



Генерация показаний наемных детекторов и поиск аномалий в данных с помощью нейронных сетей

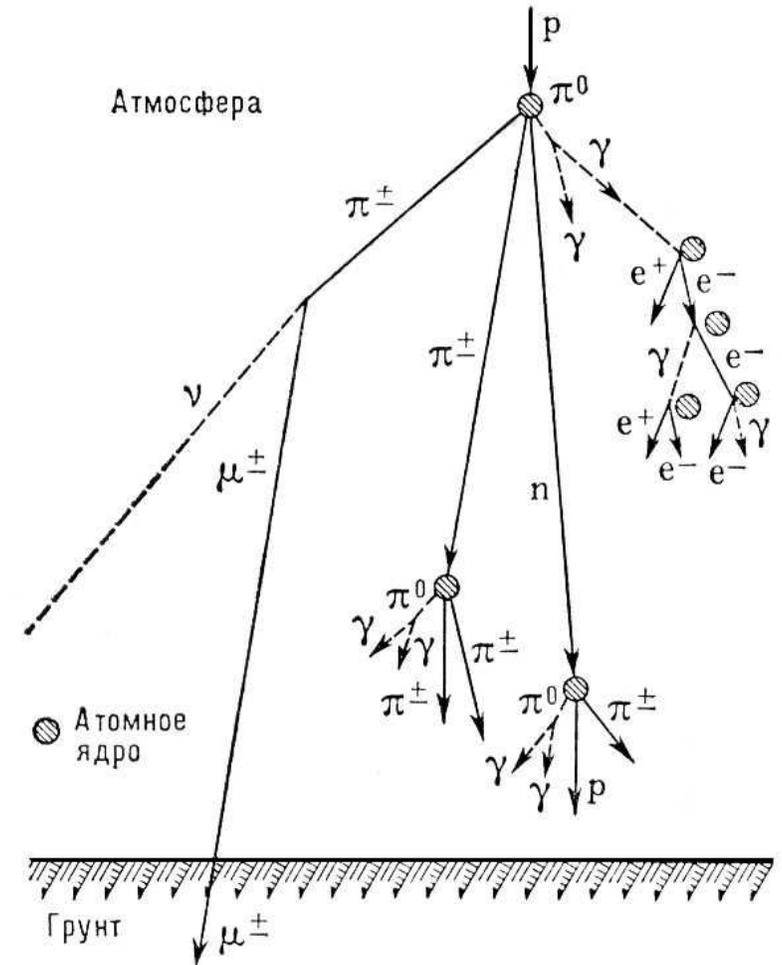
Фитагдинов Роберт
ИЯИ РАН, МФТИ
DLCP 2024

Космические лучи и широкие атмосферные ливни

Космические лучи (КЛ)- заряженные частицы, прилетающие к нам от далеких космических объектов.

Широкий атмосферный ливень(ШАЛ) - каскад вторичных частиц, вызванный КЛ из-за его взаимодействия с атмосферой Земли.

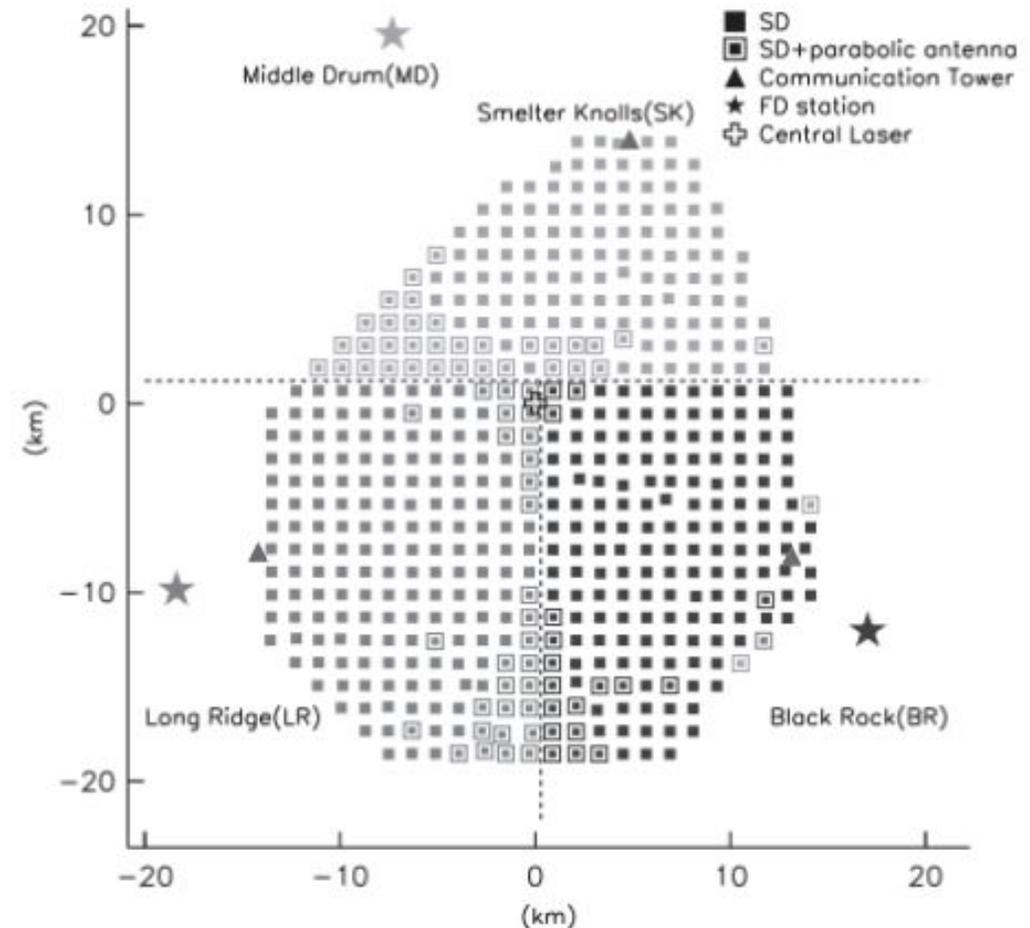
Из-за невозможности прямого детектирования КЛ и стохастичности распространения ШАЛ, исследование КЛ является тяжелой задачей.



