

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Rekonstrukce mionové produkční hloubky v atmosférických sprškách kosmického záření
Jméno autora:	Bc. Antonín Kravka
Typ práce:	Diplomová práce
Fakulta:	Fakulta jaderná a fyzikálně inženýrská (FJFI)
Katedra:	Katedra fyziky
Oponent práce:	RNDr. Roman Lysák, Ph.D.
Pracoviště oponenta práce:	Fyzikální ústav AV ČR

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání práce považuji za poměrně náročné vzhledem k tomu, že obsahuje návrh nového algoritmu rekonstrukce mionové produkční hloubky v atmosférických sprškách s použitím moderní metody strojového učení a následně jeho testování na simulovaných datech.	
Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání práce považuji za splněno bez výhrad.	
Zvolený postup řešení	vhodny
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup i metody řešení považuji za vhodné. Zvolená metoda strojového učení je moderní a hojně využívaná i v jiných oblastech fyziky. V návrhu nového algoritmu student pečlivě optimalizoval jednotlivé parametry. Testování nového algoritmu probíhalo nejprve na jednodušších simulovaných datech a následně na složitějších (generovaných s více parametry).	
Odborná úroveň	výborna
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Odbornost práce je podle mého názoru na vysoké úrovni. Práce se zabývá metodou rekonstrukce mionové produkční hloubky. Taková rekonstrukce je důležitý problém v astro-částicové fyzice, protože potenciálně může pomoci k určení kompozice spršek kosmického záření, což je jeden z hlavních problémů astro-částicové fyziky. V návrhu nové metody se vycházelo z metody dosud používané a zároveň bylo provedeno srovnání mezi oběma metodami. Samotný algoritmus využívá jeden z typů strojového učení, k čemuž bylo zapotřebí detailní nastudování této problematiky v odborné literatuře.	
Formální a jazyková úroveň	výborna
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Formální úroveň hodnotím jako výbornou a nadstandardní u daného typu práce. Obrázky obsahují všechny relevantní informace včetně srozumitelného popisu. Práce je psaná v angličtině, která je na dobré úrovni. Je zde minimum typografických chyb a překlepů.	

Výběr zdrojů, korektnost citací

vyborne

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Výběr citovaných prací považuji za vhodný a dostatečný. Obrázky a grafy převzaté z literatury správně odkazují na zdroj. Citované práce jsou uvedeny v pořadí vyskytující se v textu a bibliografické údaje jsou zcela uvedeny.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Práce se člení do několika kapitol.

Úvodní kapitola představuje pěkný a srozumitelný úvod do fyziky kosmického záření. Uvítal by jsem ale tady alespoň stručný popis modelů hadronových interakcí (EPOS, QGSJET, Sibyll) a jaké jsou mezi nimi rozdíly. Druhá kapitola popisuje současnou metodu rekonstrukce mionové produkční hloubky. Zdá se, že rovnice (2.3) není uvedena ve správném tvaru (už jen z důvodu rozměrové analýzy).

Ve třetí kapitole je popis metod strojového učení od obecných principů s následným detailnějším popisem speciálního algoritmu („gradient boosted decision tree“), který se využívá v další analýze.

Samotné vlastní výsledky jsou shrnuty v kapitolách 4 a 5. Výsledky jsou velmi dobře popsány, postupuje se systematicky od jednodušších testů po složitější. Jsou tady uvedeny poměrně obsáhlé a komplexní studie. Velké úsilí je věnováno optimalizaci parametrů metody strojového učení. Výsledky poukazují na dobrou použitelnost zvolené metody a na potenciální vylepšení oproti existující metodě. Uvítal bych větší srovnání vůči existující metodě rekonstrukce a to hlavně v kapitolách 4.2.2 a 4.2.3, kde se nová metoda studuje na simulovaných případech, které jsou blíže realitě ve srovnání s předchozím testováním.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Celkově hodnotím jak kvalitu samotné práce tak i její písemné zpracování na výborné úrovni. Student musel zvládnout nastudovat jednak problematiku spršek kosmického záření a také problematiku metod strojového učení. Dosažené výsledky poukazují na vylepšení metody rekonstrukce mionové produkční hloubky vůči v současnosti používané metodě a mohou přispět k vylepšení výsledků současného Pierre Auger experimentu, případně jeho budoucí vylepšené verze. Níže uvádím několik otázek a námětu do diskuse při obhajobě:

- Tab 4.2.: není jasné jak a na základě čeho byla provedena optimalizace hodnot hyperparametrů. Použila se k tomu stejná data jako k samotnému trénování GBDT algoritmu („training dataset“)?

- Jako vstupní proměnné GBDT se používají proměnné, které se používají i v současné metodě rekonstrukce mionové produkční hloubky (MPD). Existují i jiné, dodatečně pozorovatelné, které se měří v astročásticových experimentech, a které by mohly být citlivé na MPD? Pokud ano, byl nějaký důvod proč je nepoužít jako vstup ve Vašem GBDT modelu?

- Obr.4.3.: podle popisu jsou na pravé straně zobrazeny průměry 100 spršek, podle počtu eventů to ale vypadá, že to je suma histogramů ze 100 spršek.

- Obr.4.3.: podle popisu jsou na pravé straně zobrazeny průměry 100 spršek, podle počtu eventů to ale vypadá, že to je suma histogramů ze 100 spršek.

- Obr. 4.6.: Není jasné, proč histogramy pro 'ML' metodu jsou odlišné (i když podobné) od histogramů z Obr. 4.3. Byla použita jiná selekce eventů?

- Obr. 4.17 a 4.18: jelikož X_{max} je důležitá proměnná pro další studie, bylo by vhodné vidět pro tyto obrázky srovnání pro různé hadronizační modely a také srovnání se standardní metodou. Dělali jste takové porovnání?

Obr. 4.20 a Tab.4.1: zdá se, že je mezi obrázkem a tabulkou inkonzistence: obrázek obsahuje výsledky pro Sibyll model a ne pro EPOS model, ale podle tabulky byly eventy generovány na základě EPOS modelu a ne podle Sibyll modelu. Máte pro to vysvětlení?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm

Datum: 26.05.2023

Podpis:

