

Научная сессия к 75-летию НИИЯФ МГУ

17 февраля 2021 г.



Нейросетевые методы обработки данных физического эксперимента

С.А. Доленко

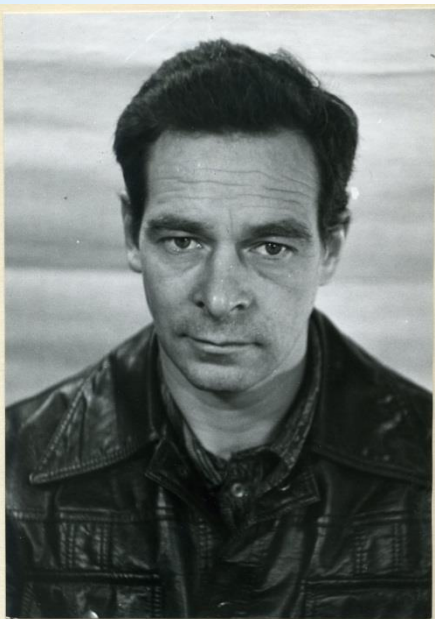
Заведующий лабораторией

Лаборатория адаптивных методов обработки данных
(ЛАМОД НИИЯФ МГУ)

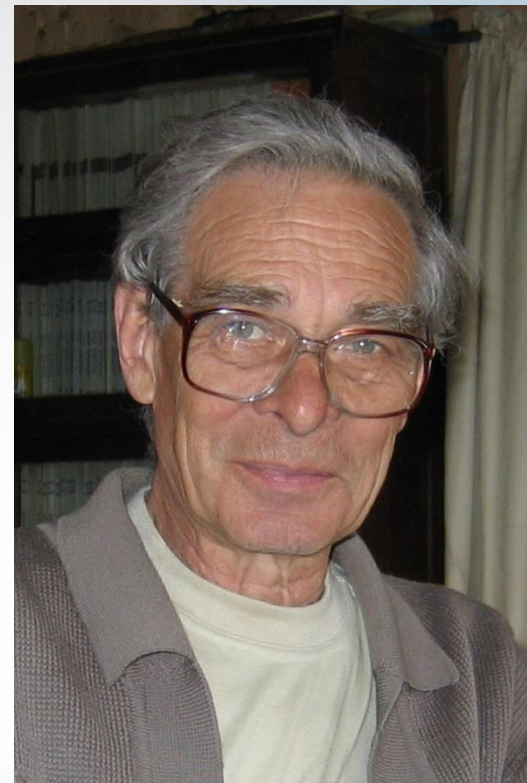
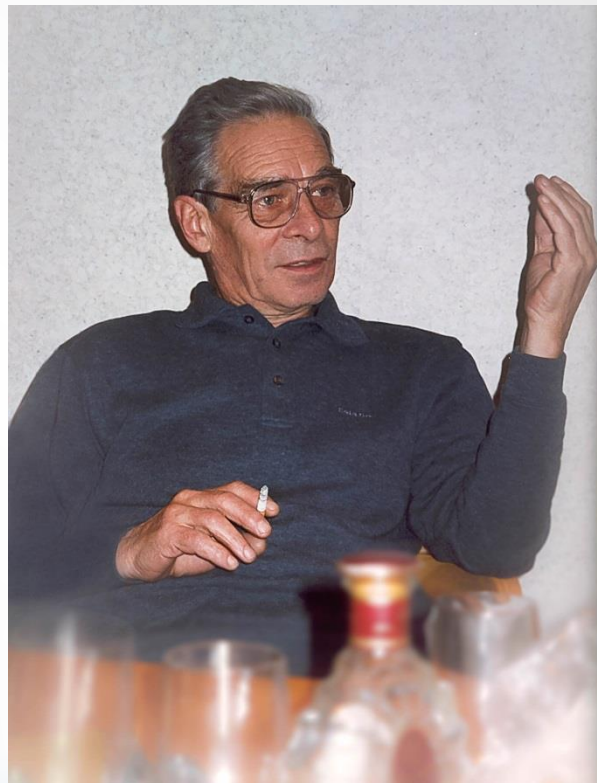
Общий план доклада

- Краткая история лаборатории ЛАМОД
- Зачем методы машинного обучения нужны физикам
- Нейронные сети и их достижения 2010-х годов: что нам подходит, а что нет, и почему
- Основные направления работы лаборатории
- Некоторые результаты, полученные за последние 5 лет
- Курс по машинному обучению и нейронным сетям

Основоположник адаптивных методов в НИИЯФ



Игорь Георгиевич
ПЕРСИАНЦЕВ
С.Н.С.



Игорь Георгиевич Персианцев

1937 – 2015

Лаборатория адаптивных методов обработки данных (ЛАМОД): история

- **1992 г.** Начало работ по нейронным сетям и адаптивным методам обработки данных в рамках лаборатории отдела микроэлектроники
И.Г. Персианцев
С.П. Ребрик
А.А. Девятов
Ю.В. Орлов
С.А. Доленко
Ю.С. Шугай
- **Ок. 2000 г.** Выделение в самостоятельную лабораторию институтского подчинения –
Лабораторию «Компьютерный класс» (ЛКК)
О.А. Агапкин
С.А. Шаров
А.Г. Гужва
- **2011 г.** Вхождение в состав отдела оперативного космического мониторинга в качестве лаборатории
В.Р. Широкий
И.В. Исаев
А.О. Ефиторов
- **2015 г.** Выделение в самостоятельную лабораторию институтского подчинения –
Лабораторию адаптивных методов обработки данных (ЛАМОД)
К.А. Лаптинский
И.В. Пластинин

Методы описания состояния объекта в физике

- **Содержательное физическое моделирование**
 - **Теоретическое** описание объекта с учётом **известных** физических закономерностей
 - Получение и использование **аналитического выражения** для интересующей исследователя физической величины
- **Численное моделирование**
 - Описание объекта **физически содержательной системой уравнений**
 - **Численное решение** системы с получением значения интересующей исследователя физической величины
- **Аппроксимационное моделирование**
 - Получение **информации об объекте в виде массива примеров**, описывающих его состояние в различных ситуациях
 - Аппроксимация описывающих объект закономерностей с получением **оценки** значения интересующей исследователя физической величины

Аппроксимационное моделирование

- Получение информации об объекте в виде массива примеров, описывающих его состояние в различных ситуациях
 - Для каждого примера должно быть известно желаемое (истинное) значение интересующей физической величины
- Выбор семейства аппроксимирующих функций, желательно из физических соображений:
 - релаксационный процесс с несколькими постоянными времени
 - разложение спектра на составляющие спектральные контуры
- Если это невозможно – выбор универсального семейства аппроксимирующих функций:
 - разложение временного ряда по базису Фурье
 - вейвлет-разложение сигналов
 - нейросетевая аппроксимация
(многослойный персептрон – универсальный аппроксиматор)

Основные классы задач обработки данных, характерных для физики

- **Задачи регрессии**
 - Задачи количественной оценки свойств
 - Прогнозирование
 - Поиск взаимосвязей и скрытых закономерностей
 - Обратные задачи
- **Задачи классификации**
 - Бинарная классификация (обнаружение событий)
 - Многоклассовая, многометочная, нечёткая классификации
- **Задачи кластеризации**
 - Моделирование по кластерам и его эффективность
- **Задачи оптимизации моделей**

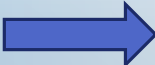
Нейронные сети и их достижения 2010-х

- Интерес к нейронным сетям носил **волнообразный характер**:
 1. Конец 1960-х (формальный нейрон, персептроны)
 2. Конец 1980-х (алгоритм обратного распространения ошибки)
 3. Вторая половина 2010-х (**глубокое обучение**)
- Некоторые **результаты третьей волны**:
 - Решена задача дикторонезависимого распознавания речи
 - Решена задача идентификации (классификации) объектов на изображениях и в видеопотоке
 - Решена задача самообучения сложным играм на основе знания лишь правил игры (шахматы, сёги, го)
- Результаты основаны на **доступности**:
 - Существенно возросших **вычислительных мощностей**
 - Практически **неограниченного объёма данных** некоторых типов

Специфика данных физического эксперимента

- **Высокая размерность** пространства входных признаков
- Сравнительно **небольшое количество примеров** (исключение – проекты класса “megascience”)
- **Нелинейность** аппроксимируемой зависимости, иногда сильная
- **Некорректность** и/или плохая обусловленность (обратные задачи)
- Существует также **специфика** конкретных типов физических задач

Всё это накладывает существенные **ограничения** на арсенал доступных методов машинного обучения. Многие «стандартные» алгоритмы глубокого обучения оказываются **непригодными**.

