на конечном этапе, либо дополнительно на этапе построения бивекторов (триплетов хитов).

// Нужно добавить больше подробностей о классификаторе - количество параметров, что это за параметры, как проходило обучение (например, размер обучающей/тестовой выборки, количество эпох, что-то из этого) // можно также добавить общее описание деревев решений

Были использованы следующие переменные для дискриминации:

1. азимутальный угол первого хита (ближайшего к мишени) на треке
2. тангенс угла импульса трека в горизонтальном направлении на первой станции трекера $T_x = p_x/p_z$, где p_x и p_z - компоненты импульса
3. $\log(p)$, где p - импульс трека
4. χ^2/NDF, $NDF = 2 \cdot N_{hits} - 5$ - число степеней свободы при фите трека
5. корреляция зарядов в хите $Q^{corr} = (Q_1 - Q_2)/(Q_1 + Q_2)$, где Q_1 и Q_2 - заряды кластеров стрипов с двух сторон считывания в детекторе, образующих данный хит
6. сумма корреляций зарядов в хитах $\sum_{i=1}^{N_{hits}} |Q_i^{corr}|$
7. максимальное число хитов в треке на последовательных станциях (без пропущенных станций)
8. степень перекрытия кластеров, образующих хит, т.е. число треков, в которых были использованы кластеры данного хита (т.к. хиты образуются при пересечении кластеров, то один и тот же кластер может образовать разные хиты, и, соответственно, участвовать в построении нескольких треков)
9. число хитов на треке $N_{hits} = 10 \cdot N_{Si} + N_{GEM}$, где N_{Si} и N_{GEM} - числа хитов на треке из кремниевых детекторов и GEM, соответственно.

Количество переменных типа 5 и 8 равно числу хитов, из которых построен трек, т.е. 4-6 в рассмотренной версии классификатора. Переменные типа 5 и 8 эффективны для идентификации фантомных хитов, т.к. такие хиты образуются пересечением стрипов, сработавших от разных частиц, и, соответственно, зарегестрировавших разное энерговыделение (заряды). Для правильных хитов декорреляция зарядов с двух сторон считывания происходит при перекрытии кластеров от близких частиц и случается с меньшей вероятностью. Также высокая степень перекрытия кластеров в хите говорит о его высокой вероятности быть фантомным. Пример распределений переменных для дискриминации приведён на рисунке 22 для 4-хитовых положительных треков.

При работе использовались параметры классификатора, установленные в обучающем макро на ROOT TMVA сайте [37]. Дополнительной оптимизации не делалось. Обучение проводилось раздельно для треков с разным числом хитов и разным знаком заряда. Оба подхода продемонстрировали высокую способность по дискриминации фейковых триплетов и треков, что показано на рисунке 23, и были интегрированы в инструментарий Vector Finder.

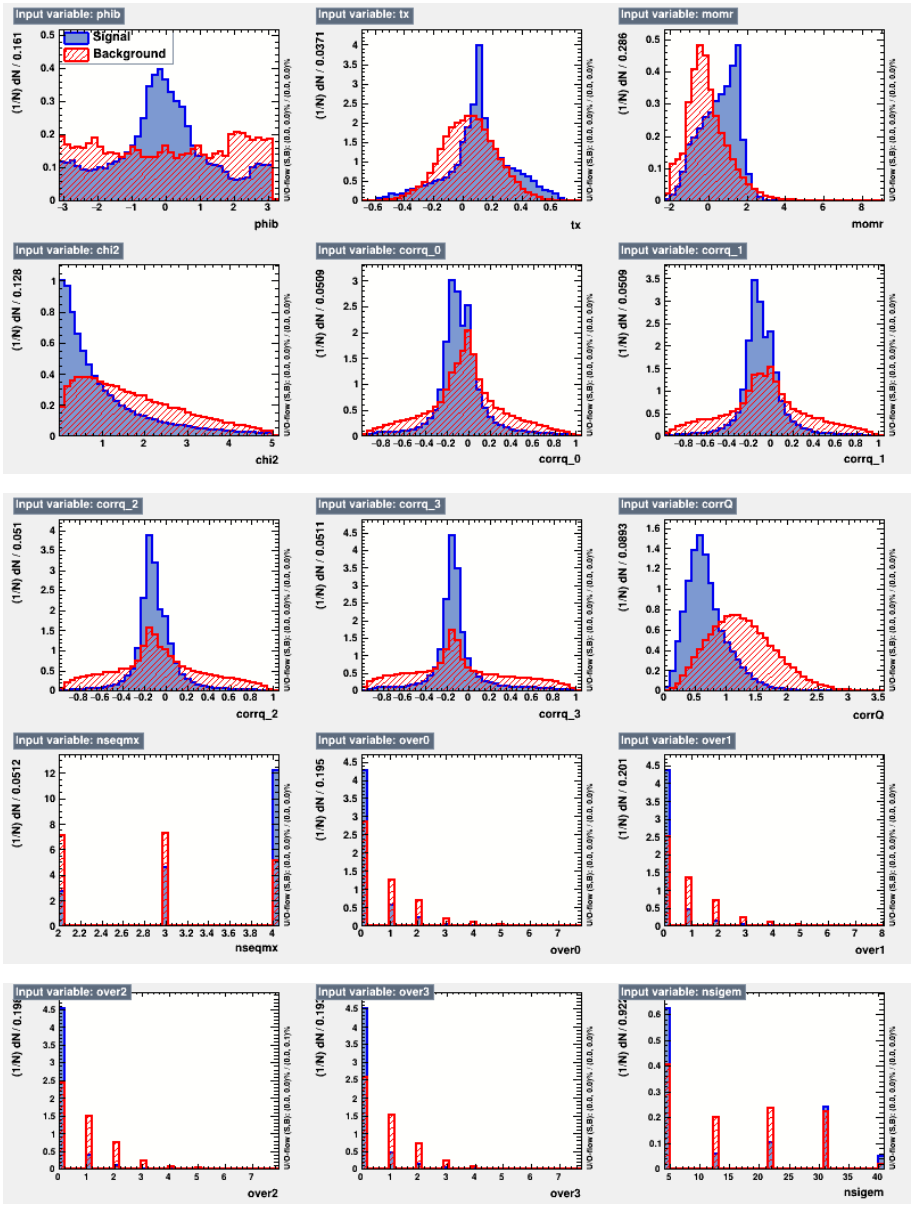


Рис. 22. Переменные, использованные для дискриминации фейковых треков с помощью TMVA для случая положительно-заряженных 4-хитовых треков (см. текст для описания). Синие гистограммы - ”хорошие” треки, красные - фейки.