривается задача **бесконтактного контроля качества алкогольных напитков**: определение концентраций этанола и вредных примесей по их спектрам комбинационного рассеяния (КР).

Определяемые компоненты:

- этанол
- метанол
- сивушное масло
- этилацетат

Постановка задачи

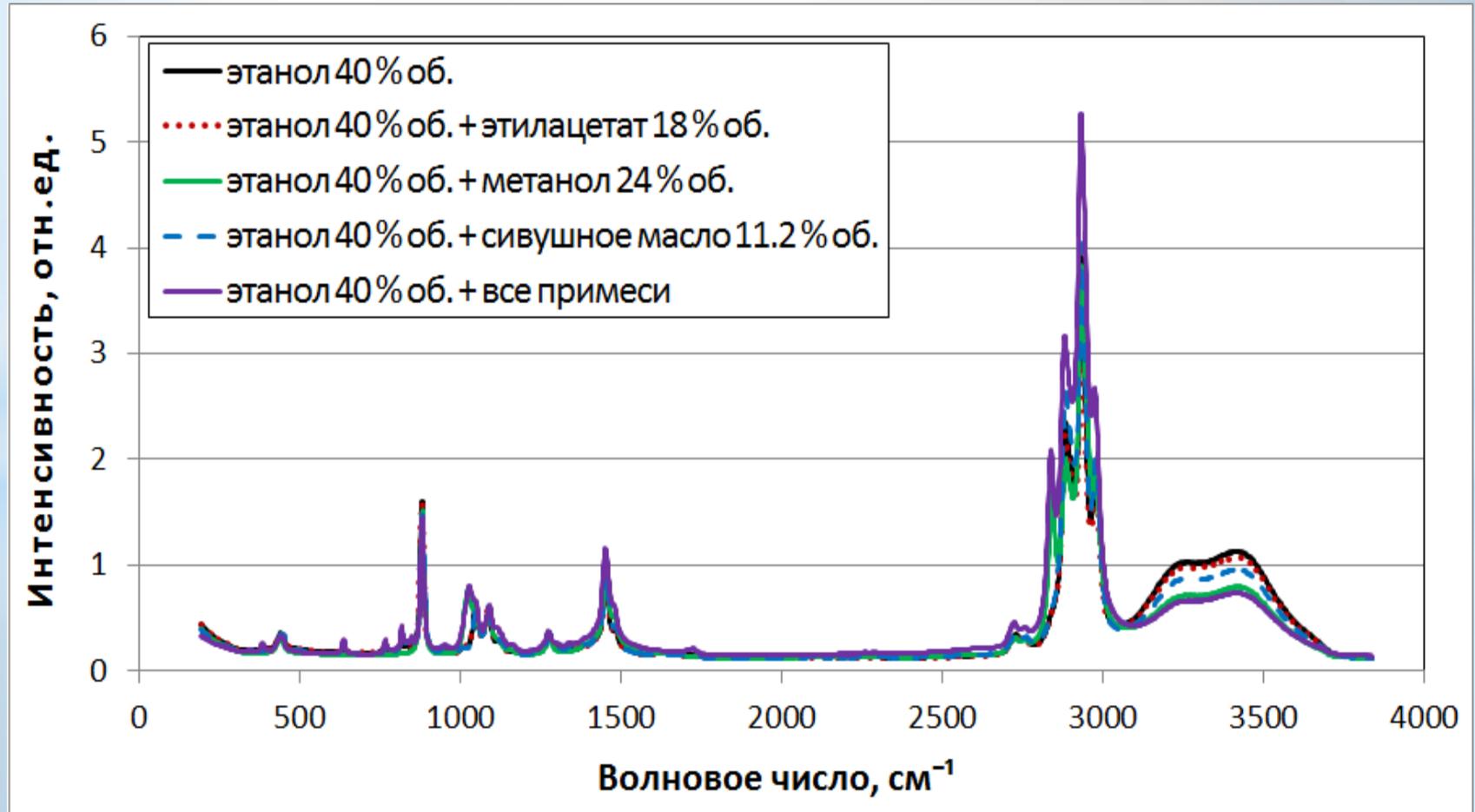
Спектры КР чистых компонентов



Характеристические линии компонентов
в значительной мере перекрываются

Постановка задачи

Спектры КР водно-этанольных растворов



Спектры КР водно-этанольных растворов оказываются **близки друг к другу**

Постановка задачи

Модельные водно-этанольные растворы

Для реализации методов машинного обучения был получен массив спектров КР водных растворов этанола с вредными примесями, моделирующих водки различной крепости.

Концентрации вредных примесей варьировались от 0 до летальной дозы.