 Требуется разработка **специальных методик** применения методов машинного обучения к данным физического эксперимента.

Основные направления работы лаборатории

- Решение многопараметрических обратных задач (ОЗ)
 - ОЗ спектроскопии (подход «от эксперимента»)
 - ОЗ разведочной геофизики (подход «от модели»)
- Решение задач прогнозирования космической погоды
 - Геомагнитные индексы, потоки заряженных частиц на геостационаре
- Участие в когнитивных исследованиях
 - Применение методов машинного обучения для анализа данных фМРТ. Задача «Когновизор».
 - Построение взаимного отображения параметров мимики и параметров эмоционального состояния (валентность, возбуждение, доминантность)
- Обработка сигналов полупроводниковых газовых сенсоров («электронный нос»)
- Разработка алгоритмов и методик решения различных типов задач

Сотрудничество с ЛКФИ ОКН НИИЯФ, физическим, химическим факультетами МГУ, РГГРУ, ИИКС НИЯУ МИФИ, обсуждение сотрудничества с ИПИМ МГУ, НИЦ «КИ»

Обсуждение участия в школах МГУ по искусственному интеллекту и когнитивным наукам

Отбор и преобразование входных признаков

- Адаптивные методы **отбора** признаков (плюс – интерпретируемость!):
 - Отбор существенных признаков по кросс-корреляции, кросс-энтропии
 - Нелинейные адаптивные методы отбора (в осн. нейросетевые)
 - Часто методы отбора требуют задания порога отбора

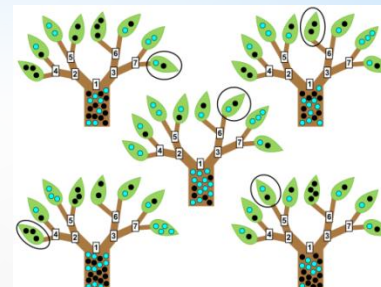
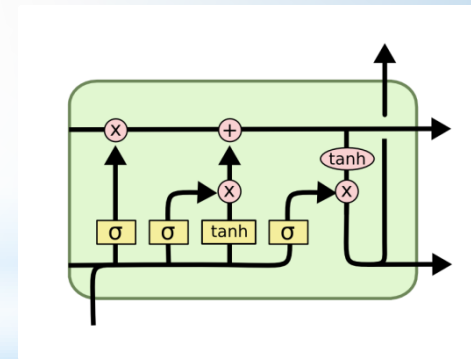
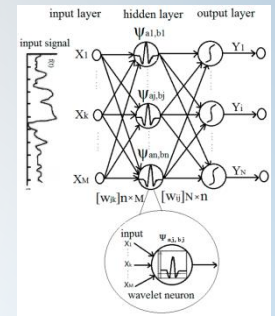
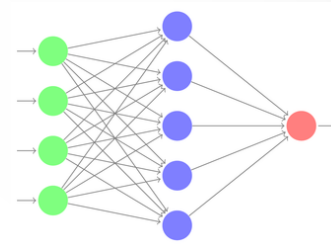
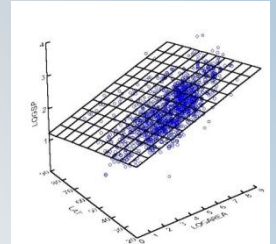
- **Зачем** нужны адаптивные методы отбора?

Нельзя ли осуществить отбор из физических соображений?

- Причинно-следственная связь не обязательна, нас устраивают **предвестники** (крик петуха годится для прогнозирования восхода)
- Разные входные признаки могут нести **частично совпадающую** информацию (физическая взаимосвязь, геометрическая близость)
- Методы **преобразования** входных признаков:
 - Анализ главных компонент (АГК), линейный и нелинейный, и методы на его основе (метод проекций на латентные структуры)
 - Отбор Фурье- и вейвлет-коэффициентов, Фурье- и вейвлет-фильтрация
 - Агрегация признаков

Используемые методы машинного обучения

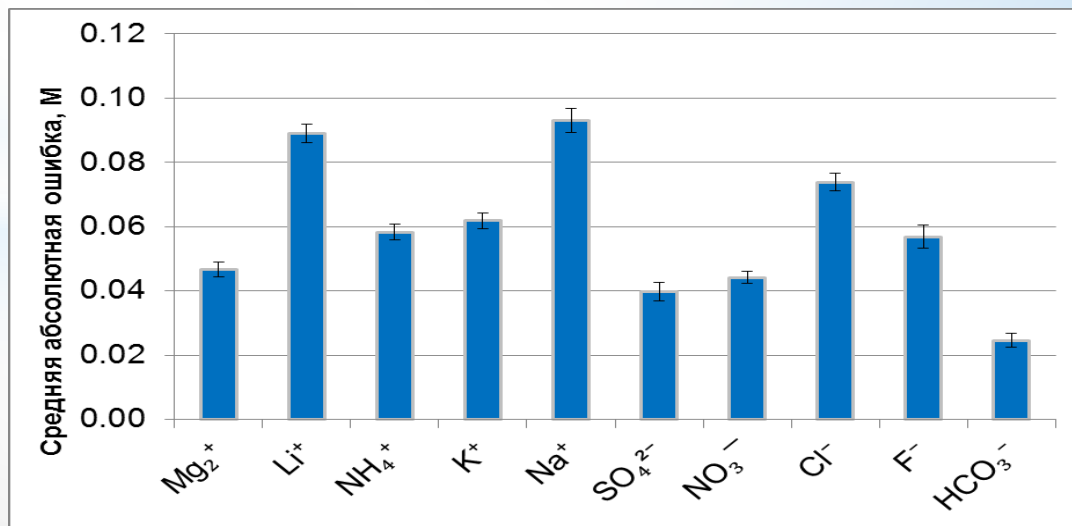
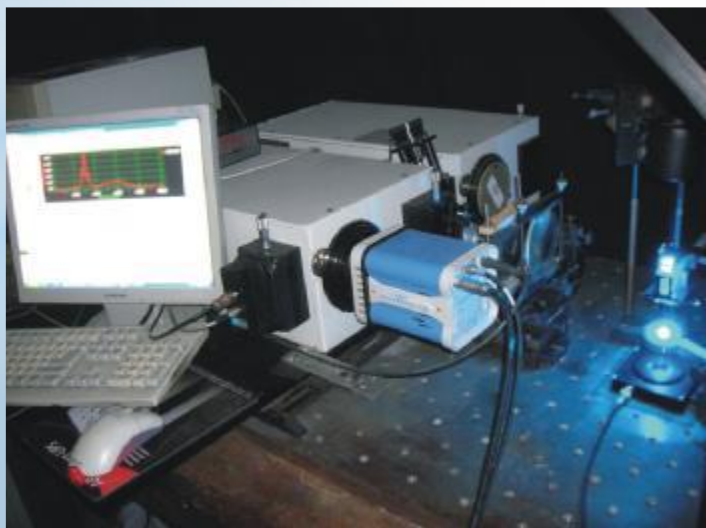
- Линейная регрессия
 - Линейная регрессия в линейном и нелинейном базисе
 - Метод проекций на латентные структуры
- Искусственные нейронные сети без обратных связей
 - Многослойные перцептроны
 - Вейвлет нейронные сети
- Рекуррентные нейронные сети
 - Сети с длинной краткосрочной памятью (LSTM)
- Метод группового учёта аргументов (МГУА)
- Адаптивные нейро-нечёткие системы (ANFIS)
- Деревья решений и алгоритмы на их основе
 - Алгоритм случайного леса
 - eXtreme Gradient Boosting
- Комитеты (ансамбли) предикторов