Эксперимент Telescope Array

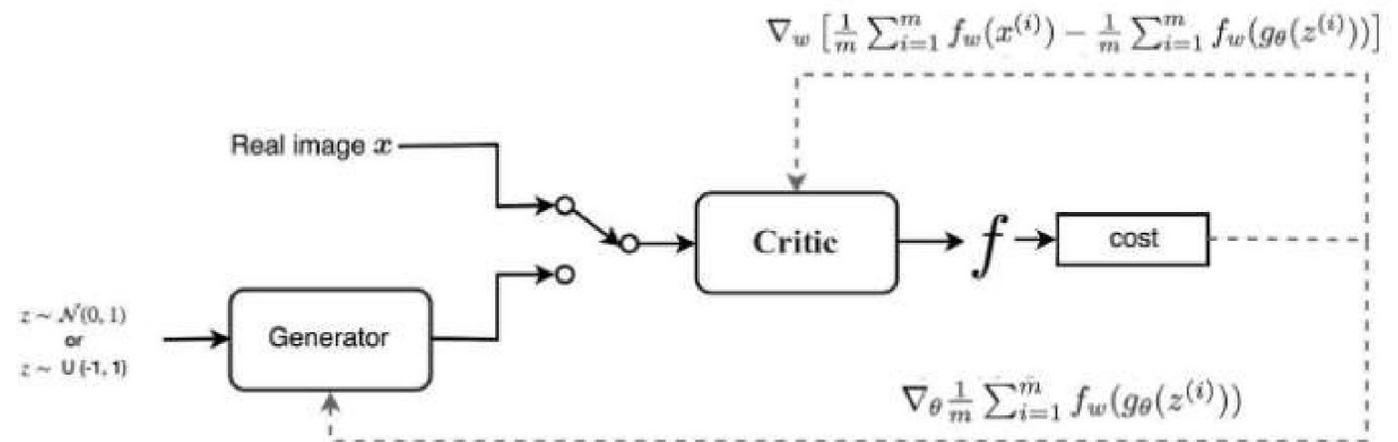
Эксперимент состоит из 507
сцинтилляционного детектора и
3 флуоресцентных детектора.

Поверхностные детекторы
равномерно распределены по
решетке 762 км^2 с расстоянием
1,2 км между каждым
детектором.



Генеративно-состязательная сеть вассерштейна с градиентным штрафом (WGAN-GP)

- WGAN состоит из двух сетей: генератор и критик. Суть обучения GAN состоит в конкуренции этих двух моделей. Генератор производит реалистичные данные, критик пытается отличить реальные от сгенерированных.

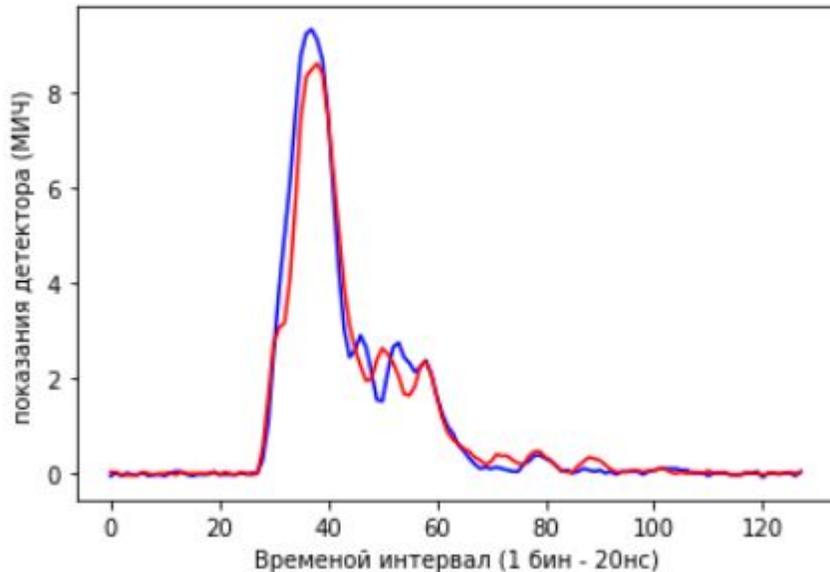


$$L = \underbrace{\mathbb{E}_{\hat{x} \sim P_g} [D(\hat{x})] - \mathbb{E}_{x \sim P_r} [D(x)]}_{\text{Original critic loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x} \sim P_m} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}_{\text{Our gradient penalty}}.$$

Задачи

1. Генерация показаний и поиск аномальных событий на данных полученных с наиболее активного детектора в каждом событии
2. Генерация данных нескольких детекторов в каждом событии

Данные наиболее активного детектора



- Данные имеют размер (128, 2)

2 - показания двух каналов детектора

128- количество бинов, на которые разбивается интервал в 2560 нс.

