



ВОССТАНОВЛЕНИЕ ПОЛОЖЕНИЯ ИСТОЧНИКА ИЗЛУЧЕНИЯ В СЦИНТИЛЛЯЦИОННОМ ДЕТЕКТОРЕ ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Выполнил студент М25-112

Козлов А.

Научный руководитель:

Попов Д.В.

В экспериментальные задачи современной физики элементарных частиц входит восстановление точки (или вершины) взаимодействия. В случае с антинейтрино реакторного происхождения и ОБР - точки, в которой родился позитрон и нейтрон. Данная задача является одной из основополагающих для нейтринного сцинтилляционного детектора на основе пластиковых сцинтилляторов, разрабатываемого в МИФИ для реакторного мониторинга

С решением данной задачи могут помочь современные методы машинного обучения, активно применяемые в физике последние несколько десятков лет. Машинные алгоритмы эффективно справляются с восстановлением координат различных событий при наличии информации о геометрии детектора и иных вспомогательных данных.

В связи с этим разработка собственного алгоритма по восстановлению положения источника излучения становится актуальной и востребованной задачей в рамках разработки нового нейтринного детектора.

Цель работы

Реализация метода локализации источника излучения методами машинного обучения на основе данных моделирования

Задачи:

- Изучить данные, полученные при моделировании в Geant4
- Провести обработку данных, конструирование признаков (feature engineering)
- Провести исследование эффективности нескольких популярных методов машинного обучения в рамках задачи восстановления координат источника
- Найти наиболее оптимальные гиперпараметры моделей и применить их в решении задачи по локализации точечного источника

Данные моделирования

Данные для обработки были получены при помощи моделирования в среде GEANT4 для заданной геометрии детектора

Прототип детектора представляет собой 9 пластиковых сцинтилляторов размерами 50 x 50 x 700 мм, изготовленных из полистирола с добавками п-терфенила и ПОПОП