Определение ионного состава раствора

Решение обратной задачи определения солевого и **ионного состава многокомпонентного раствора** неорганических солей по спектрам комбинационного рассеяния света

- Растворы содержали от 1 до 5 солей, или от 2 до 10 различных ионов.
- Использование адаптивных методов понижения входной размерности данных позволяет снизить погрешность решения задачи и его вычислительную стоимость.
- Разрабатываемый метод позволит осуществлять дистанционную диагностику минеральных, технологических и сточных вод.



Обратные задачи и добавление шума

Повышение **качества** нейросетевого **решения** многопараметрической обратной задачи и его **устойчивости к шумам** разнообразной природы путём **добавления шума** в процессе обучения

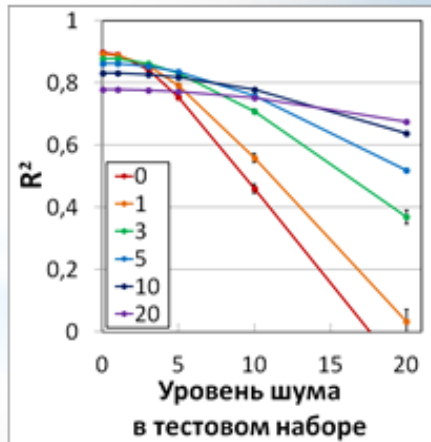
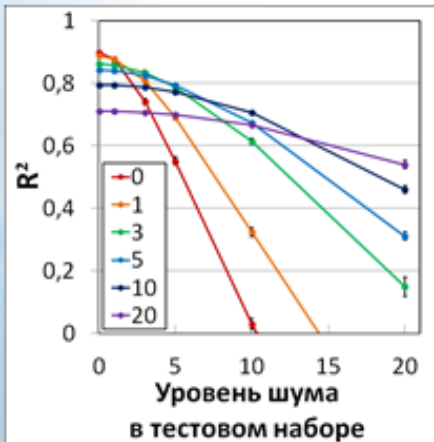
- При этом **максимальный** эффект достигается, если тип, характеристики и уровень шума при обучении соответствуют тем, которыми характеризуются независимые данные при применении обученной ИНС.

Зависимость качества решения ОЗ МТЗ от уровня шума в тестовом наборе данных для сетей, обученных на тренировочных наборах данных с различным уровнем шума

Параметр №17

Аддитивный гауссовый

Аддитивный равномерный

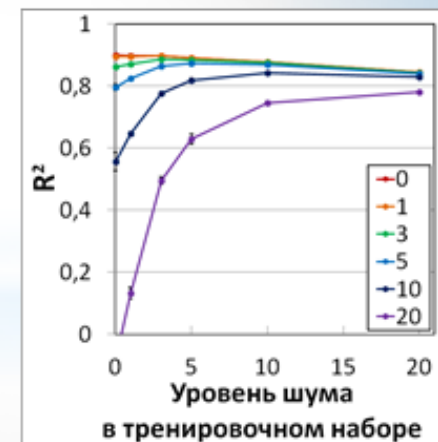
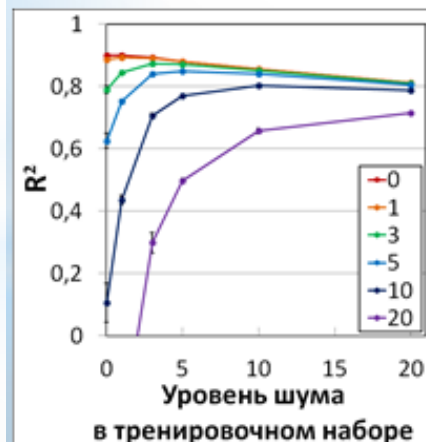


Зависимость качества решения ОЗ от уровня шума в тренировочном наборе данных для различных уровней шума в тестовом наборе данных

Параметр №83

Мультипликативный гауссовый

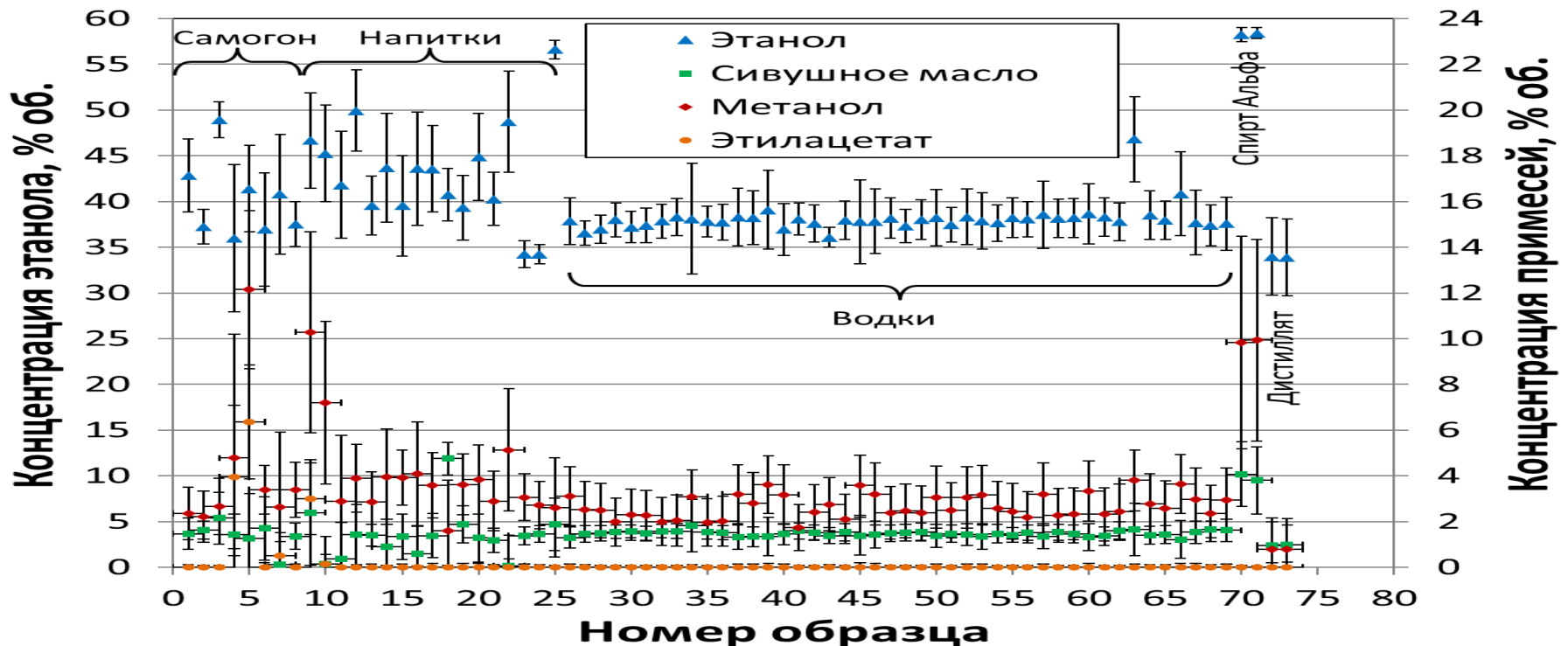
Мультипликативный равномерный



Диагностика алкогольных напитков

Продемонстрированы **перспективы** применения спектроскопии комбинационного рассеяния света с обработкой результатов с помощью искусственных **нейронных сетей** для диагностики водно-этанольных растворов и **алкогольных напитков**