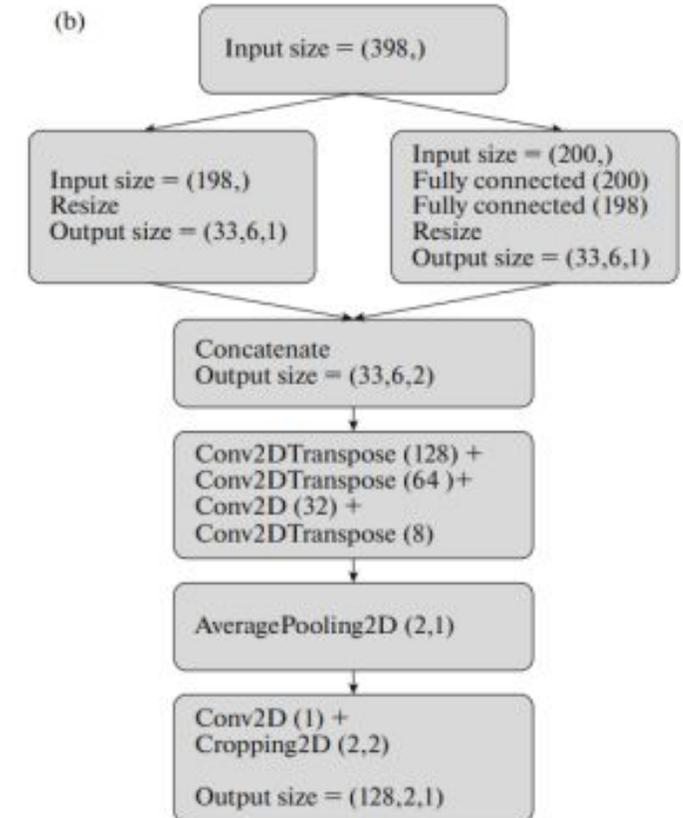
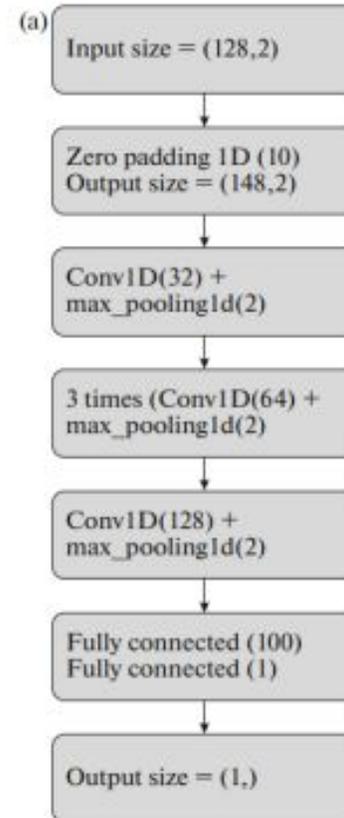
Всего мы располагали около 500к данных, полученных с помощью метода МК, использующего в своей основе различные модели высокоэнергетических взаимодействий.