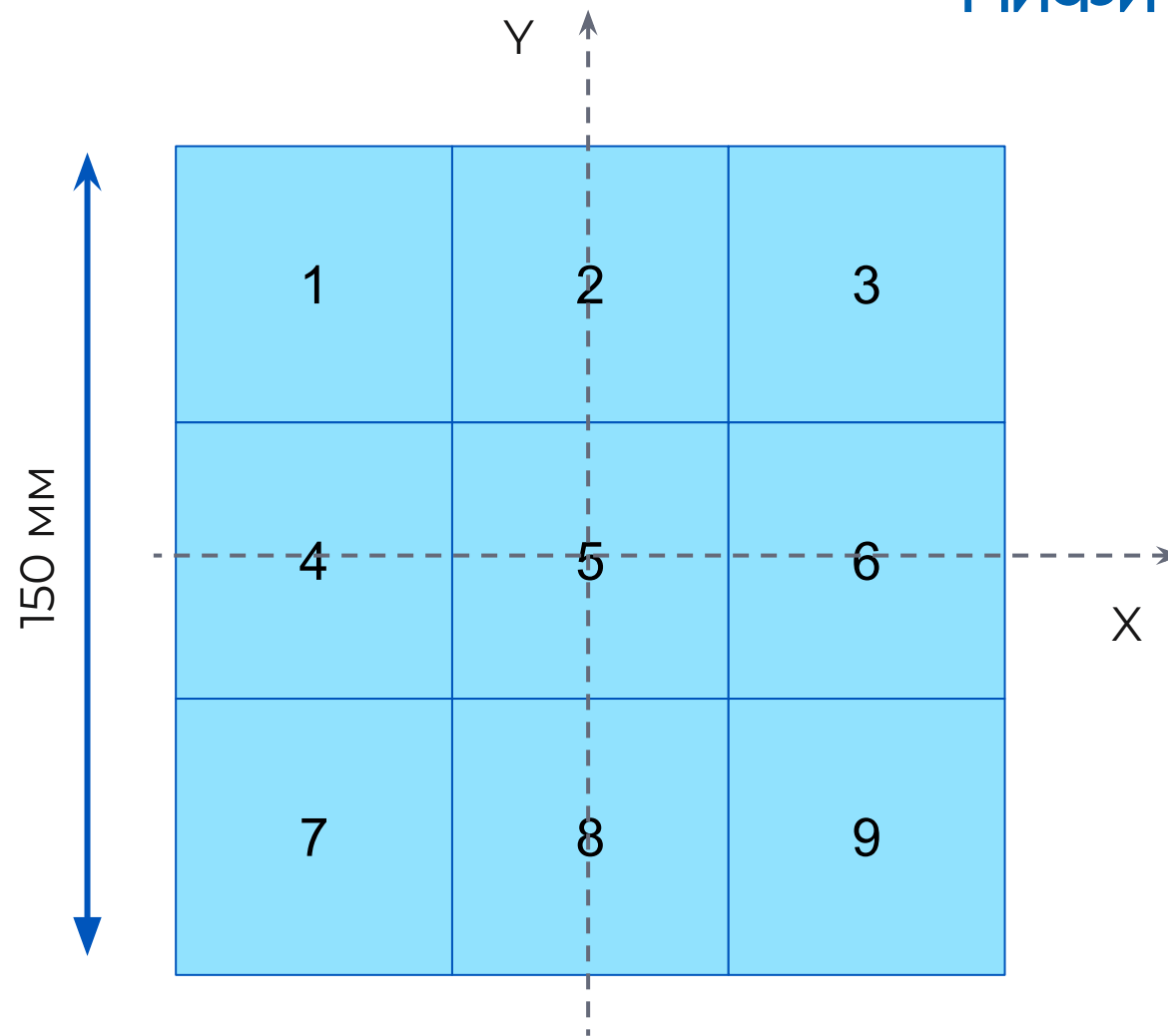


Рис. 1 Схема геометрии детектора

Данные моделирования

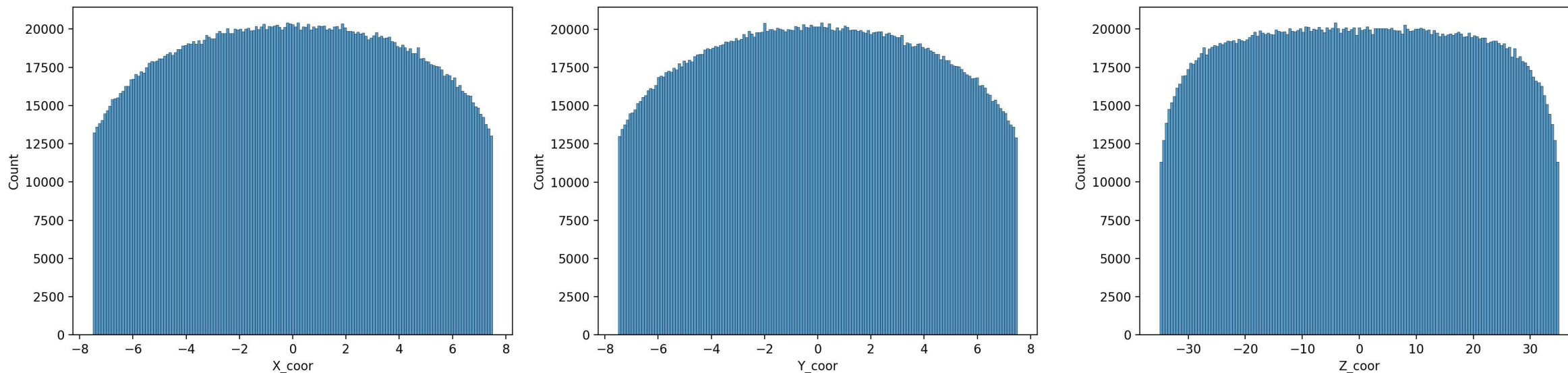


Рис. 2 Распределение координат
розыгрыша гамма-квантов в объёме

В экспериментальном объёме симуляции разыгрывались гамма-кванты с равномерным распределением по трём координатам: X, Y, Z

Энергия частиц разыгрывалась в диапазоне от 0 до 10 МэВ

Всего было разыграно 10 миллионов событий, после обработки осталось 2 812 362 события

Данные для обучения модели

Данные из исходного моделирования:

1. Энерговыделение в отдельных сцинтилляторах, МэВ
2. Изначальная энергия частицы, МэВ
3. Координаты розыгрыша частицы: X, Y, Z, см

	Edep_1	Edep_2	Edep_3	Edep_4	Edep_5	Edep_6	Edep_7	Edep_8	Edep_9	E_init	X_coor	Y_coor	Z_coor
0	0.0	2.888909	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	3.462055	1.053565	-2.606599	-7.634737
1	0.0	0.055129	0.0	5.941811	0.06118	0.000000	0.0	0.145499	0.02678	7.227068	-5.611222	-4.459059	14.563748
2	0.0	0.000000	0.0	0.000000	0.000000	0.487303	0.0	0.000000	0.000000	1.607376	4.716731	-1.495950	-26.777709

Рис. 3 Типичное представление данных, полученных после моделирования

Идея обучения: по имеющимся данным об энергии и энерговыделении восстановить координаты X, Y, Z с максимально возможной точностью

Конструирование дополнительных признаков

На определении координат X,Y,Z может положительно сказаться прямое указание модели на геометрию нашей установки – для этого был введён ряд дополнительных искусственных признаков:

- Center of Gravity (CoG, центр тяжести) для осей X и Y

$$\text{CoG}_x = \frac{\sum E_{dep_i} * X_i}{\sum E_{dep_i}}$$

X_i – координата центра i-огосцинтиллятора

E_{dep_i} – энергосвыделение в i-ом сцинтилляторе

- $\text{grad}(x)$ и $\text{grad}(y)$ – разность энергий между противоположными краями детектора, нормированная на полную энергию

$$\text{grad}(x) = \frac{E_1 + E_4 + E_7 - E_3 - E_6 - E_9}{E_{total}}$$

$$\text{grad}(y) = \frac{E_1 + E_2 + E_3 - E_7 - E_8 - E_9}{E_{total}}$$

- Асимметрия центра – отношение E_{dep5} / E_{total}
дополнительное указание для модели на геометрию события
- Доли энергии в каждом отдельном детекторе, нормированные на E_{total}
- Логарифмы энергосвыделения для учёта краевых эффектов
- Разброс распределения энергии: $\sigma(E_{dep}) / E_{total}$

Модель на основе градиентного бустинга

Главная концепция – последовательное обучение ряда слабых предсказателей на основе результатов работы каждого предыдущего предсказателя

В качестве моделей градиентного бустинга применялась модель **HistGradientBoostingRegressor** из библиотеки `scikitlearn`, обеспечивающая достаточную гибкость настройки и хорошую интерпретируемость результатов