Определяемые компоненты:

- Этанол – 35..57 %
- Метанол – 0..24 %
- Этилацетат – 0..18 %
- Сивушное масло (моделируется смесью изоамилового и изопропилового спиртов) – 0..11.2 %

Физический эксперимент

Модельные водно-этанольные растворы

- Набор данных получен **экспериментально**.
- Количество примеров: 4046 примеров
- Разбиение на наборы:
 - ✓ Тренировочный набор 70% 2800 примеров
 - ✓ Валидационный набор 20% 800 примеров
 - ✓ Тестовый набор 10% 446 примеров
- Размерность данных:
 - ✓ По входу 2048 признаков
 - ✓ По выходу 4 параметра

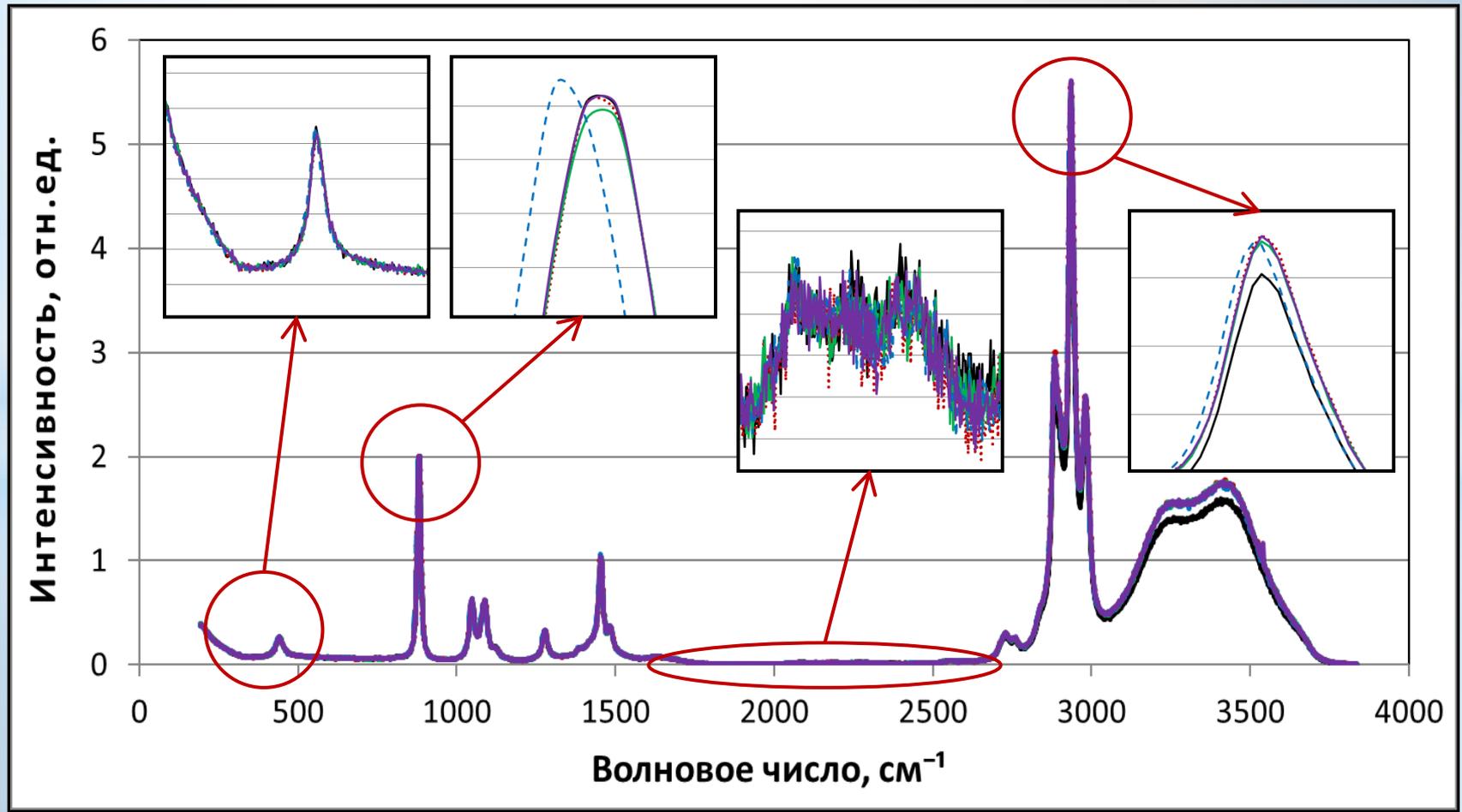
Физический эксперимент

Реальные алкогольные напитки

- Набор данных получен экспериментально.
- Количество примеров: 69 примеров
- Алкогольные напитки:
 - водка
 - джин
 - текила
 - ликеры
 - самогон
 - и т. д.