Модели

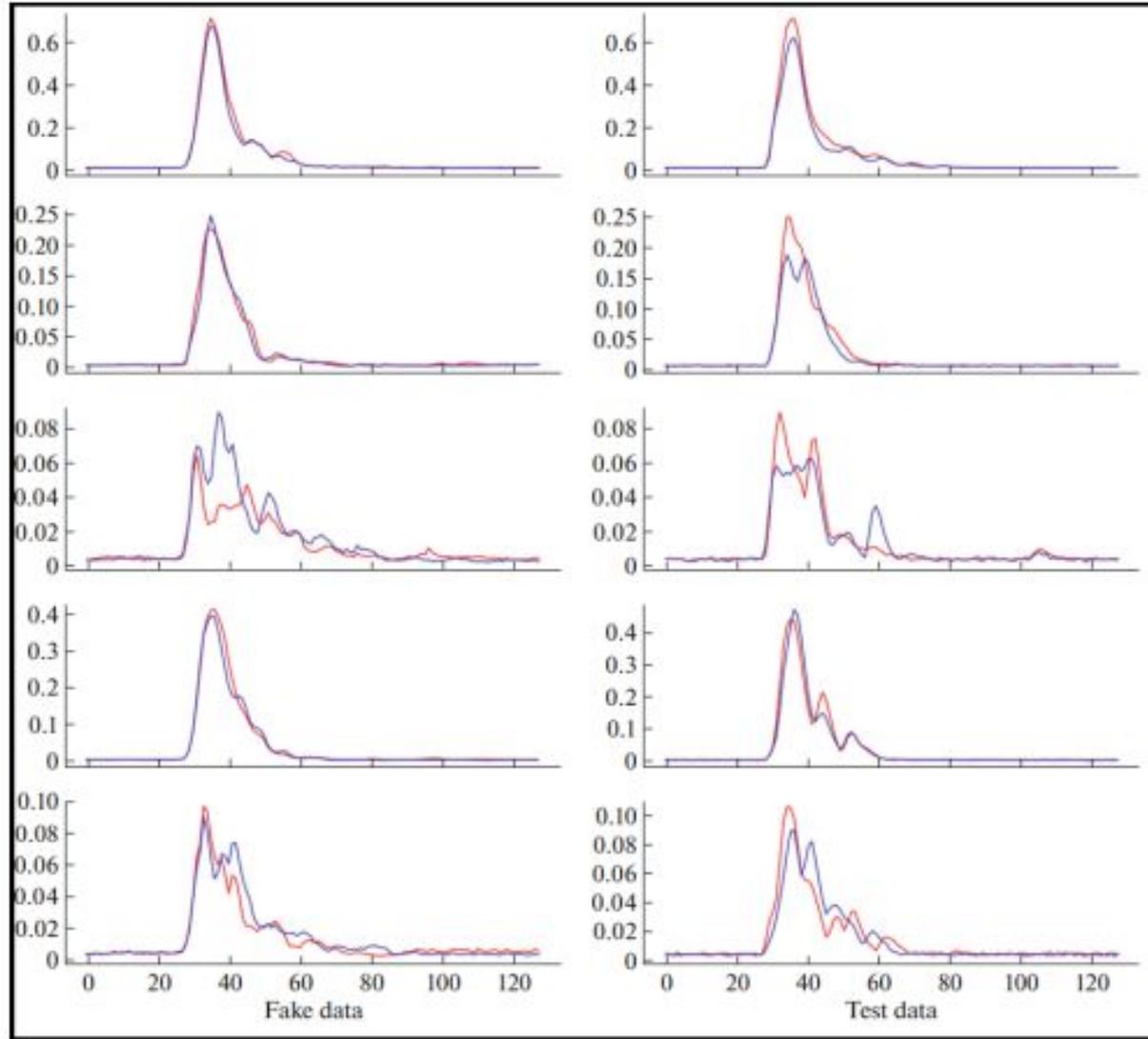
Схема генеративно-состязательной сети для задачи генерации наиболее активного детектора.

а) - схема дискриминатора (150к обучаемых параметров);

б) - схема генератора (180к обучаемых параметров);



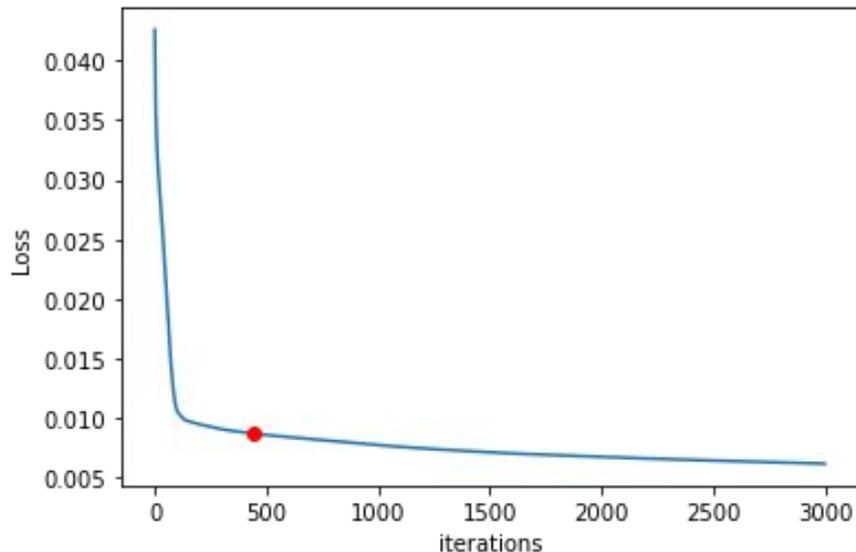
Результаты генерации



Алгоритм поиска аномалий

$$L = L_{pixel}(real, fake) + w * L_{disc}(real, fake)$$

- L_{pixel} Сумма по всем элементам разницы реальных и сгенерированных данных
- L_{disc} разница показаний дискриминатора.
- w - вес ошибки 2. В нашем случае 0,001