Gradient Boost

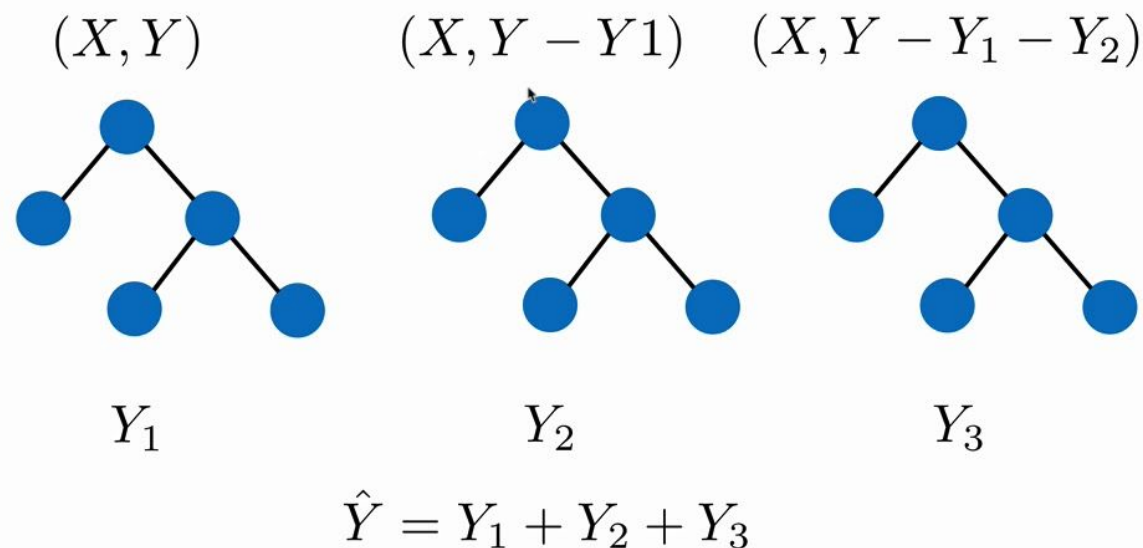


Рис. 4 Схематическое объяснение работы алгоритма градиентного бустинга

Подбор гиперпараметров модели

Для модели HistGradientBoostingRegressor при помощи файнтюнинга были определены следующие оптимальные гиперпараметры:

- `n_estimators`: 50 – число слабых предсказателей
- `max_depth`: 6 – максимальная глубина слабого дерева
- `learning_rate`: 0.05 – шаг обучения
- `tree_method`: hist – метод для ускорения работы с табличными данными
- `early_stopping_rounds`: 50 – число итераций для прекращения обучения при отсутствии результата

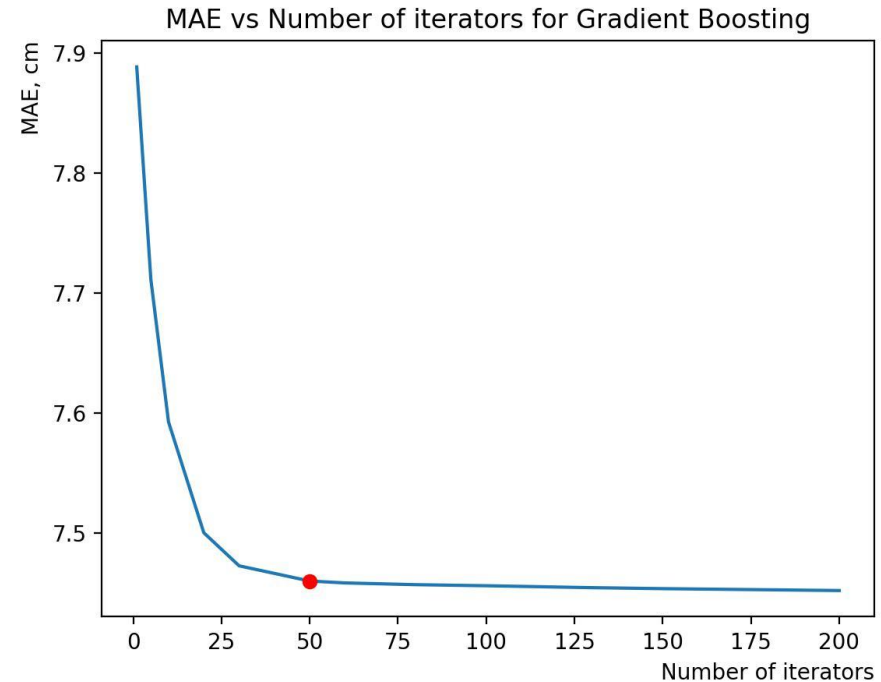


Рис. 5 График зависимости модуля средней ошибки предсказания координат от числа предсказателей в модели

Признак	Важность признака для определения MAE
CoG	1.13
E_{init}	0.03
asymetry center	0.0012
$E_{dep_4}ratio$	0.007
$E_{dep_6}ratio$	0.003

Таблиц 1 Признаки и их важности для модели градиентного бустинга

При помощи метода перестановок (перемешивания признаков) была оценена важность признаков для модели на основе градиентного бустинга

Таким образом, наиболее содержательным и полезным признаком оказался центр масс (CoG_x), на основе которого модель делает наиболее точные предсказания

Количество признаков для обучения и предсказания было уменьшено до 5 наиболее значимых, что позволит избежать переобучения и ускорит процесс обучения модели

Результаты работы градиентного бустинга

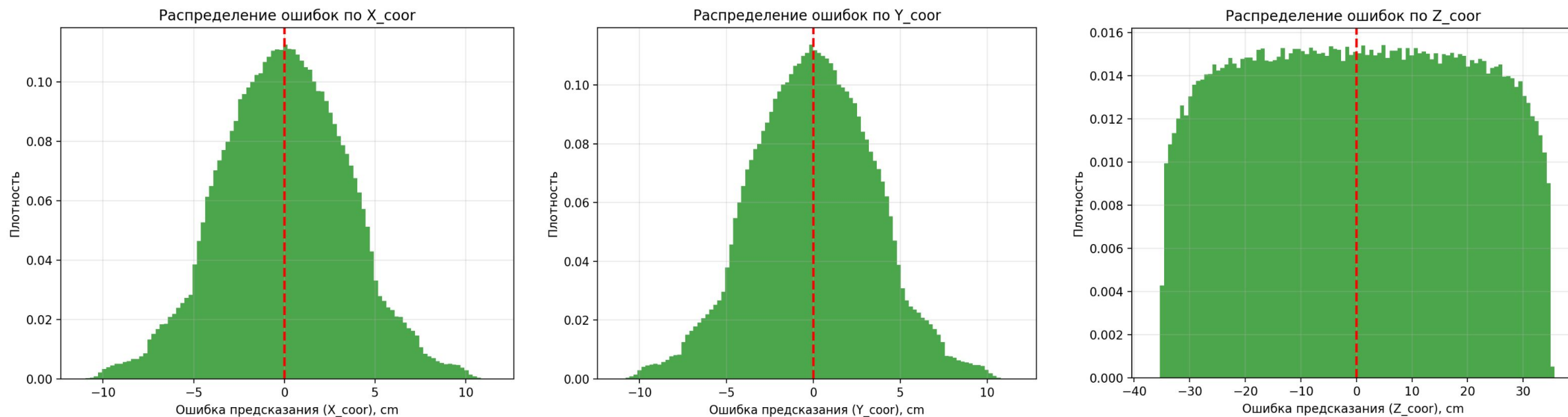


Рис. 7 Распределение ошибок по координатным осям X, Y, Z на основе предсказаний, сделанных моделью

