

# Raport z ćwiczenia z Pracowni Specjalizacyjnej

Agnieszka Ilnicka

## 1 Pi of the Sky

Projekt "Pi of the Sky" działa od 2004 roku, oraz prowadzony jest przy współpracy wielu jednostek w Polsce, m. in.: Wydziały Fizyki UW i PW, Centrum Badań Kosmicznych PAN i Instytut Problemów Jądrowych PAN. Projekt jest nastawiony głównie na zbieranie informacji na temat szybkozmiennych zjawisk astronomicznych, takich jak błyski gamma. Jest to możliwe dzięki budowie teleskopu, który posiada duże pole widzenia, oraz dodatkowo umieszczony jest na stelarzu, który umożliwia szybką zmianę obserwowanego fragmentu nieba w wypadku pojawienia się zjawisk. Oprócz zbierania informacji na temat błysków gamma, "Pi of the Sky" zbiera również informacje na temat innych obiektów astronomicznych, takich jak wybuchy nowych i supernowych czy gwiazdy zmienne.

Celem "Pi of the Sky" jest jak największa automatyzacja, który dzięki zastosowanym algorytmom on-line mógłby reagować na pojawiające się błyski. Również spora część analizy uzyskiwanych zdjęć następuje automatycznie bezpośrednio po odczycie.

Do niedawna detektor eksperymentu znajdował się na terenie obserwatorium Uniwersytetu Warszawskiego w Las Campanas Observatory na pustyni Atacama w Chile, jednak pod koniec 2010 roku nowy zestaw detekcyjny został zainstalowany w centrum INTA na wybrzeżu Oceanu Atlantyckiego w Hiszpanii. Dane, które podczas ćwiczenia były poddawane analizie pochodzą z tego właśnie nowego detektora i zostały zebrane na początku 2011 roku.

## 2 Opis i cel ćwiczenia

Celem ćwiczenia było stworzenie programu, który umożliwiłby analizę danych dotyczących gwiazd, które zostały zebrane przez nowy detektor "Pi of the Sky" w INTA w trakcie nocy 2/3, 3/4 i 4/5 stycznia 2011. W tym czasie wyłączono system śledzący możliwe kierunki pojawiających się błysków gamma, aby zebrać dużą ilość zdjęć jednego obszaru nieba. W skład omawianej analizy wchodziły dane dotyczące około 70 000 gwiazd, które zostały zebrane podczas tych trzech nocy. W każdym pliku znajdowało się około 250-300 pomiarów pochodzących z 20 zsumowanych zdjęć o czasie ekspozycji równym 10 sekund. Informacje z każdego pomiaru zawierały informacje na temat czasu zbierania danych, magnitudo gwiazdy (również katalogowej, jeśli jest dostępne), jej położenia oraz numeru kamery, na której była zarejestrowana. Te dane były podstawą do wykonania ćwiczenia.

Celami, które miał realizować program było:

- Stworzenie efektywnego algorytmu umożliwiającego odcinanie długich "ogonów" (spowodowanych zakłuceniami w pomiarach - np. przejściem chmur), który umożliwiłby wyznaczenie dla każdej gwiazdy najbardziej prawdopodobnej wartości magnitudo oraz błędu pomiaru,
- Przedstawienie zależności odchylenia standardowego od wartości magnitudo,
- Sporządzenie profili, czyli zależności uśrednionych wartości odchylenia po przedziałach magnitudo.

Dwa ostatnie podpunkty należało zrobić z podziałem na poszczególne kamery, a następnie zbiorczo dla wszystkich gwiazd. Jako dane pomocnicze, które pomagały przy optymalizacji procedury odcinania ogonów posłużyły dane na temat czasowego rozkładu ilości rekonstruowanych gwiazd.