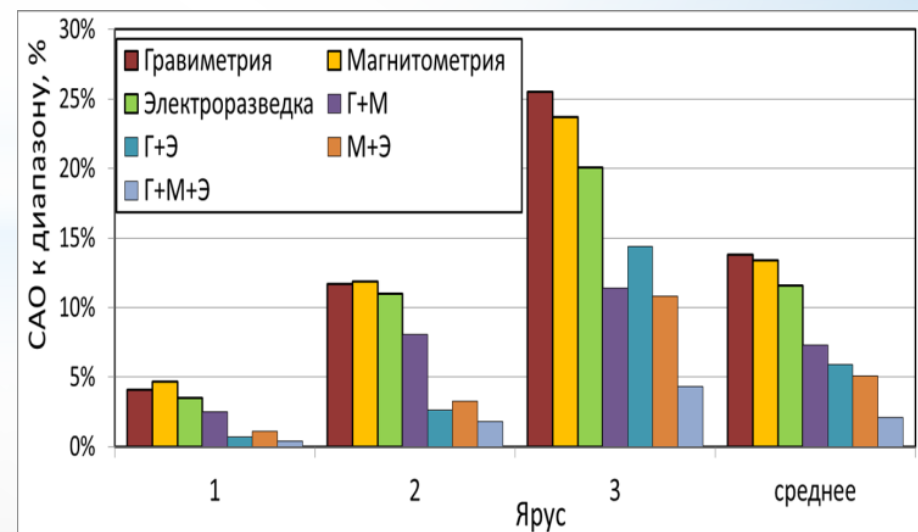
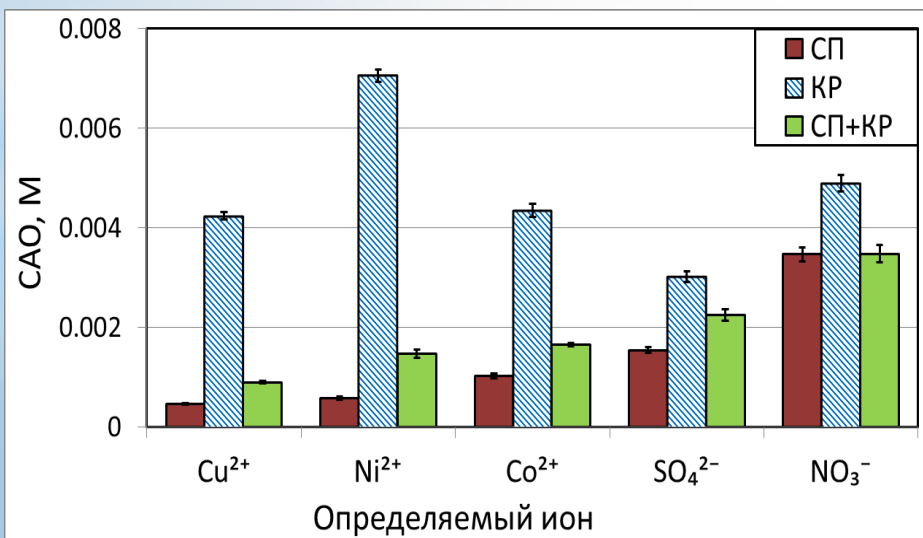
- Результаты определения концентрации этанола и примесей в алкогольных напитках с помощью НС, натренированных с добавлением шума.



Комплексирование физических методов

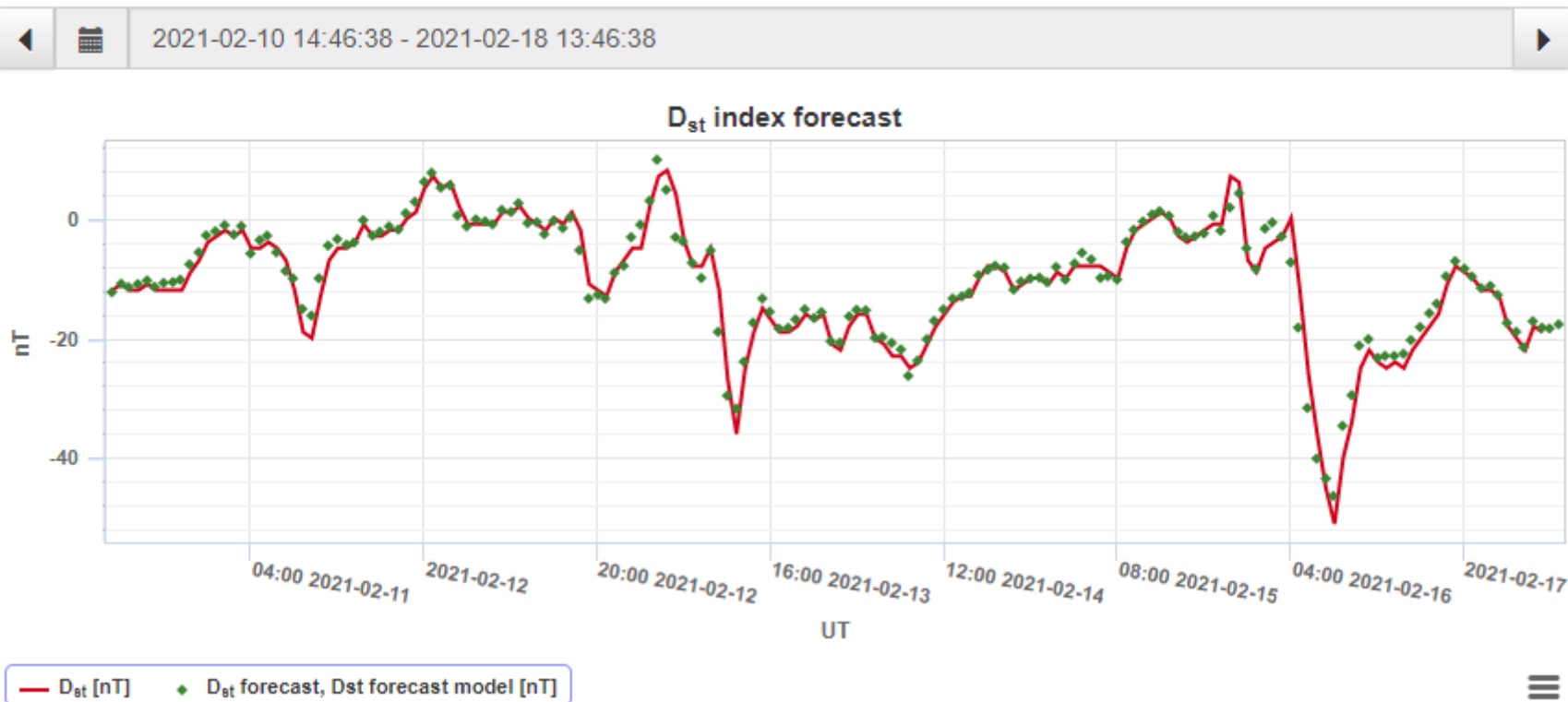
Можно ли **повысить качество** решения многопараметрической ОЗ методами машинного обучения, если решать **одновременно** несколько ОЗ с **одинаковыми** определяемыми параметрами, основанных на результатах различных **физических методов** измерений?