MAE = 2.8 см
RMSE = 3.8 см

MAE = 2.8 см
RMSE = 3.5 см

MAE = 16.8 см
RMSE = 19.5 см

Равномерное распределение ошибки по оси Z – следствие физического ограничения на возможность различать события вдоль сцинтиллятора

Модель KNN

Алгоритм К ближайших соседей (K Nearest Neighbors) является классическим алгоритмом машинного обучения. В задаче регрессии объекту присваиваются свойства на основе значения аналогичных свойств у его N ближайших соседей в многомерном пространстве признаков

Ключевая особенность – модель не учится на части выборки, а запоминает её целиком, что положительно влияет на учёт потенциальных аномалий и выходов за заложенные в симуляции пределы величин

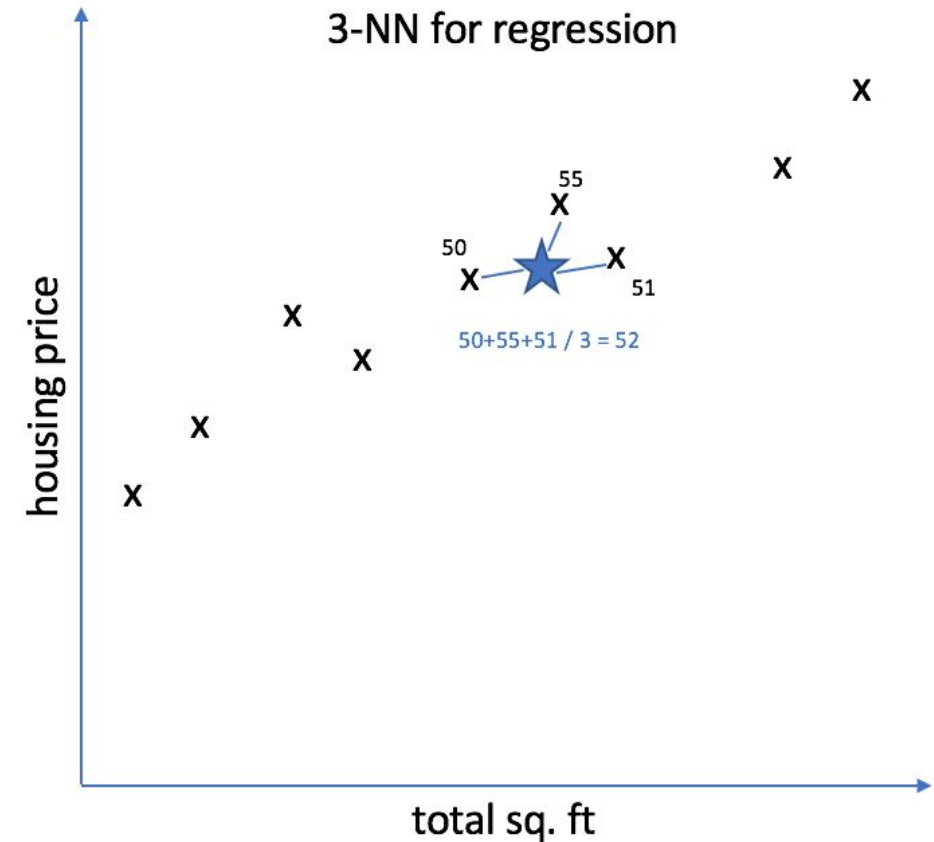


Рис. 8 Схема работы метода KNN с тремя соседями для задачи регрессии

Результаты работы KNN-модели

N соседей – 350

	KNN-модель	Градиентный бустинг
MAE X, см	2.8	2.8
MAE Y, см	2.8	2.8
MAE Z, см	16.8	16.8
RMSE X, см	3.5	3.8
RMSE Y, см	3.5	3.5
RMSE Z, см	19.5	19.5

Таблица 2 – Сравнение метрик для моделей KNN и градиентного бустинга

Более простая модель KNN не уступает в решении данной задачи регрессии модели градиентного бустинга

Проверка с моделированием точечного источника

В качестве проверки эффективности рассмотренных моделей в задачах реконструкции был проведён эксперимент с моделированием точечного источника гамма-квантов

Точечную модель источника цезий-137 разместили на середине второго сцинтиллятора. Было получено 3 628 721 полезных событий в объёме детектора

Для этих событий модель градиентного бустинга восстанавливала координату источника излучения