## 3 Wykonanie ćwiczenia

Efektom ćwiczenia jest program używający bibliotek pakietu ROOT. Program można podzielić na dwie zasadnicze części: algorytm odcinania "ogonów" oraz tworzenie wykresów dla całości lub części (podział na kamery) danych.

### 3.1 Algorytmy odcinania "ogonów"

Podstawą do działania poszczególnych algorytmów wymienionych poniżej było obliczenie podstawowych parametrów niezmodyfikowanych danych, takich jak średnia, odchylenie standardowe, wartości największa i najmniejsza.

#### 3.1.1 Metoda procentowego odcinania ogonów od zmienianej średniej

Pierwsza zaproponowana metoda (Metoda I) polega na liczeniu średniej po czym odrzuceniu jednego ze skrajnych punktów, który znajduje się od niej dalej, po czym liczeniu nowej średniej:

$$M_{new} = \frac{M_{old} \cdot i - x_{re}}{i - 1}$$

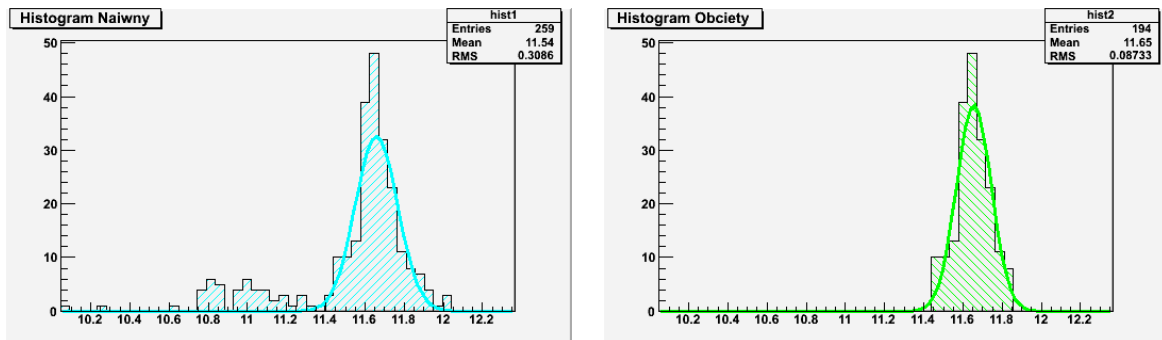
, dzięki czemu przy każdej iteracji odrzucana wartość jest odnoszona do nowej, poprawionej średniej, która potencjalnie lepiej oddaje wartość po oddzieleniu szumu. Odcinanie kończy się, gdy stosunek pozostających danych do początkowych osiągnie zadany przez użytkownika procent. Jest to prosty i szybki algorytm, aczkolwiek dzięki przesuwaniu średniej przy każdej iteracji, daje pożądane efekty. Jego mankamentem może być fakt, iż procent odcinanych danych jest niezależny od początkowego rozkładu, co może spowodować zarówno zbyt nadmierną jak również niewystarczającą odrzucenie punktów.

W poniższym opracowaniu odrzucano 15%, 10% i 5% danych.

Aby zapobiec sztuczemu polepszaniu wyników poprzez odcinanie "poprawnie" zmierzonych wartości zastosowano poprawki (przez które wymnażane były wartości dyspersji):