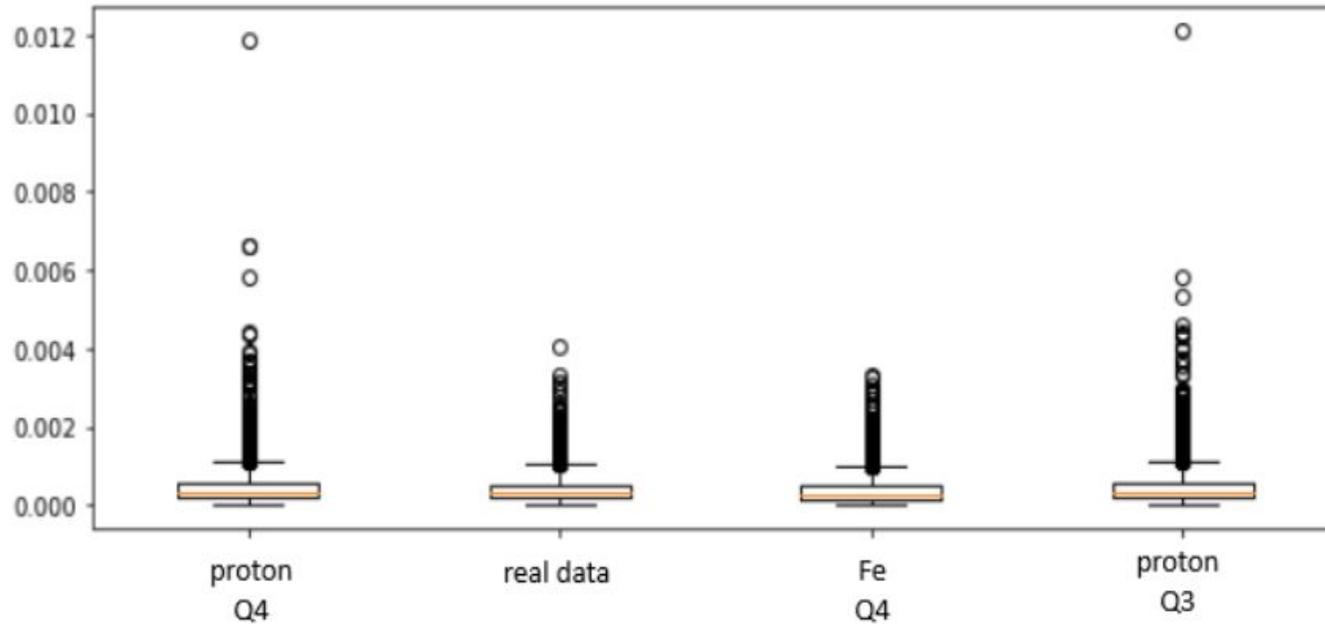
1. Создается случайный шум и подается на вход генератору.

2. Считается функция ошибок от выходных данных генератора и данных переданных алгоритму.

3. С помощью градиентного спуска и автодифференцирования, оптимизируется шум, подаваемый на вход генератору.

4. Пункты 2-3 повторяются некоторое количество раз (количество шагов оптимизации или условия ранней остановки являются гиперпараметрами алгоритма)

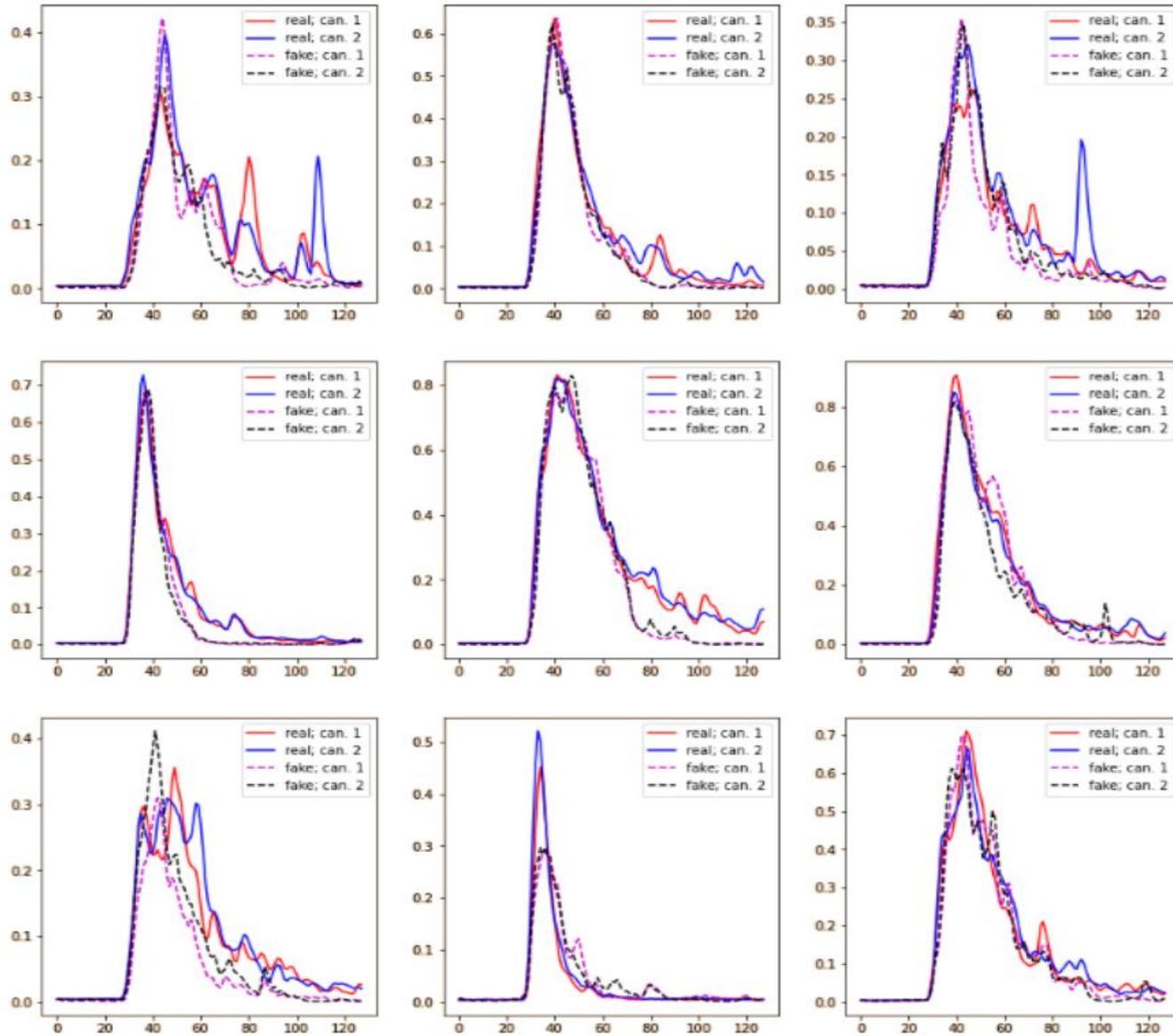
Поиск аномалий (Результаты)