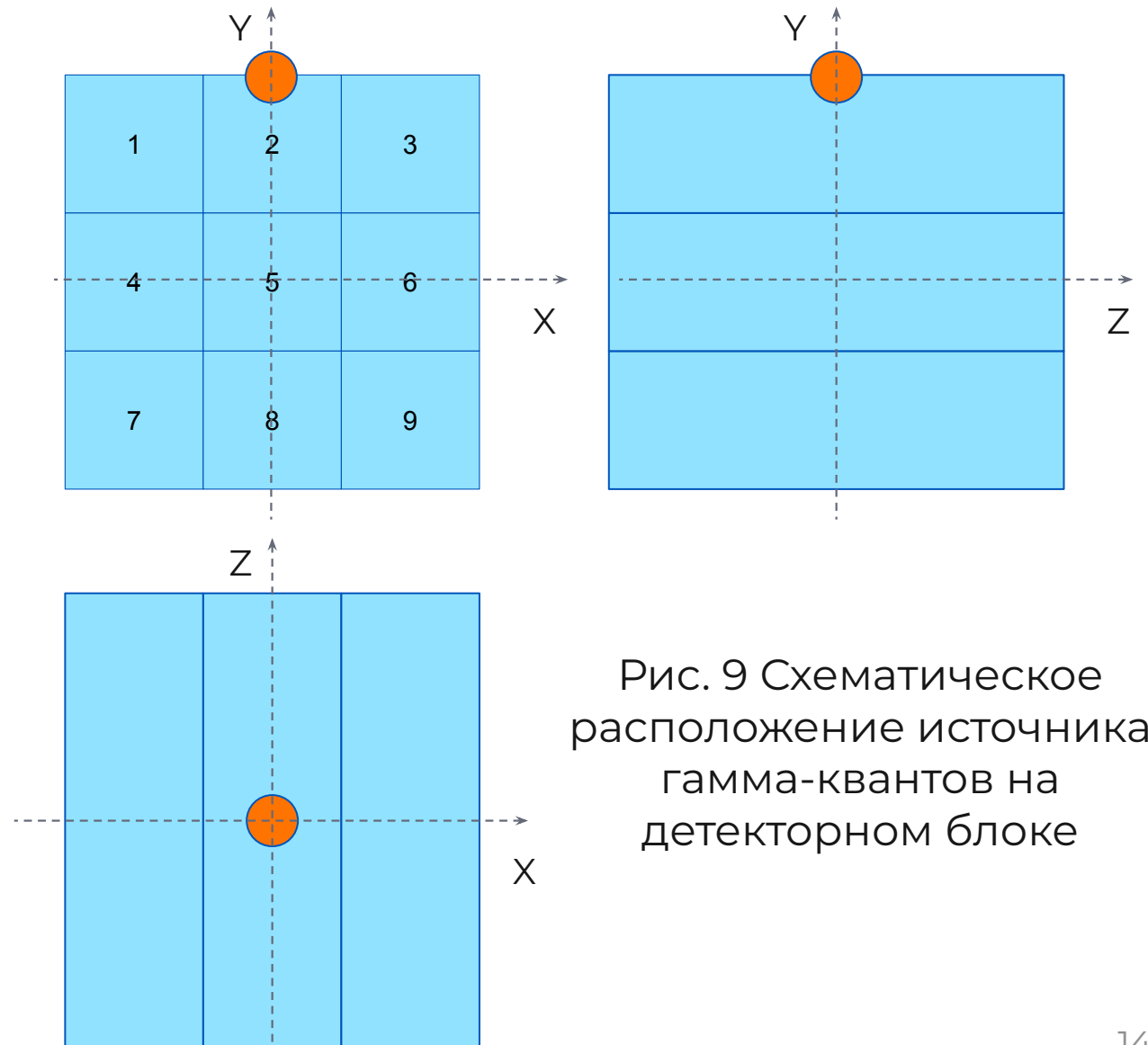


Рис. 9 Схематическое расположение источника гамма-квантов на детекторном блоке

Проверка с моделированием точечного источника

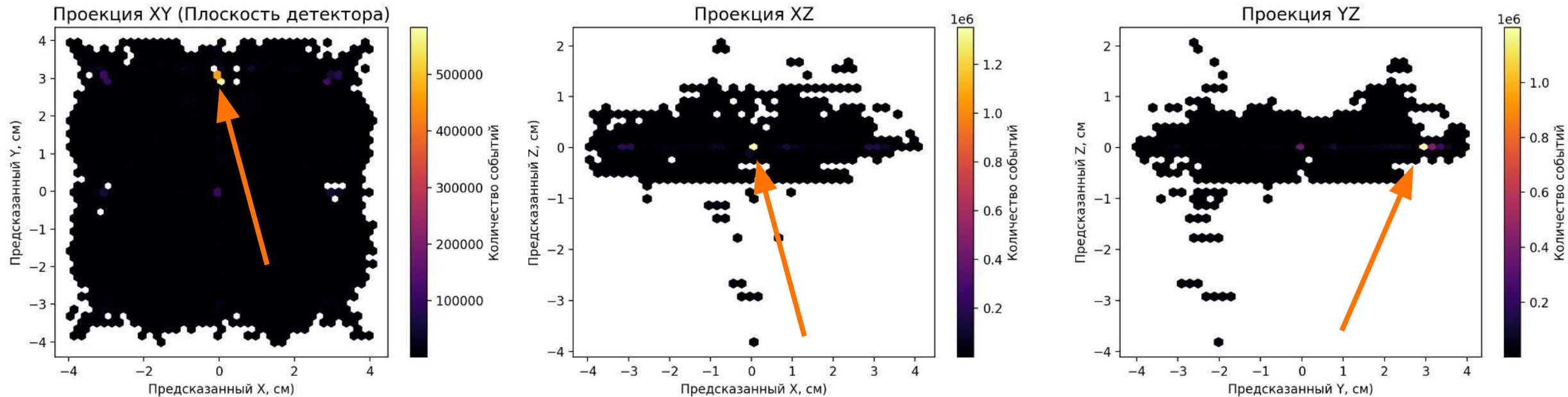


Рис. 10 Тепловая карта реконструкции координаты положения источника излучения моделью на основе градиентного бустинга

- MAE X = 1.3 см
- RMSE X = 1.9 см
- Отклонение средней оценки X = 0.05 см

- MAE Y = 3.12 см
- RMSE Y = 3.6 см
- Отклонение средней оценки Y = 1.8 см

- MAE Z = 35 см
- RMSE Z = 35 см
- Отклонение средней оценки Z = 35 см

При помощи данных, полученных из моделирования в среде GEANT4, были решены следующие задачи:

- Построение моделей для предсказания положения источника излучения на основе метода градиентного бустинга и метода К ближайших соседей (KNN)
- Оценка точности предсказаний модели для равномерного розыгрыша частиц в экспериментальном объёме. Обе модели продемонстрировали одинаковую эффективность (MAE X, Y = 2.8 см, RMSE X, Y = 3.5)
- Проверка работы модели в эксперименте по локализации точечного источника гамма-квантов (MAE X = 1.3 см, RMSE X = 1.9 см; MAE Y = 3.12, RMSE Y = 3.6)

Следующие шаги развития эксперимента:

- Добавление в моделирование новых частиц с уникальной физикой (нейтроны, позитроны)
- Работа с экспериментальными данными, полученными с установки
- Разработка ансамбля моделей для классификации частиц и обнаружения их источника