Постановка задачи

Спектры КР водно-этанольных растворов

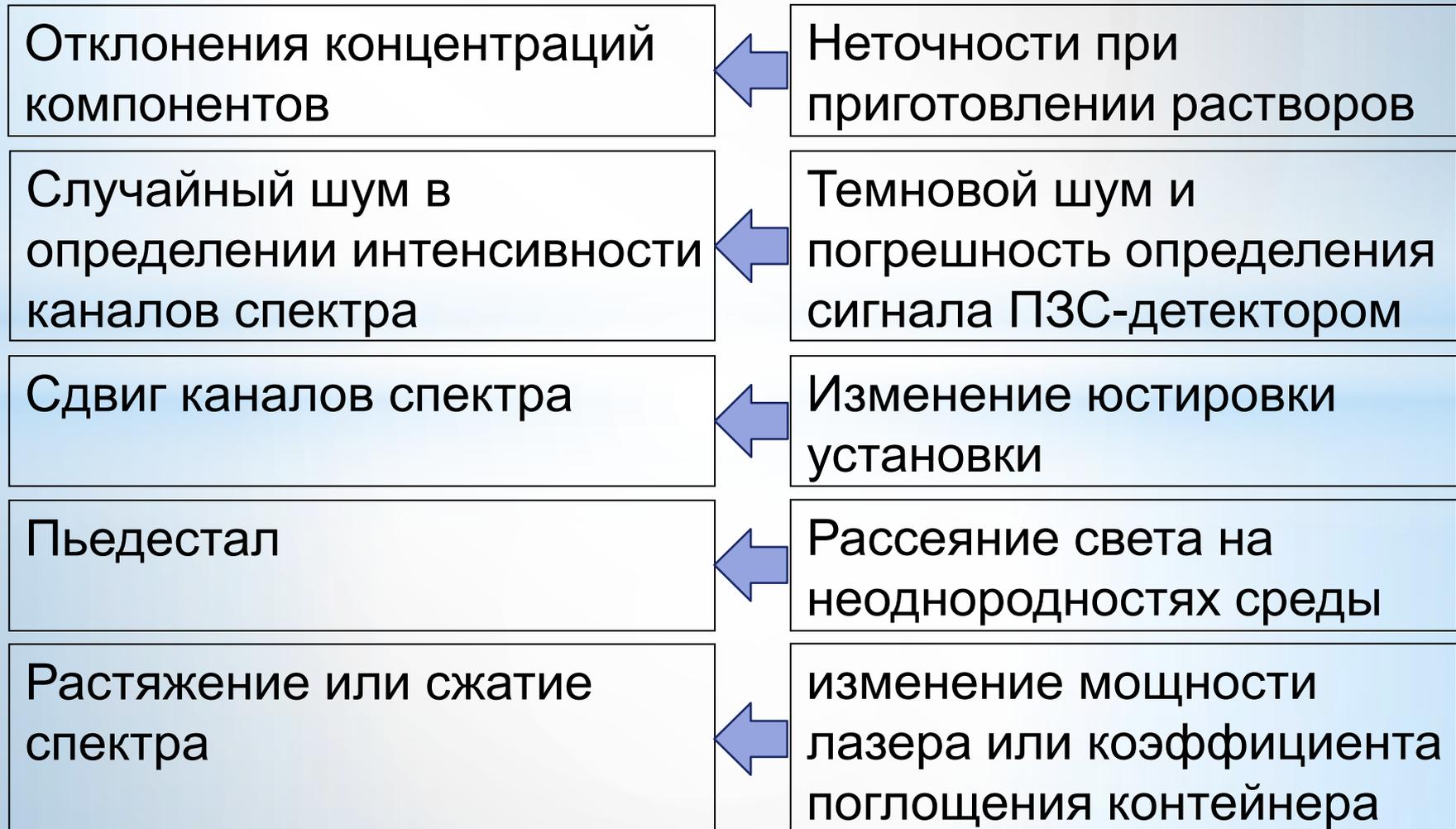


Спектры КР одного образца могут **отличаться** друг от друга

Постановка задачи

Описание искажений в данных

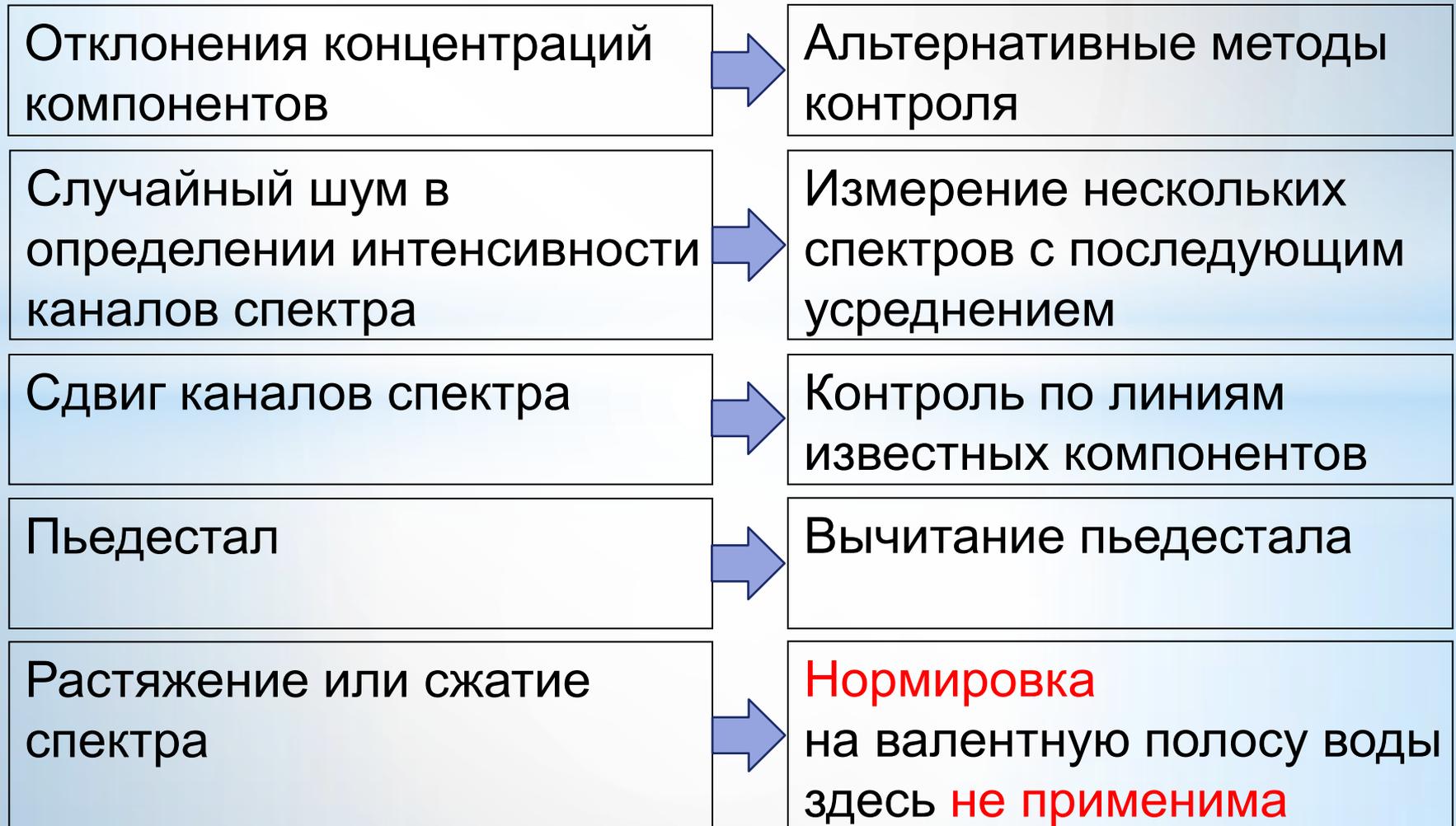
Типы искажений, содержащиеся в данных спектроскопии:



Постановка задачи

Описание искажений в данных

Типы искажений, содержащиеся в данных спектроскопии:



Постановка задачи

Цель работы

В предыдущих работах было показано, что **добавление шума** во время обучения НС позволяет **повысить устойчивость решения** к выбранному типу и уровню шума для различных ОЗ.

Цель настоящего исследования состояла в том, чтобы **проверить применимость** данного метода повышения устойчивости нейросетевого решения ОЗ к шумам в данных применительно **к ОЗ спектроскопии** водных растворов этанола в отношении такого типа искажений, как **растяжение / сжатие** спектра.

Постановка задачи

Метод добавления шума при обучении НС

В предыдущих работах было показано, что **оптимальным режимом** обучения с добавлением шума оказался такой, когда обучение проводилось на тренировочном наборе, который содержал шум, а остановка обучения осуществлялась по валидационному набору без добавления шума.

В этом случае **качество** решения было **выше,** а **время** обучения - **ниже.**

Постановка задачи

Метод добавления шума при обучении НС

Шум добавлялся в тренировочный набор перед обучением НС. Каждый пример исходного обучающего набора был представлен в 10 шумных реализациях.

Рассматривались два вида статистик искажений: использовались искажения, имеющие равномерное распределение и нормальное распределение.

Сети, обученные на тренировочном наборе с определенным уровнем шума, были применены к тестовым наборам всех уровней шума той же статистики.

Вычислительный эксперимент

Параметры нейронных сетей

- Топология: персептрон с 1 скрытым слоем.
- Количество нейронов в скрытом слое: 32.
- Передаточная функция:

скрытый слой	сигмоидальная
выходной слой	линейная
- Остановка обучения:

500 эпох без улучшения результата на валидационном наборе.
- Каждая сеть тренировалась по 5 раз с различными случайными комплектами начальных весов с последующим усреднением статистик результатов применения.

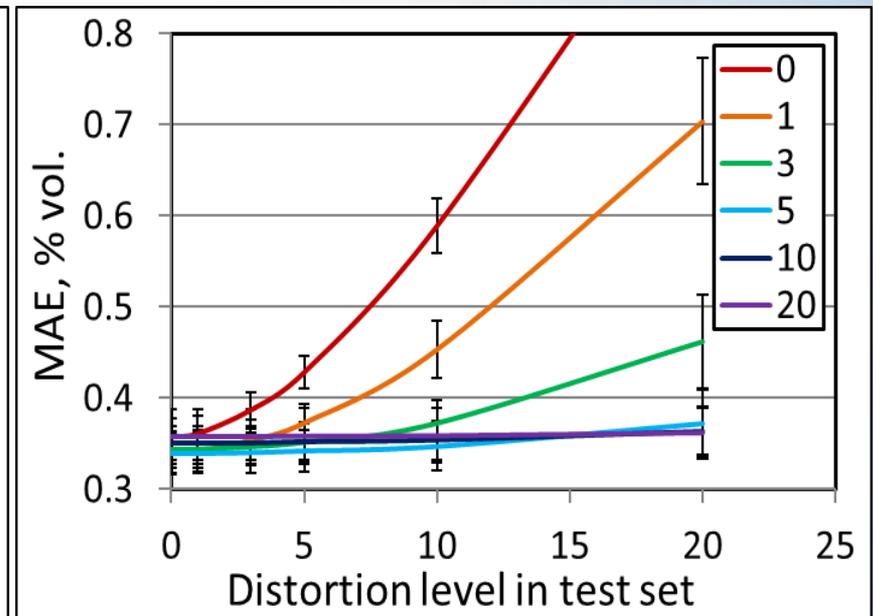
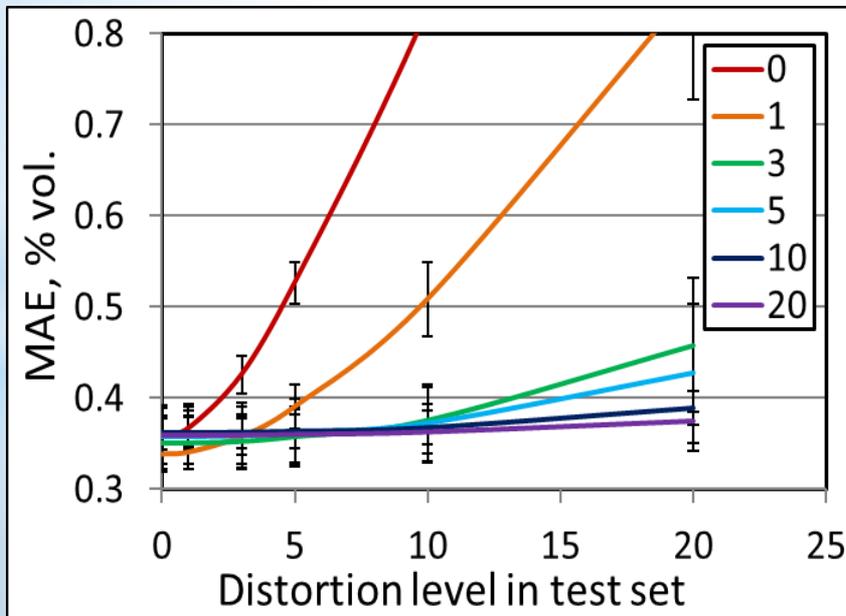
Результаты