- 1.1479 - przy odrzucaniu 5% przypadków
- 1.2693 - przy odrzucaniu 10%

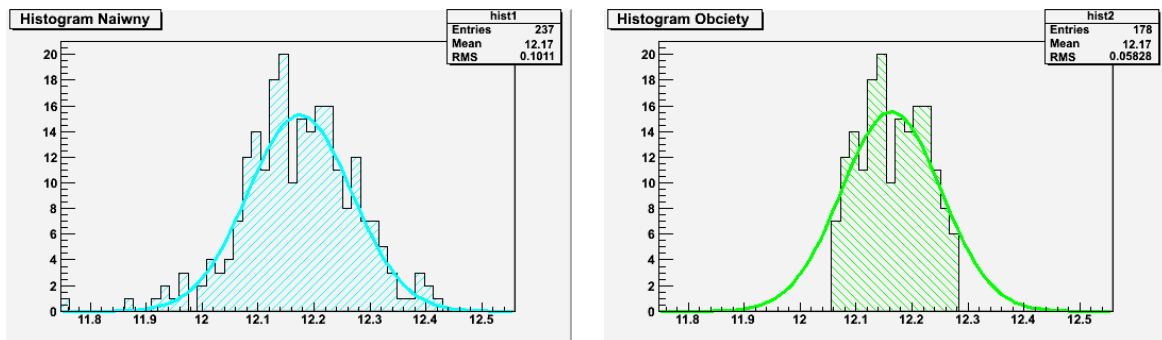
- 1.3857 - przy odrzucaniu 15%



(a) przed poprawką:

(b) po poprawce:

Rysunek 1: Przykład działania metody I



(a) przed poprawką:

(b) po poprawce:

Rysunek 2: Przykład kiedy zawodzi metoda I

### 3.1.2 Metoda odcinania wielokrotności standardowego odchylenia od średniej

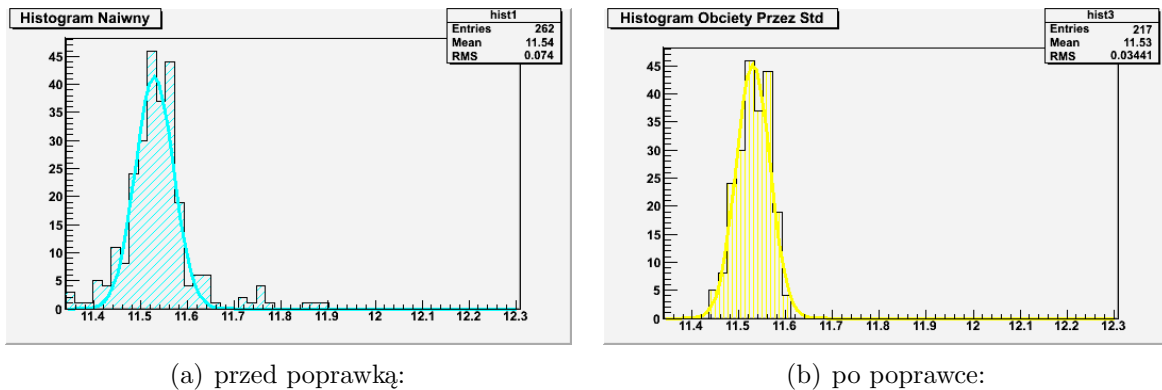
Kolejna metoda (Metoda II) polegała na odcinaniu punktów, które znajdują się dalej od średniej wyliczonej z pierwotnych danych niż pewna wielokrotność odchylenia standardowego. Jest to szybka i prosta metoda, posiadająca jednak poważny mankament - średnia dla rozkładów o długim "ogonie" może poważnie przesunąć się w stronę tego "ogona", co w wyniku rekonstrukcji doprowadzi do otrzymania rozkładu, który w widoczny sposób odbiega, od oczekiwanego. Podjęto również próbę poprawienia poprzez używanie wartości średniej oraz STD otrzymanych z dopasowania funkcji gaussowskiej do danych pierwotnych. Jednak w niektórych przypadkach doprowadzało to do otrzymania bezsensownych wyników z powodu np. niemożności dopasowania funkcji, bądź dopasowania, w którym wartość średnia znajduje się poza zakresem danych lub też odchylenie standardowe jest bardzo duże (nawet przekraczające zakres danych).

Zastosowane wielokrotności STD to 1, 1.5 oraz 2.