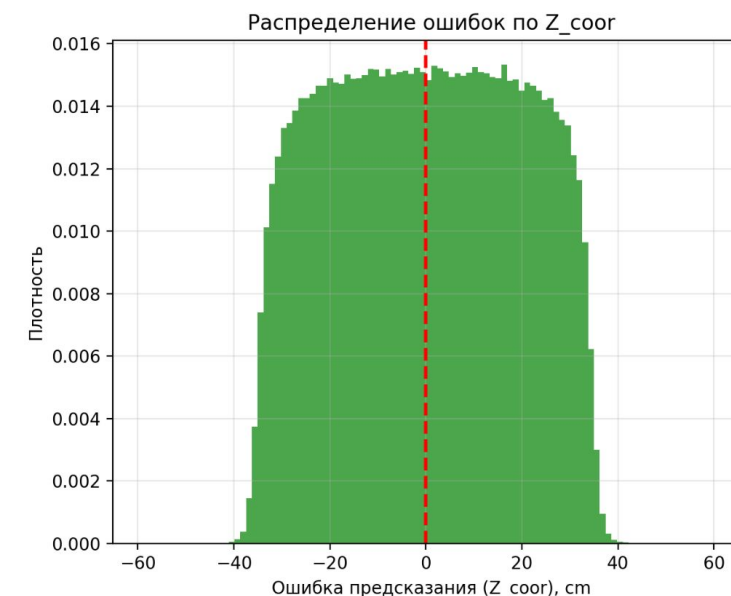
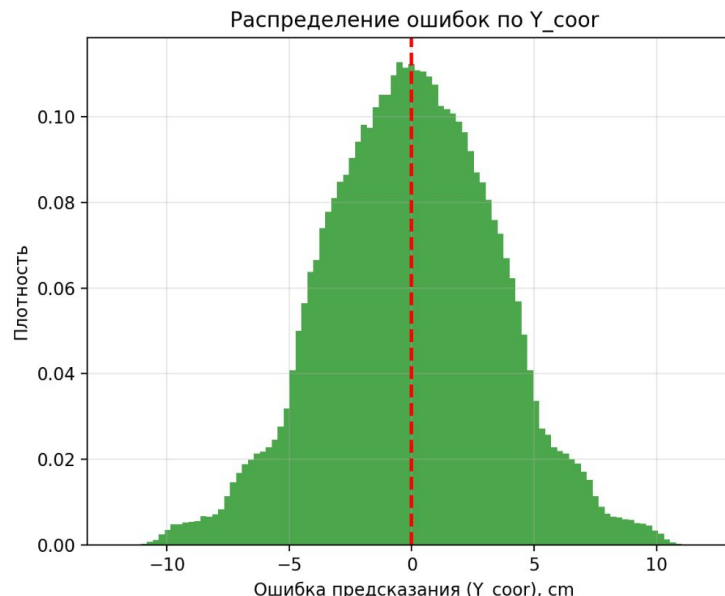
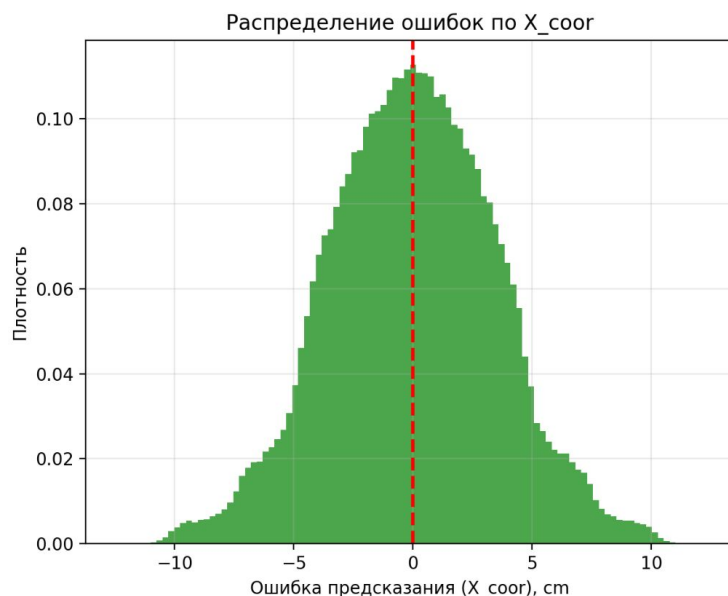


Спасибо за внимание!

Козлов А.А.

andcauselove@yandex.ru

Дополнительный слайд №1



Распределение ошибок по координатным осям X, Y, Z на основе предсказаний, сделанных моделью KNN

MAE = 2.8 см
RMSE = 3.5 см

MAE = 2.8 см
RMSE = 3.5 см

MAE = 16.8 см
RMSE = 19.5 см

Равномерное распределение ошибки по оси Z – следствие физического ограничения на возможность различать события вдоль сцинтиллятора