Как видно из гистограмм, наиболее аномальные события присутствуют в наиболее легких частицах. Проанализировав детальней эти события, было обнаружено, что наиболее аномальные события имели низкий азимутальный угол падения космического луча (почти вертикальное падение)

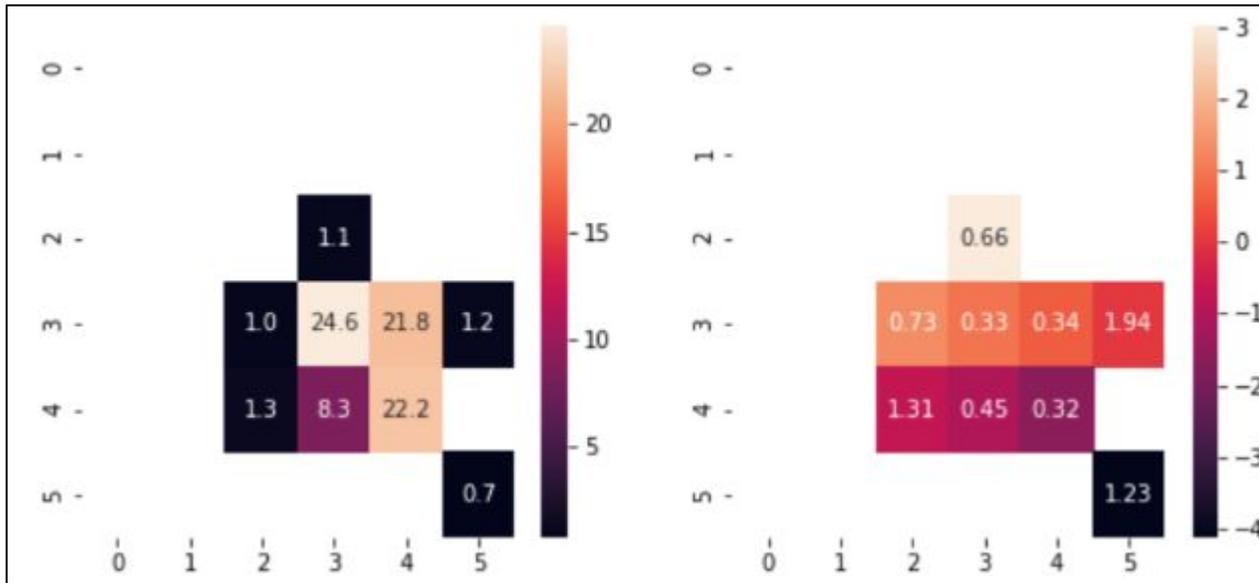
Результаты для протонов в качестве космического луча с использованием моделей QGSJET-II-03 (Q3 на гистограмме) и QGSJET-II-04 (Q4 на гистограмме) (1 и 4 гистограммы), для реальных данных с эксперимента (2ая гистограмма) и для ядер железа в качестве космического луча (3ая гистограмма)

Примеры наиболее аномальных событий



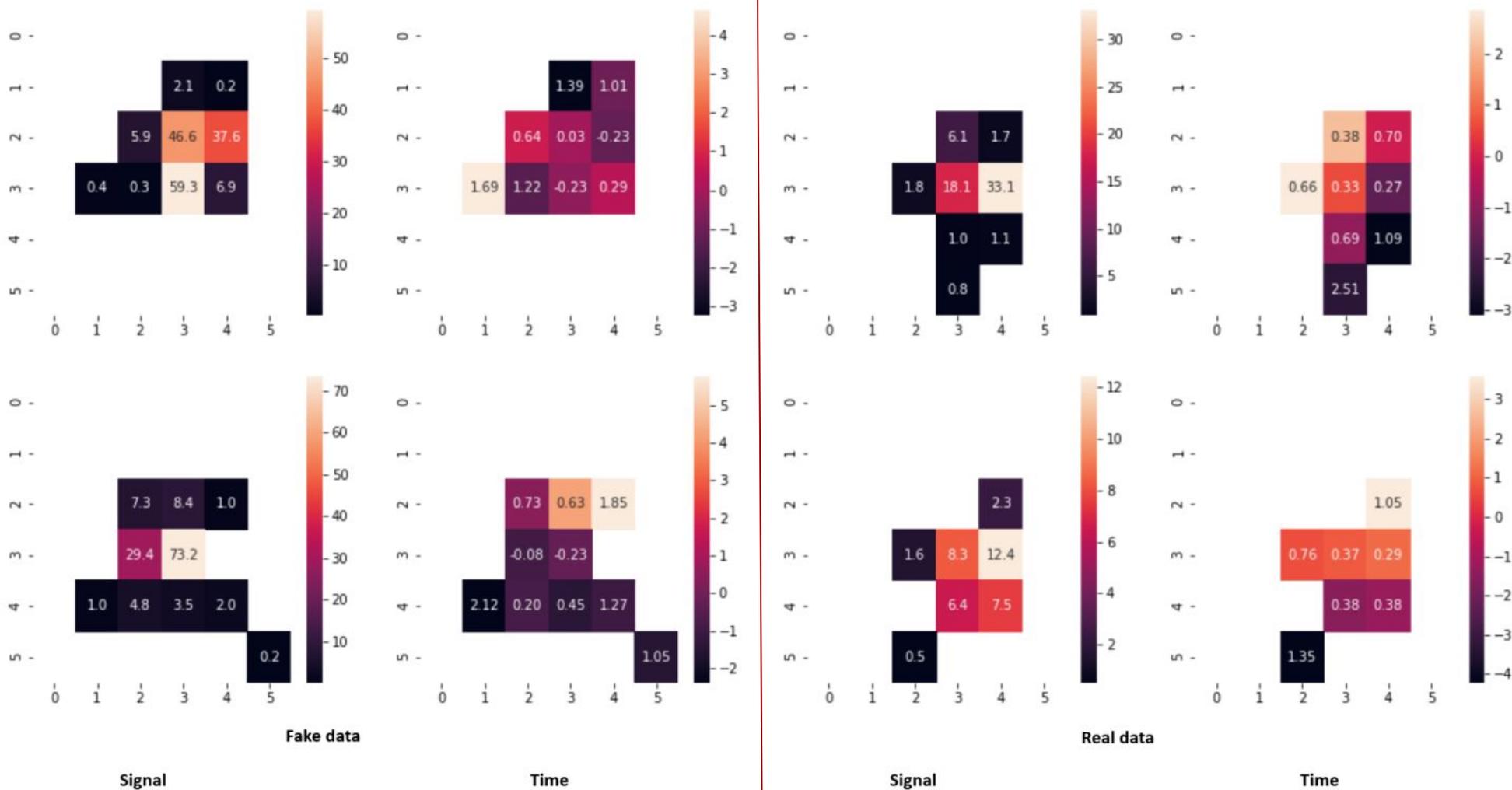
Генерация нескольких детекторов (данные)

Данные по каждому детектору имеют 4 канала



- **сигнал интегральный** (слева)
- **время реконструированного плоского фронта** (справа, цвет)
- **разница реального время детекции ШАЛ и плоского фронта** (справа, числовые значения)
- **маска сработавших детекторов** (nan/ not nan)

Генерація нескольких детекторов (Пример)



Дополнительная функция ошибок

Наша функция ошибок была основана на критерии правдоподобности хи-квадрат фитирования времени детекции ШАЛ.

Параметры ШАЛ оптимизировались путем минимизации величины хи-квадрат для подсчета ошибки для каждого батча.

$$t(\vec{R}) = t_0 + t_{\text{plane}}(\vec{R}) + a \times (1 + r/R_L)^{1.5} LDF(r)^{-0.5}$$

Где

$$t_{\text{plane}}(\vec{R}) = \frac{1}{c} \vec{n} \cdot (\vec{R} - \vec{R}_{\text{core}})$$

$$LDF(r) = \left(\frac{r}{R_m}\right)^{-1.2} \left(1 + \frac{r}{R_m}\right)^{-(\eta-1.2)} \left(1 + \frac{r^2}{R_1^2}\right)^{-0.6}$$

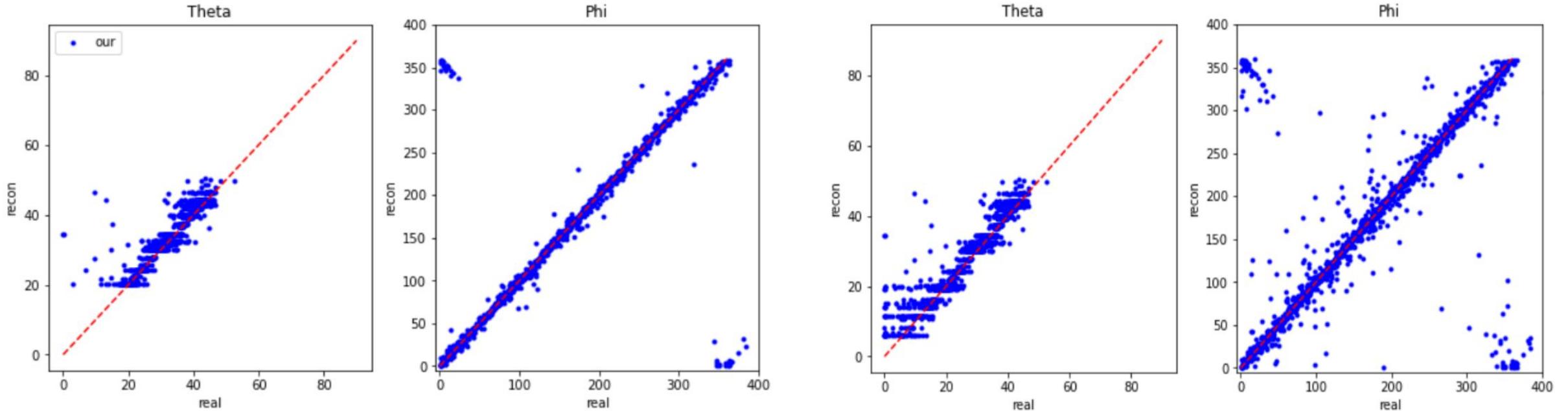
при

$$r = \sqrt{(\vec{R} - \vec{R}_{\text{core}})^2 - (\vec{n} \cdot (\vec{R} - \vec{R}_{\text{core}}))^2}$$