Набор данных модельных растворов

Этанол

Нормальное

Равномерное



Чем выше уровень шума в тренировочном наборе, тем хуже сеть работает с данными без шума, но **тем медленнее деградирует качество решения** с увеличением уровня шума в тестовом наборе.

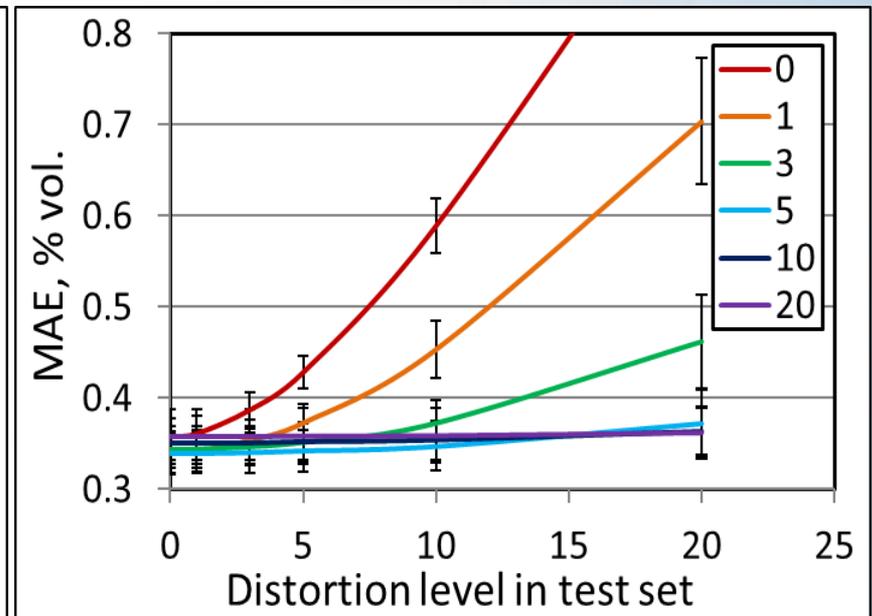
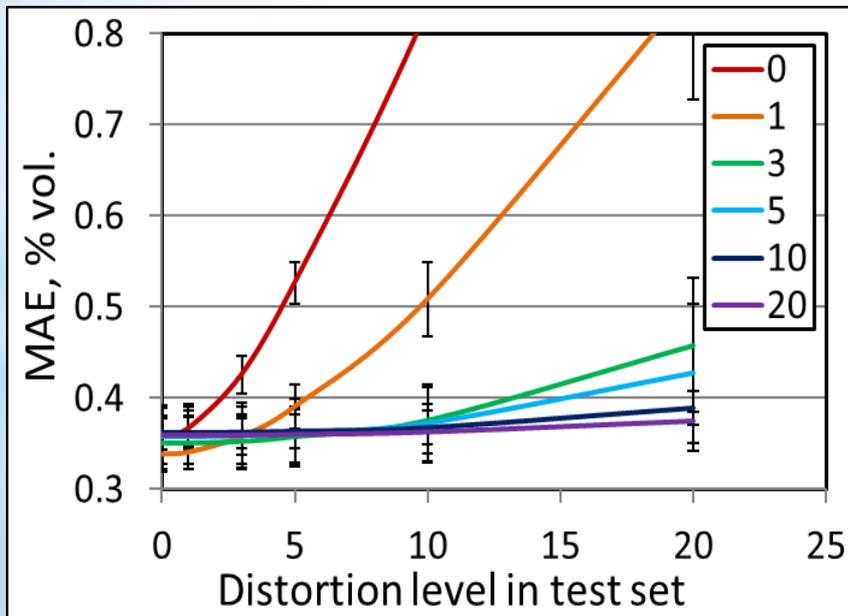
Результаты

Набор данных модельных растворов

Этанол

Нормальное

Равномерное



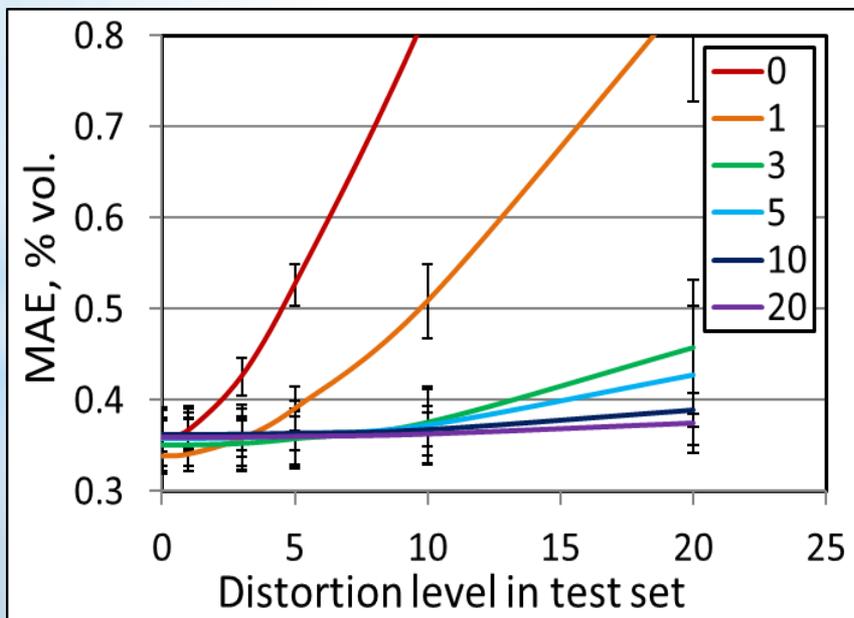
Устойчивость решения к искажениям данных **выше** для искажений, имеющих **равномерное распределение**, чем для искажений, имеющих нормальное распределение.