Дополнительный слайд №2

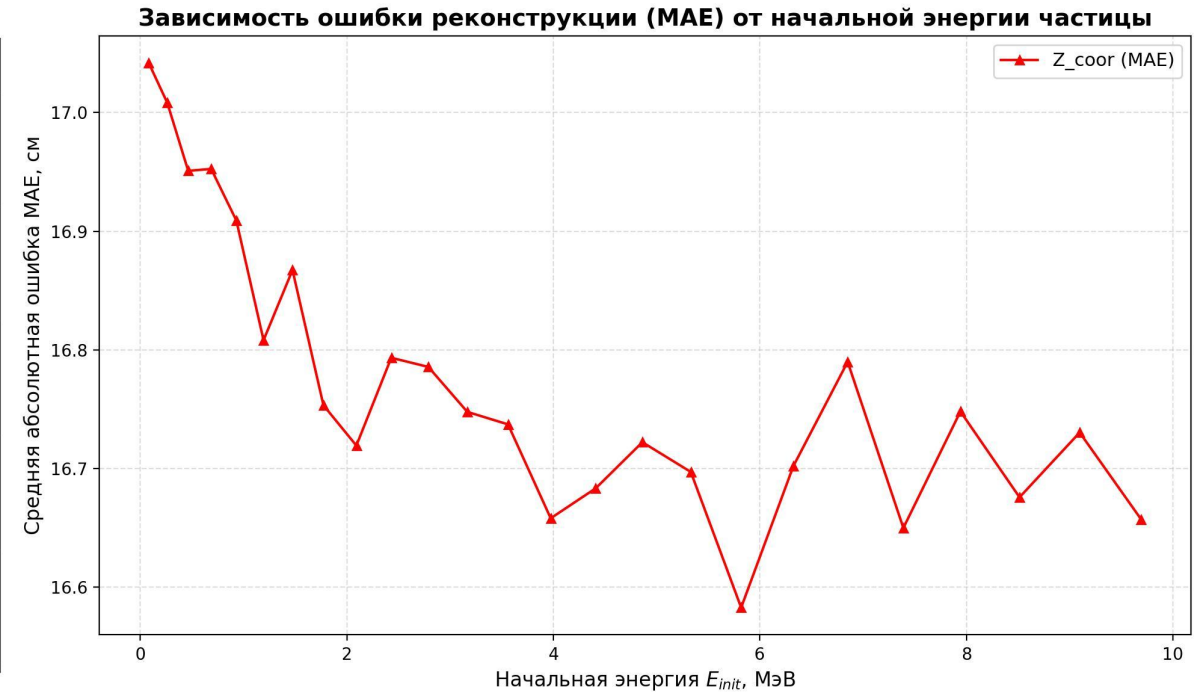
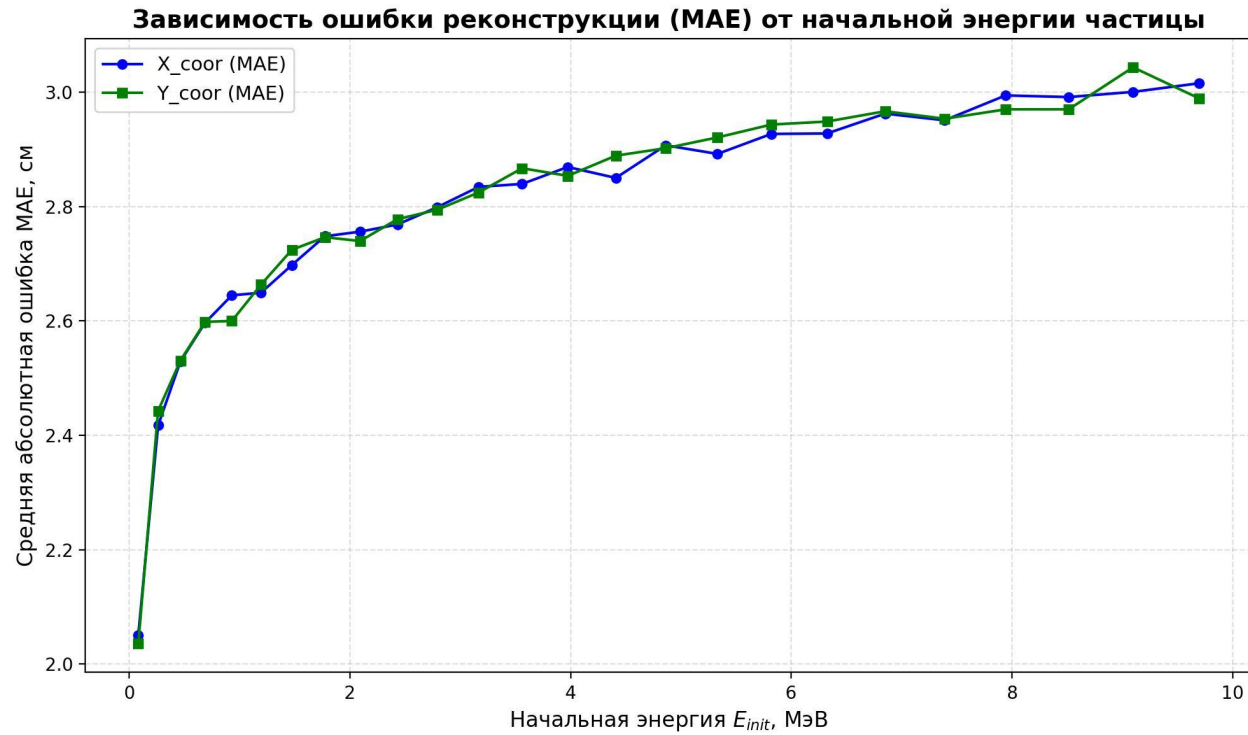


График зависимости ошибки предсказания MAE от начальной энергии частицы для модели GBT