- Если при использовании физических методов по отдельности **один** из них дает **существенно меньшую** погрешность, чем остальные, то совместное применение **не позволяет** ее уменьшить.
- Если же методы по отдельности дают **сравнимую** погрешность, то их совместное применение **позволяет** улучшить результат.



Прогноз часовых значений индекса Dst

Прогноз Dst



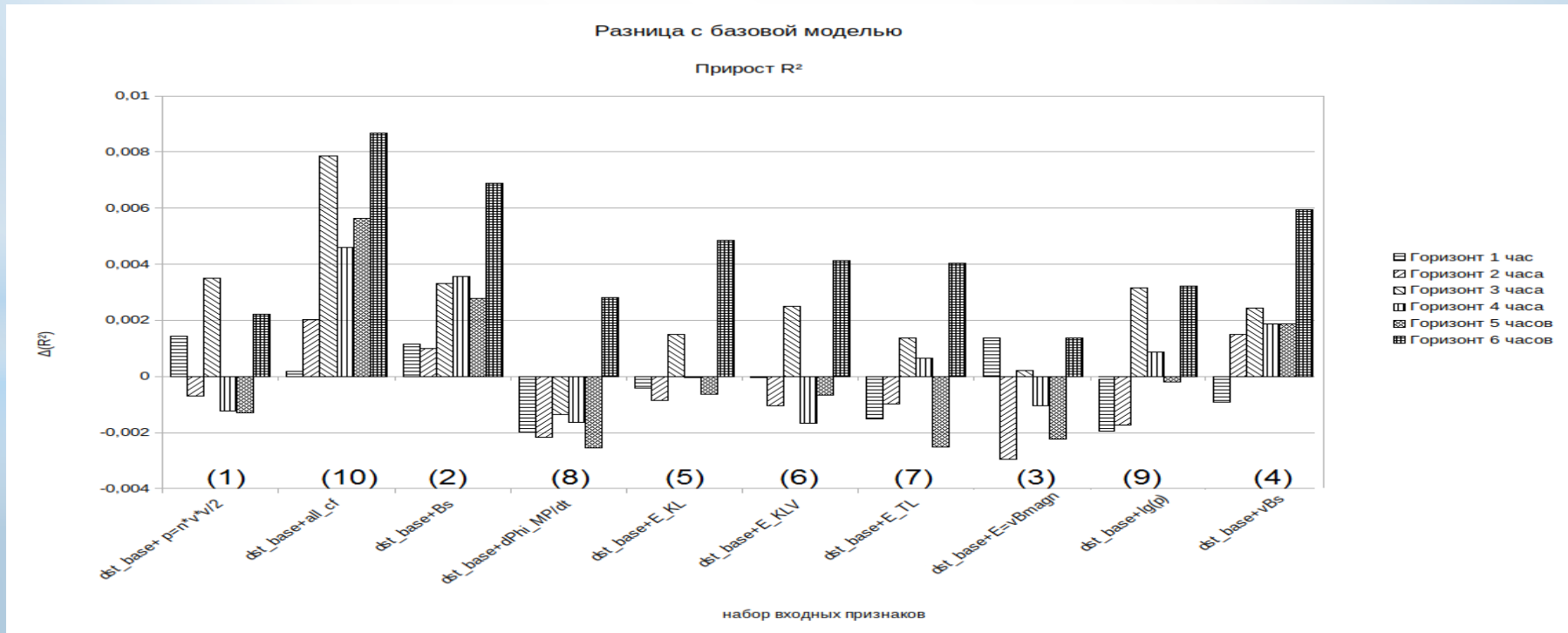
Онлайн прогноз Dst-индекса на один час вперед на сайте ЦОКМ НИИЯФ МГУ

<http://swx.sinp.msu.ru/models/dst.php?gcm=1>

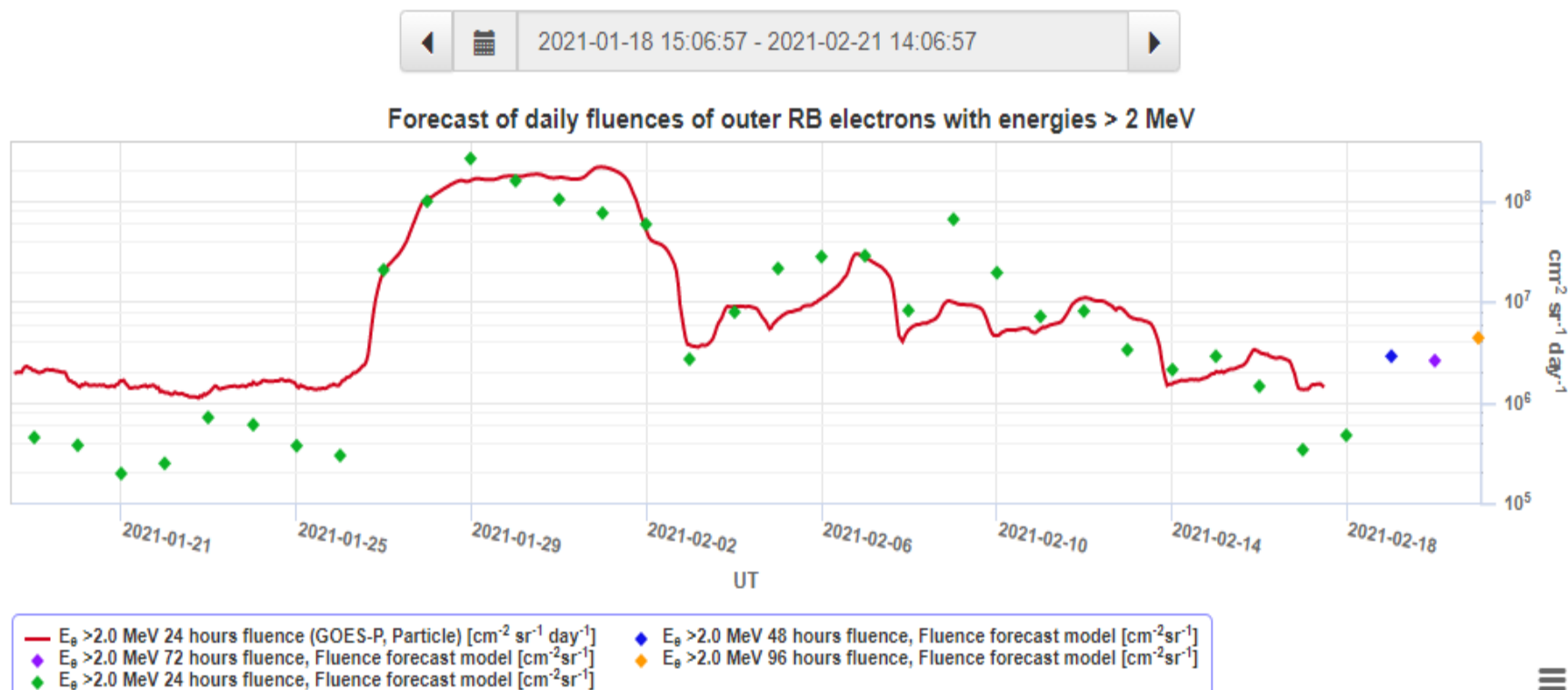
Использование «функций связи»

Использование т.н. **функций связи** (coupling functions) при прогнозировании **Dst-индекса** адаптивными методами позволяет **повысить качество прогноза**

- **Наибольший вклад** в улучшение качества прогноза вносят функции **Bs** и **vBs**, а также использование одновременно нескольких ФС.
- **Горизонт** удовлетворительной точности прогноза составляет **3 часа**.