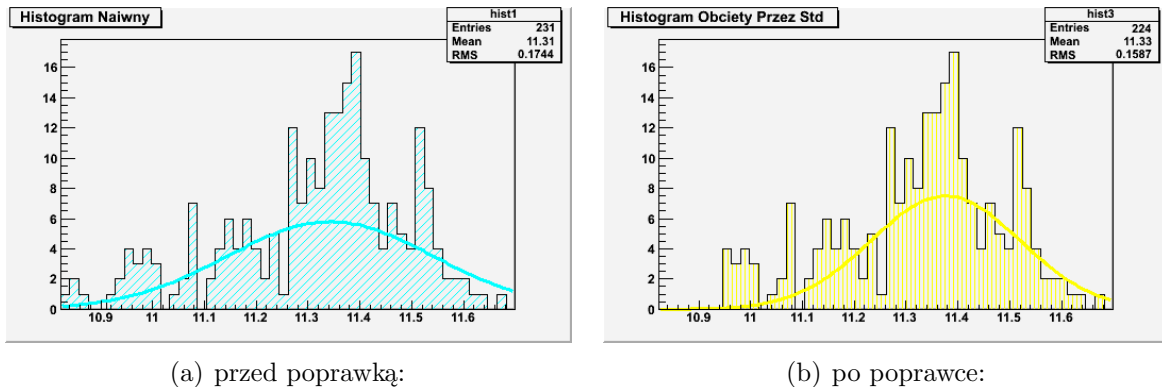
Poprawki stosowane dla tej metody:

- 1.1368 - przy odcinaniu ogonów od  $\pm 1 \sigma$

- 1.3465 - przy odcinaniu od  $\pm 1.5 \sigma$
- 1.8534 - przy odcinaniu od  $\pm 2 \sigma$



Rysunek 3: Przykład działania metody II

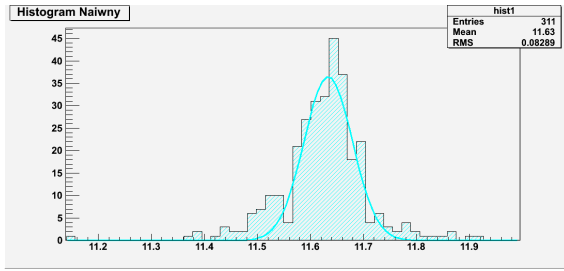


Rysunek 4: Przykład kiedy zawodzi metoda II

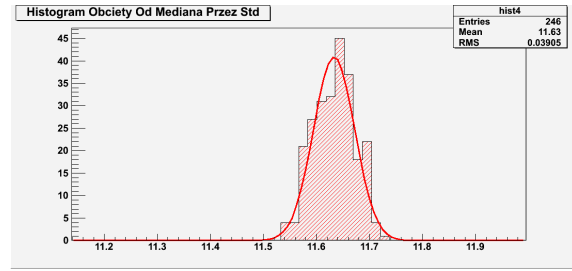
### 3.1.3 Metoda odcinania wielokrotności standardowego odchylenia od mediany

Kolejna metoda (Metoda III) jest modyfikacją Metody II, polegająca na zastąpieniu średniej medianą. Modyfikacja ta motywowana jest z tym, że średnia bardziej ulega zmianie przy pojawieniu się długich (a niekoniecznie licznych) "ogonów". Jednak również ta metoda może zawieść dla rozkładów o silniej rozbudowanych ogonach. W praktyce okazało się, że Metoda II i III w prawie wszystkich przypadkach daje takie same efekty.

Zastosowano takie same wielokrotności oraz poprawki jak w metodzie II.