Результаты

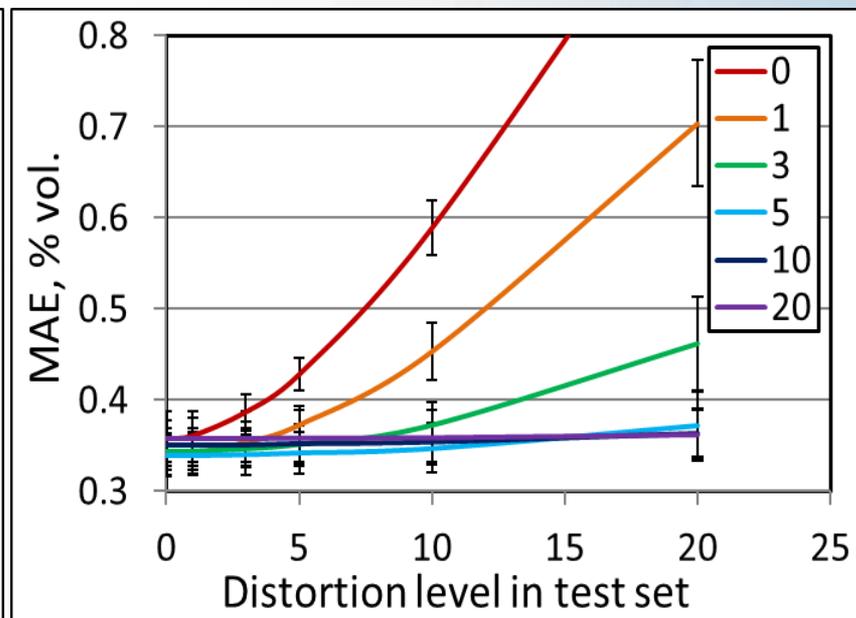
Набор данных модельных растворов

Этанол

Нормальное



Равномерное

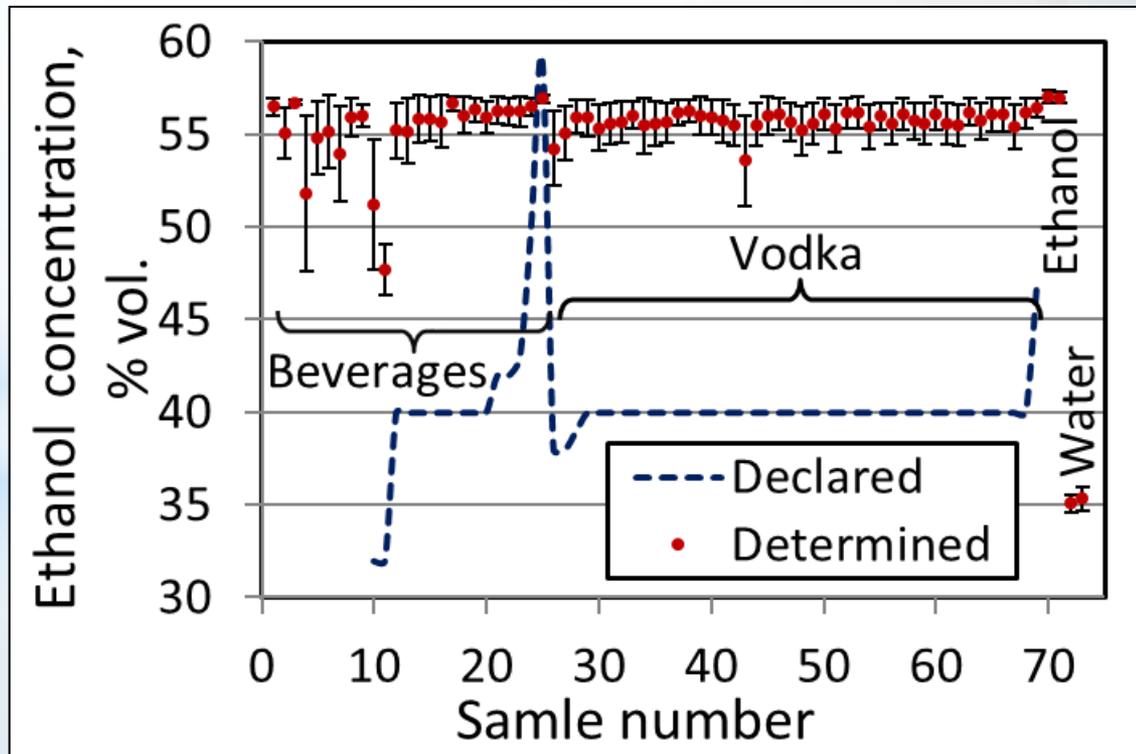


Низкий уровень ошибки может косвенно указывать на то, что набор данных является репрезентативным

Результаты

Набор данных реальных алкогольных напитков

Сети, обученные **без добавления** шума

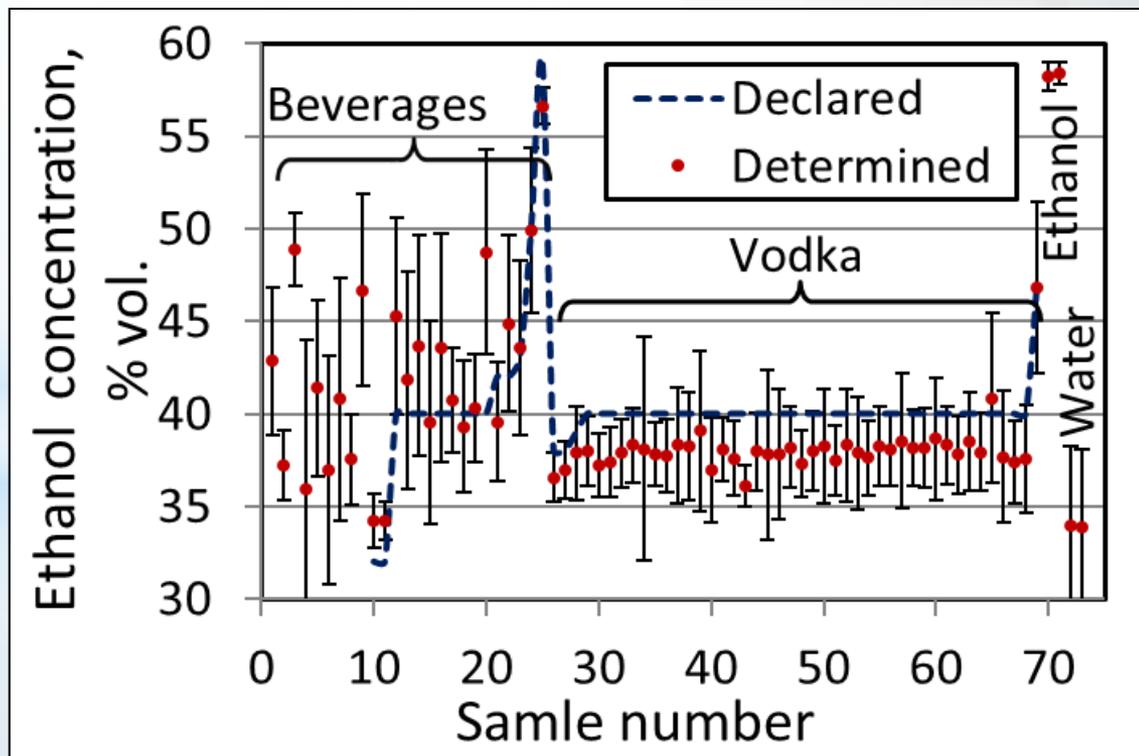


Результаты практически для всего набора данных **ушли в насыщение** - показали нижний или верхний уровень концентраций в тренировочном наборе