Прогноз суточных флюенсов РЭ ВРПЗ



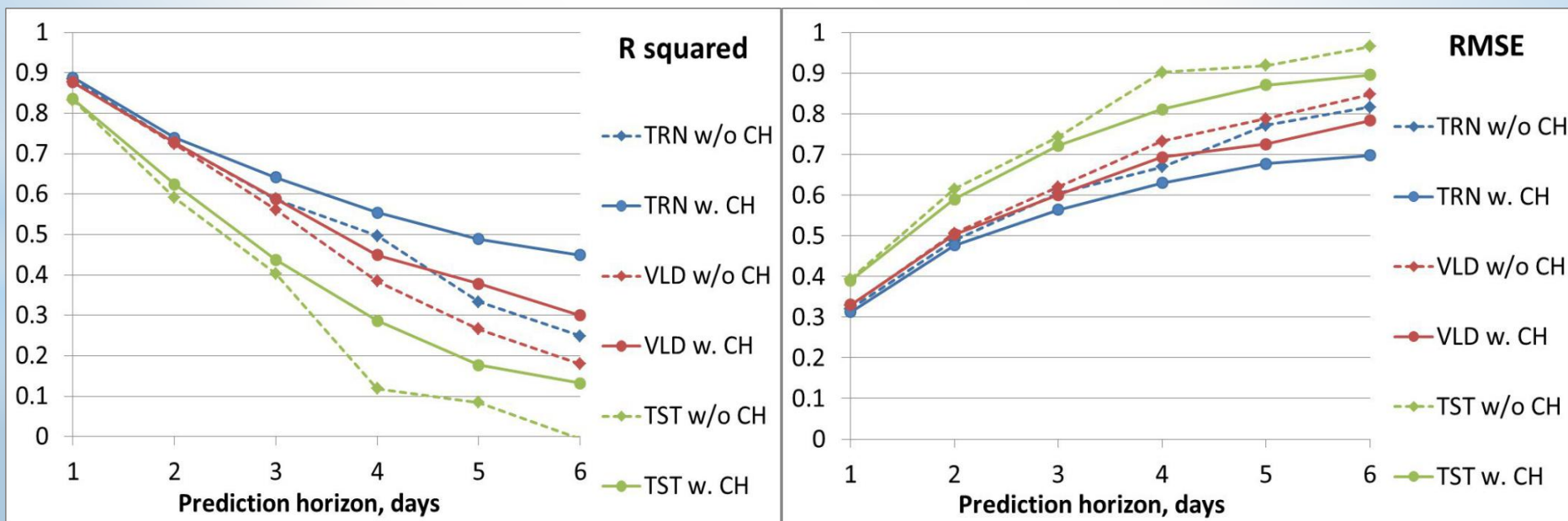
Онлайн прогноз суточных флюенсов РЭ ВРПЗ на 1-4 суток вперед
на сайте ЦОКМ НИИЯФ МГУ

http://swx.sinp.msu.ru/models/rb_electrons/index.php?gcm=1

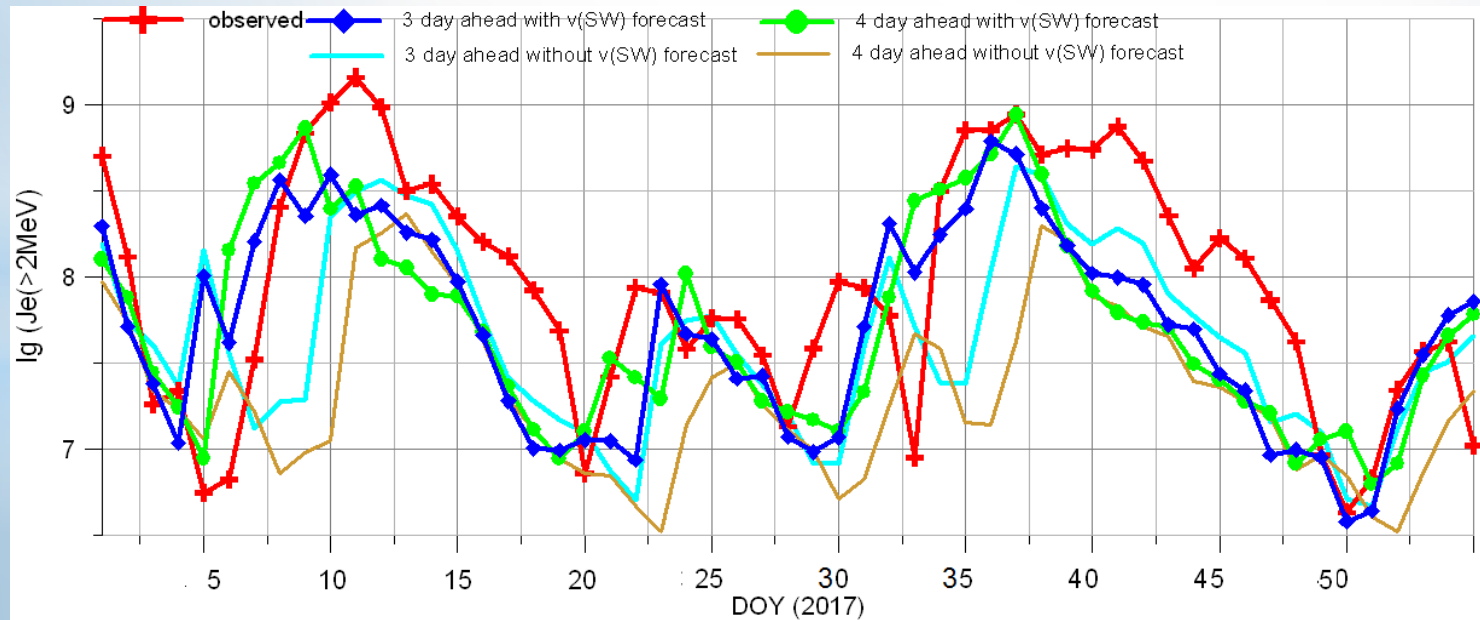
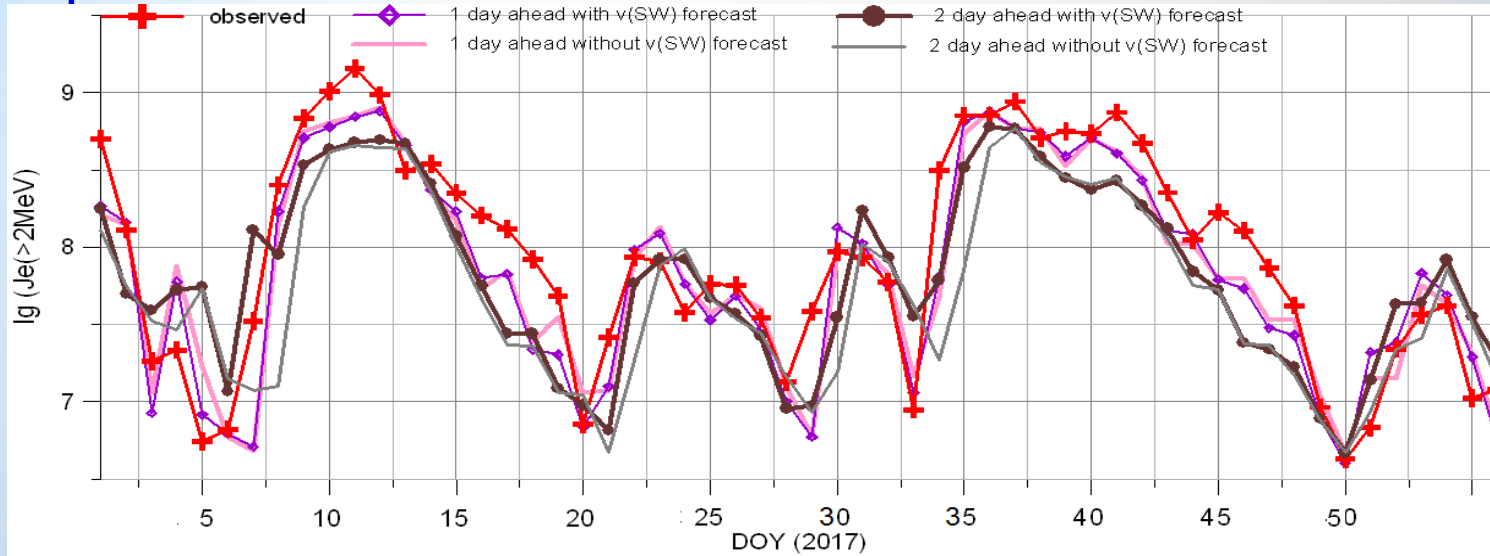
Учёт прогноза скорости солнечного ветра