(a) przed poprawką:



(b) po poprawce:

Rysunek 5: Przykład działania metody III

### 3.1.4 Metoda odrzucania przedziałów czasowych

Metoda ta polegała na odrzucaniu wszystkich pomiarów z godzin, w których odrzucane było w Metodzie I przy 10% odrzucanych danych więcej niż 10% pomiarów z danego okienka półgodzinnego. Wycinane były pomiary z okresów: pierwszej nocy: 19-20 i 1-5, drugiej: 19-21 i 2:30-5, trzeciej: 23:30-1:30 i 3-4, czyli czas, w którym pomiary nie są uwzględniane stanowi około 30%. Wadą tej metody jest odrzucanie dużej ilości pomiarów, a ze względu na charakterystykę wykresów z Rysunku 7 trudno o rozsądny dobór mniej restrykcyjnego cięcia. Jednak ze względu na metodologię odrzucania punktów, rozkład nie powinien ulegać sztucznemu zawężeniu.

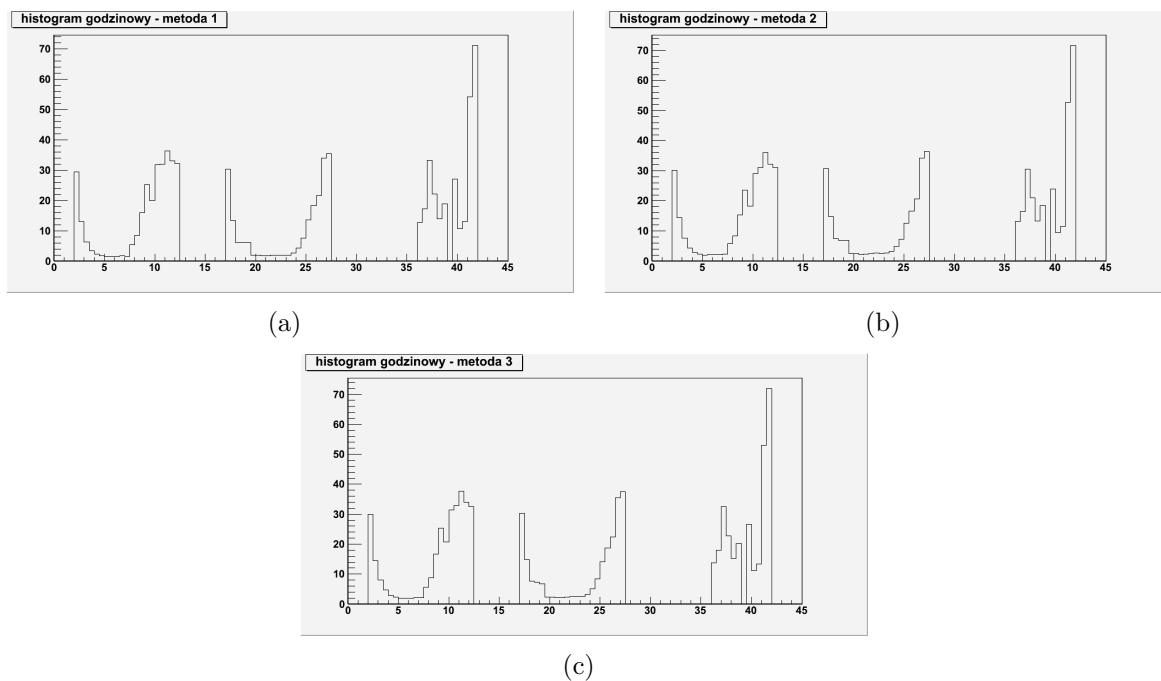
## 3.2 Wyniki zbiorcze

Po zaprogramowaniu algorytmów odrzucania "ogonów" kolejnym krokiem było sporządzenie zbiorczych wykresów wartości odchylenia standardowego względem magnitudo oraz profili, powstałych w wyniku uśredniania STD po przedziałach magnitudo. Wykresy zostały wykonane przy użyciu klas TH2F oraz TProfile dostępnych w pakiecie ROOT. Do ich sporządzenia użyto danych bez poprawek, jak również danych z poprawkami uzyskanymi różnymi metodami.