Результаты

Набор данных реальных алкогольных напитков

Сети, обученные с добавлением 20% Гауссова шума

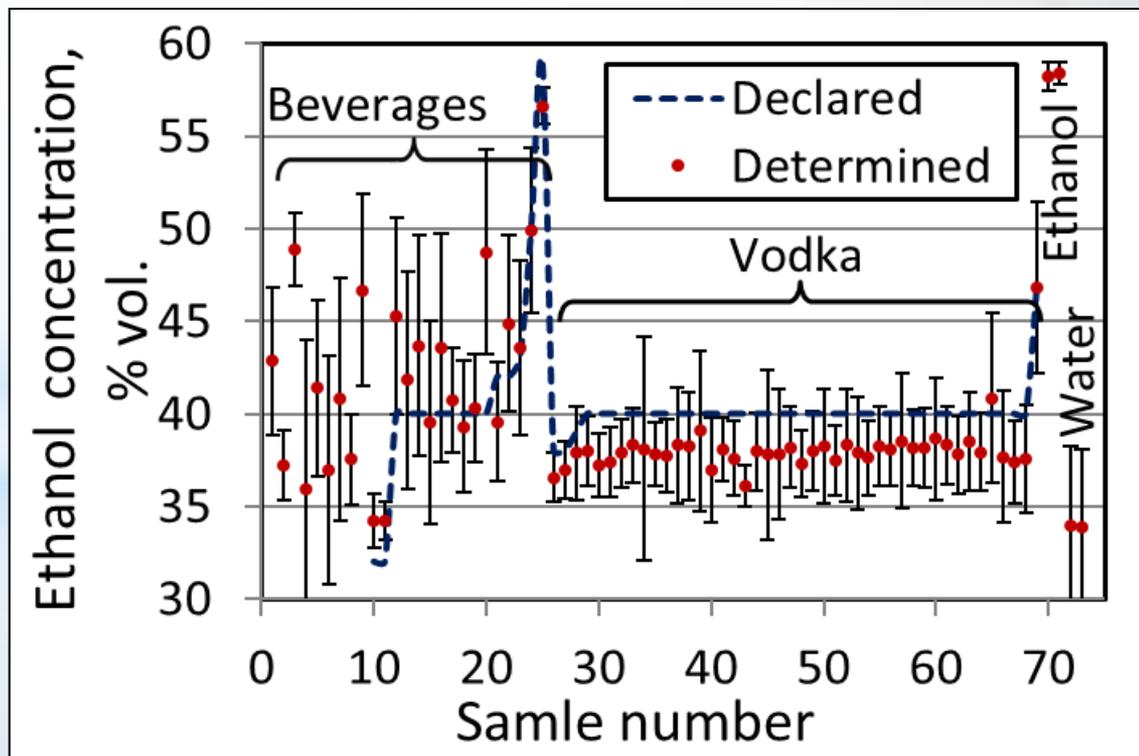


При использовании НС, обученных с добавлением шума, **результаты** определения концентраций оказались **близки к заявленным** производителями.

Результаты

Набор данных реальных алкогольных напитков

Сети, обученные с добавлением 20% Гауссового шума

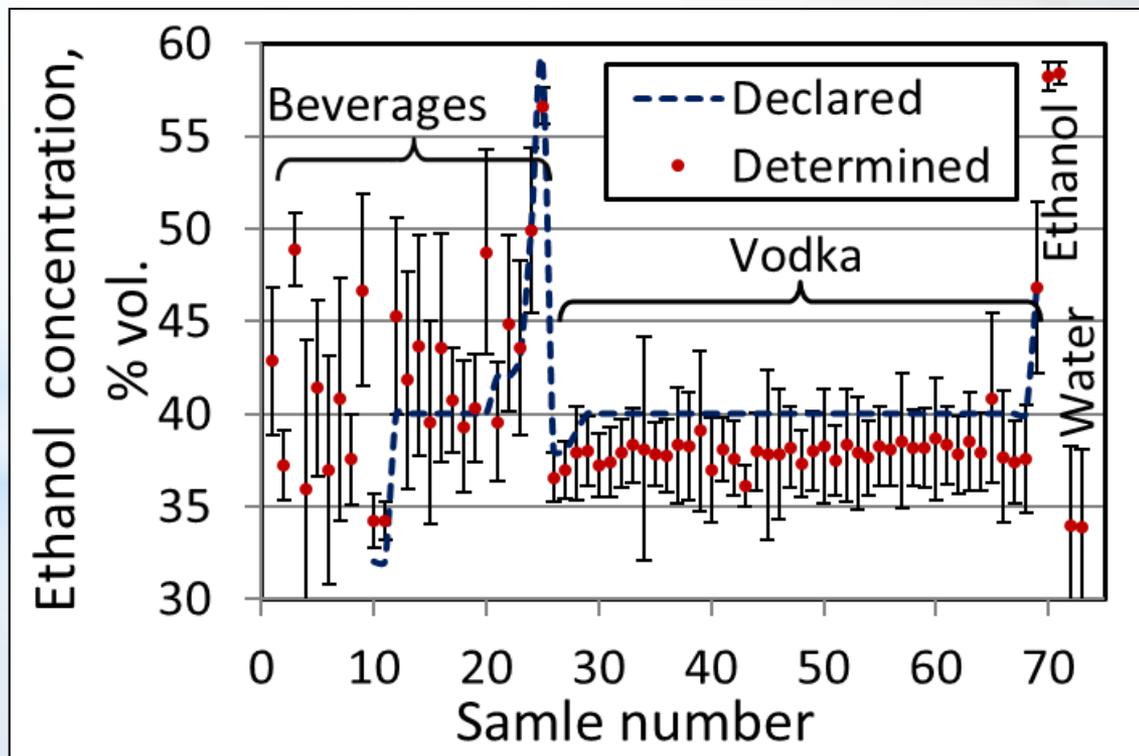


Наблюдаемая систематическая ошибка косвенно свидетельствует о наличии неучтенных искажений в наборе.

Результаты

Набор данных реальных алкогольных напитков

Сети, обученные с добавлением 20% Гауссового шума



Среднее отклонение концентрации составило 2,07% об.

Выводы

Общие выводы

- Чем выше уровень шума в тренировочном наборе, тем хуже сеть работает с данными без шума, но тем **медленнее деградирует** качество решения **с увеличением уровня шума** в тестовом наборе.
- **Устойчивость** решения к искажениям данных **выше** для искажений, имеющих **равномерное** распределение, чем для искажений, имеющих нормальное распределение.

Выводы

Частные выводы

- Спектры **реальных алкогольных напитков** **существенно отличаются** от спектров модельных растворов.
- Как следствие, сети, обученные **без добавления** шума, **не смогли** показать адекватные результаты.
- Для сетей, обученных **с добавлением** 20% Гауссового шума, **среднее отклонение** концентрации составило **2,07% об.**

Спасибо за внимание!