Учёт прогноза скорости солнечного ветра по площадям корональных дыр на изображениях Солнца на 1-4 суток:

- Позволяет повысить качество прогноза суточных флюенсов РЭ ВРПЗ с горизонтом 3-4 суток
- Горизонт удовлетворительной точности прогноза составляет 3 суток.
- Сплошные линии – с учётом прогноза скорости СВ, пунктирные – без учёта:



Прогноз суточных флюенсов РЭ ВРПЗ

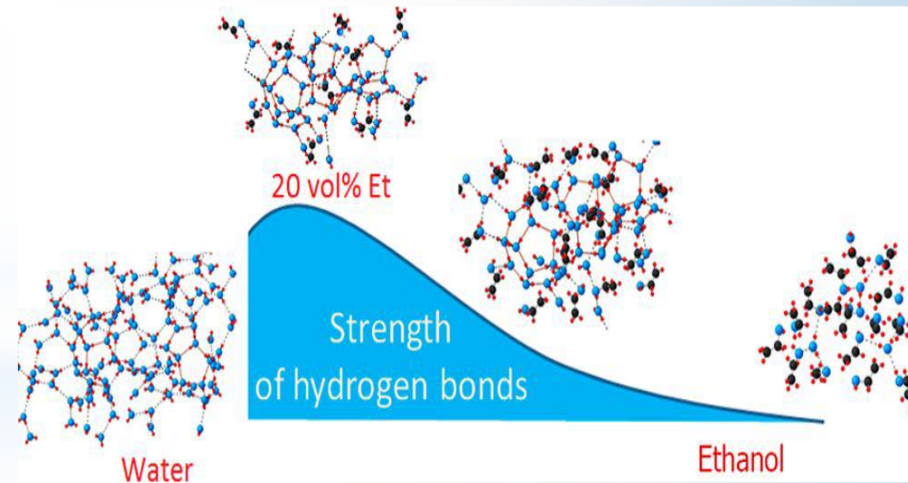
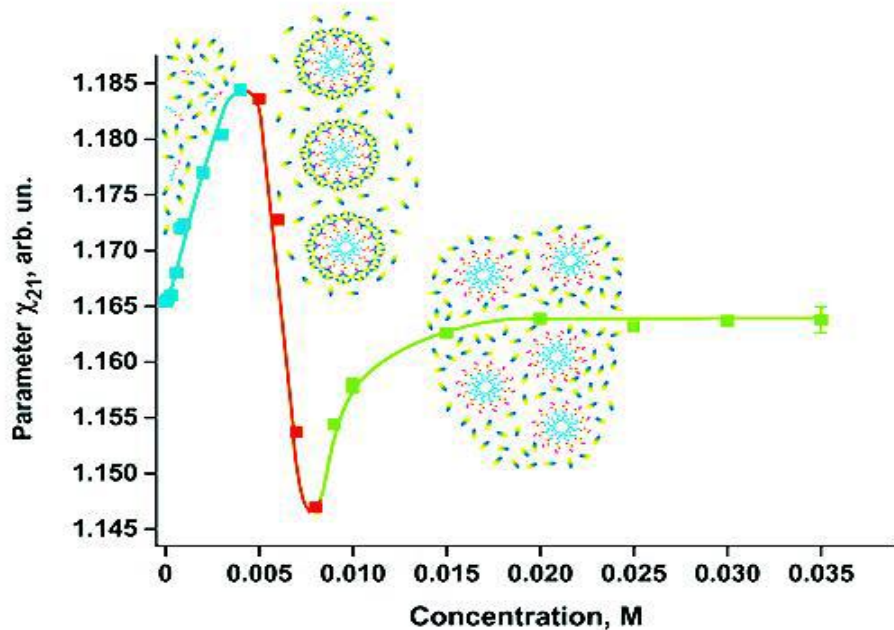


Сравнение качества прогнозов суточных флюенсов РЭ ВРПЗ, сделанных без учёта и с учётом прогноза скорости СВ по площадям корональных дыр на изображениях Солнца

Метод многомерного разрешения кривых

Применение метода **многомерного разрешения кривых** (MCR) для получения физически значимых результатов путём разложения спектра на составляющие.

- Использован для обнаружения **присутствия** в водных растворах SDS и этанола различных **типов клатратов** и определения их спектров
- Применение метода позволило **выделить** для каждого спектра вклады молекул с различной силой водородных связей

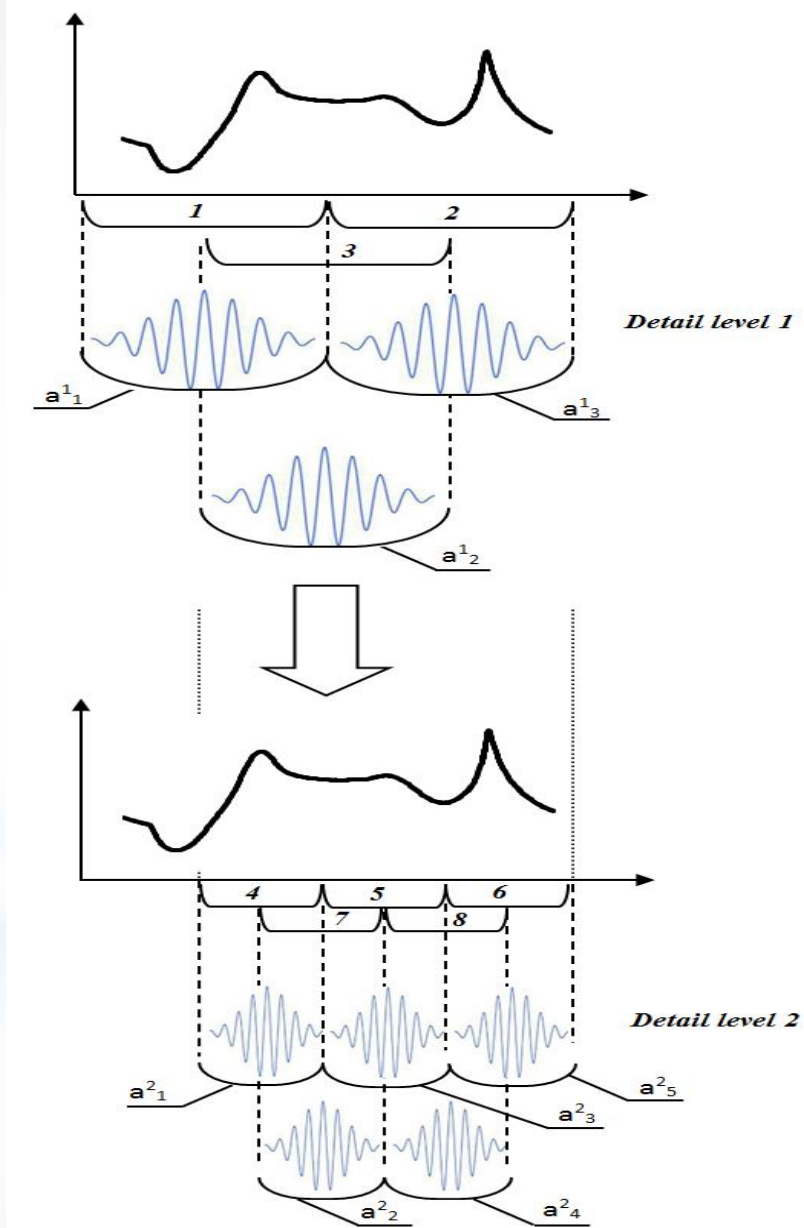


Преобладающие типы клатратов в зависимости от концентраций SDS (слева) и этанола (вверху)

Новый тип вейвлет-нейронной сети

Предложен **новый тип** вейвлет-нейронной сети – ВНС с **адаптивными окнами**, предназначенный прежде всего для **обработки сигналов**.