$$R_m = 90.0 \text{ m}, R_1 = 1000 \text{ m}, R_L = 30 \text{ m}$$

$$\eta = 3.97 - 1.79(\sec(\theta) - 1).$$

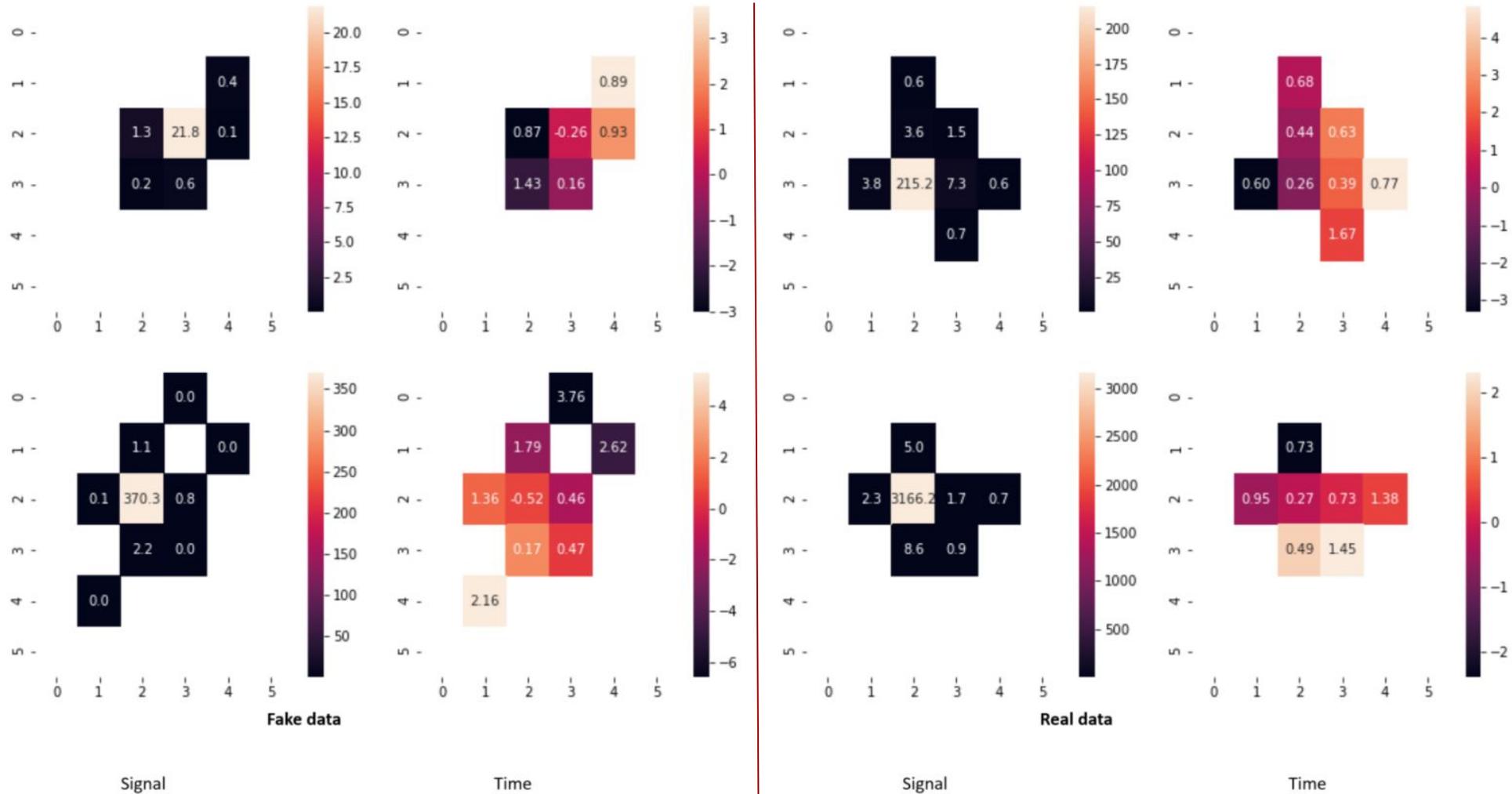
Точность фитирования параметров ШАЛ



Фитирование углов наклона оси ливня для всех событий (справа) и для событий с зенитным углом более 20 градусов (слева).

Таким образом данная функция ошибок применима только для событий с зенитным углом более 20 градусов

Генерация нескольких детекторов с использованием физической функции ошибок (Пример)



Спасибо за внимание