Дополнительный слайд №3

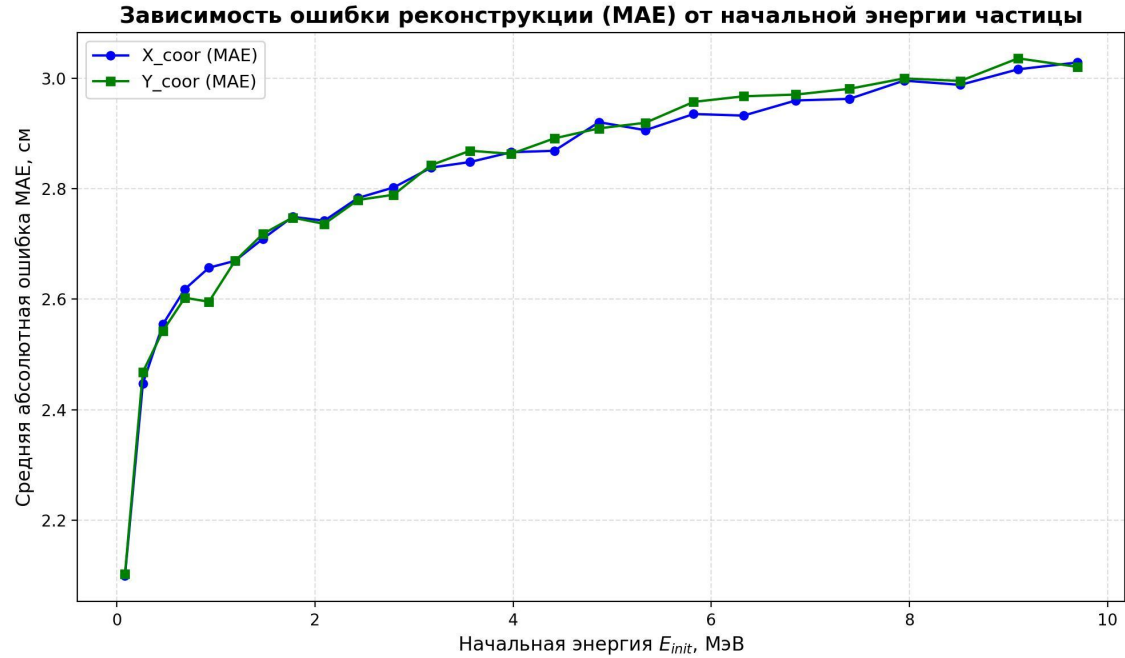
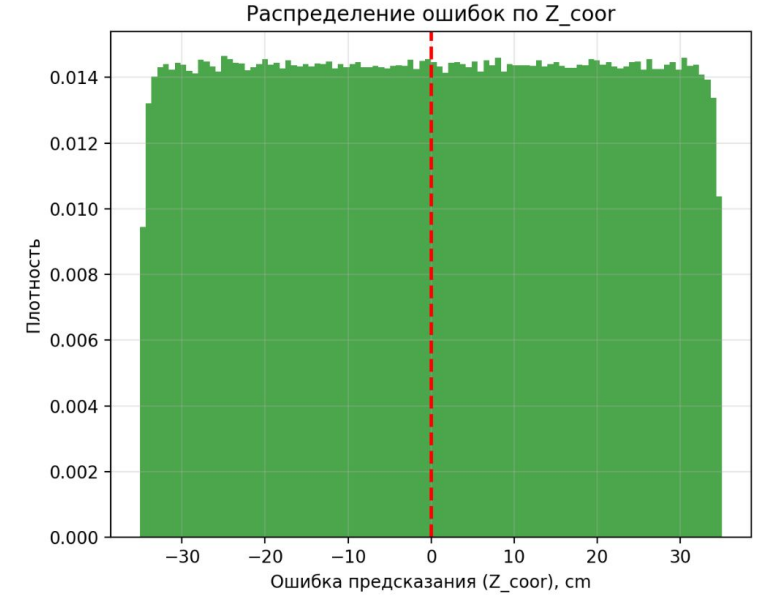
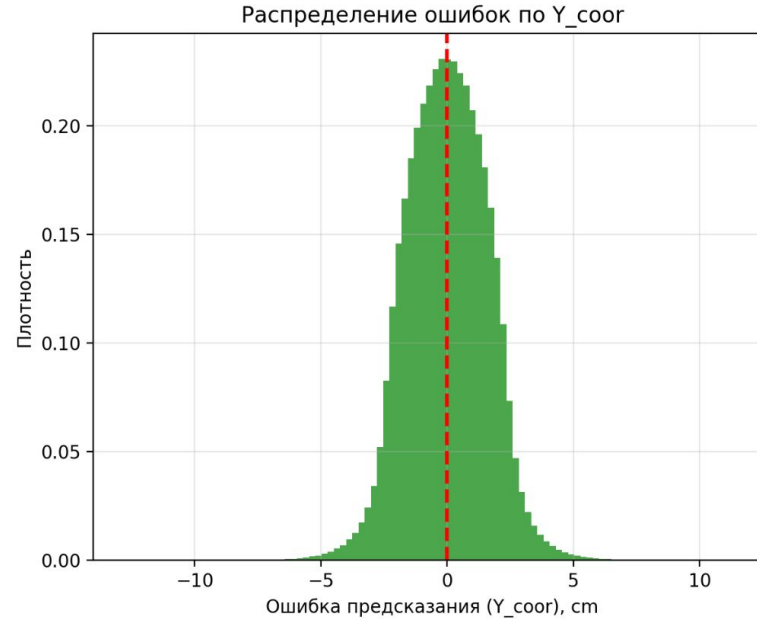
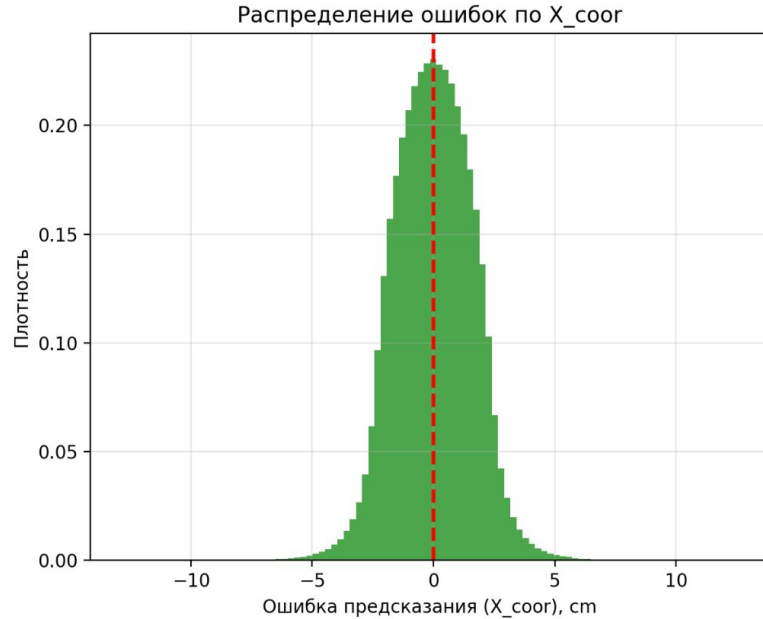


График зависимости ошибки предсказания MAE от начальной энергии частицы для модели KNN

Дополнительный слайд №4



Распределение ошибок по координатным осям X, Y, Z на основе предсказаний для нейтронов, сделанных моделью GBT

MAE = 1.3 см
RMSE = 1.6 см

MAE = 1.3 см
RMSE = 1.6 см

MAE = 17.4 см
RMSE = 20.1 см