- Если не требуется обратное преобразование, значения коэффициентов **сдвига и масштаба** могут определяться в процессе обучения сети, а окна, соответствующие различным вейвлет-функциям, могут **перекрываться**.
- Положения окон и уровни вейвлетов определяются с помощью специальной итерационной процедуры



Обработка данных газовых сенсоров

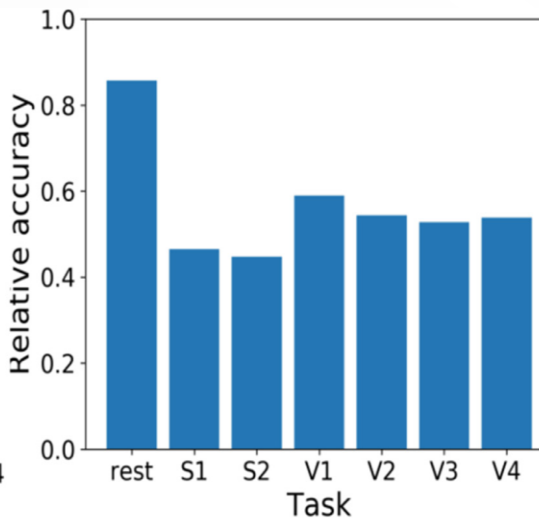
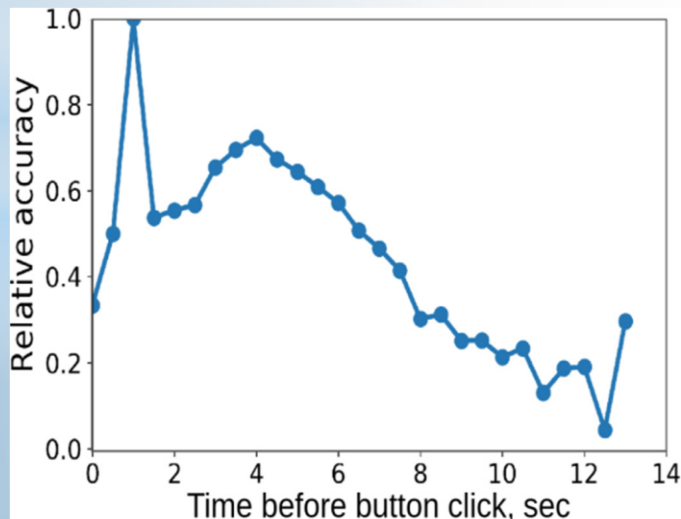
Продемонстрирована **высокая эффективность** нейросетевых методов при обработке данных **полупроводникового газового сенсора**, работающего в динамическом температурном режиме

- Исследовано **высокоселективное обнаружение** различных отдельных газов (CO , H_2 , CH_4 , C_3H_8 , NO , NO_2 , H_2S , SO_2) при низких концентрациях (0.01–667 ppm) в воздухе с помощью **одного** датчика из оксида металла на основе SnO_2 (MOX-sensor).
- Датчик работает в **динамическом** температурном режиме в сочетании с рядом **адаптивных алгоритмов** обработки сигналов.
- **Идентификация** отдельных газов **одним** датчиком, усредненная по всем газам и концентрациям, привела к ошибкам только в 13,2% случаев.
- Значительное **улучшение** точности **количественной** оценки (снижение относительной погрешности с 7 до 2,5%) было достигнуто за счет использования **массива** из 3 датчиков в сочетании с адаптивным алгоритмом обработки данных, по сравнению с использованием 1 датчика.
- Показано заметное **негативное** влияние **влажности** (Rh 30%, 25 °C).

Классификация многомерных данных фМРТ

Участники эксперимента должны были решать **задачи 6 типов**, требующих активизации **различных типов мышления**.

- 4D фМРТ-временные ряды с временным разрешением 0,5 с
- Классификация по типу задач; бинарная классификация (активность/покой) с помощью модели случайного леса на основе платформы Spark
- Для решения однотипных задач требуется некоторое ограниченное время, а если тратится больше времени, то это связано с тем, что мозг не сразу включается в работу над предложенной задачей, а еще какое-то время пребывает в состоянии покоя.

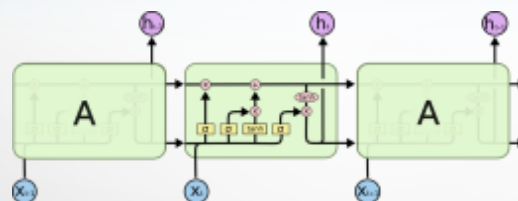
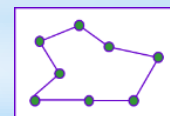
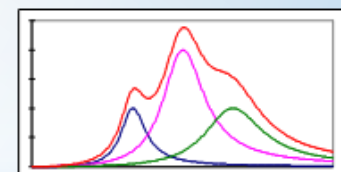
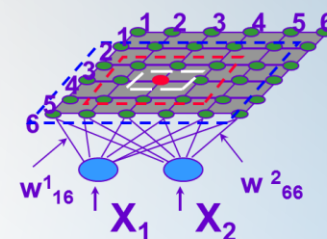
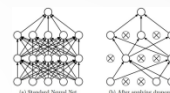


Зависимость точности классификации от времени, остающегося до нажатия кнопки испытуемым, и от типа решаемой задачи (rest – состояние покоя)

КПК по машинному обучению и нейросетям

Курс «Машинное обучение. Искусственные нейронные сети и генетические алгоритмы» существенно обновлён в 2020 году

- Искусственные **нейронные сети** и другие алгоритмы машинного обучения
- Основы **предобработки данных**
- Генетические** алгоритмы, генетическое программирование
- Нечёткая логика**, метод группового учёта аргументов
- Современный **инструментарий** для практической работы по анализу данных
- Практические** занятия с использованием языков **R** и **Python**
- Самостоятельная работа под руководством преподавателя
- Сайт курса: <http://kpk-nnga.sinp.msu.ru/>
- Для студентов, аспирантов и **сотрудников МГУ** участие **бесплатное**
- В этом году первое занятие **2 марта**, запись **открыта**





Спасибо за внимание!

Благодарю коллег и соратников:

- И.Н.Мягкова, О.Г.Баринов, В.В.Калегаев – ЛКФИ ОКН
- С.А.Буриков, Т.А.Доленко, О.С.Сарманова – физический факультет МГУ
- И.Е.Оборнев, Е.А.Родионов, Е.А.Оборнев, М.И.Шимелевич – РГГРУ
- В.В.Кривецкий, М.В.Румянцева, **А.М.Гаськов** – химический факультет МГУ
- В.Л.Ушаков, В.А.Орлов – НИЦ «Курчатовский институт»
- А.В.Самсонович, Д.В.Тихомирова – ИИКС НИЯУ МИФИ