### 3.2.1 Porównanie rozkładu czasowego odrzucanych pomiarów z liczbą rekonstruowanych gwiazd

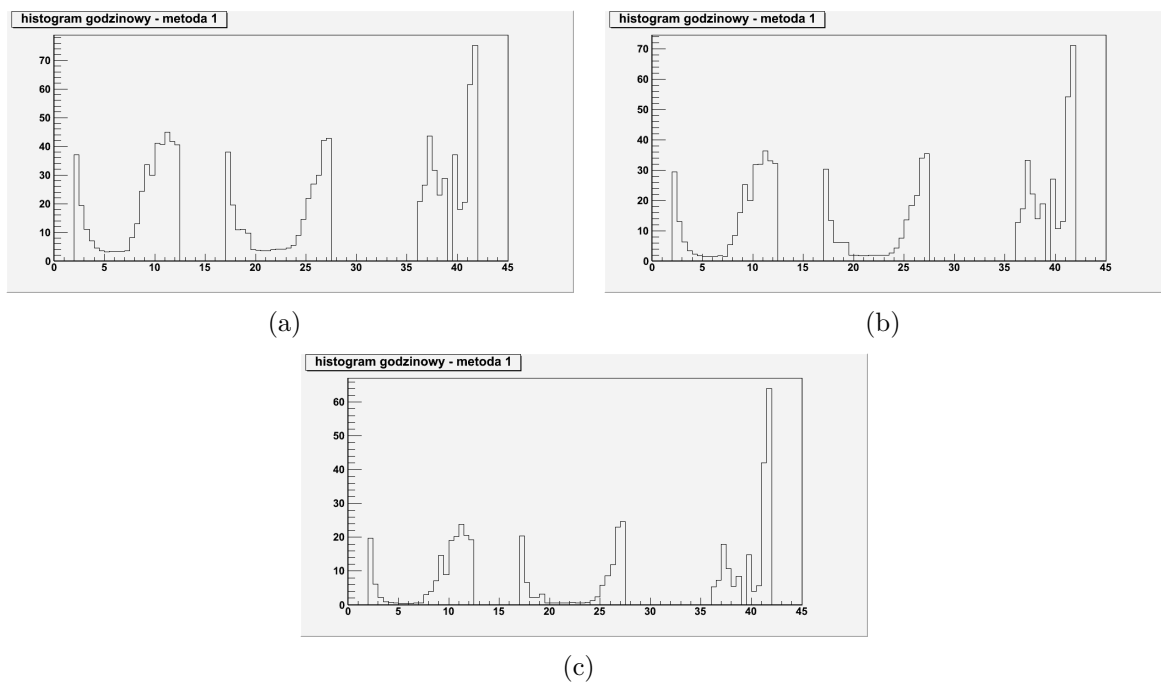
Jednym z aspektów ćwiczenia było również porównanie rozkładu czasowego odrzucanych pomiarów z dostępnymi informacjami na temat ilości rekonstruowanych gwiazd w ciągu nocy. Na wykresach widać rozkład odrzucanych pomiarów (dla parametru 85% dla Metody I oraz 1 dla Metod II i III). Zbieżność tych obu parametrów została wybrana jako test na poprawność działania metod odrzucania "ogonów", oraz jest pomocna przy dobieraniu najoptymalniejszych parametrów.

Warto zwrócić uwagę, że w półgodzinnym okienku czasowym widocznym na histogramach jest około 420 000 pomiarów. Jak widać wszystkie histogramy wykazują podobną tendencję, z wyraźnymi maksimumami, które zbiegają się w czasie z minimumami w ilości rekonstruowanych gwiazd. Jednak na pokazanych wykresach, mimo podobnej wysokości maksimumów, wyraźnie widać, iż na wykresie pochodzącym z Metody I tło poza pikami jest wyraźnie niższe. Potwierdza to wybór tej metody jako najbardziej optymalnej.



Rysunek 6: Porównanie rozkładów czasowych odrzucanych punktów w trzech metodach

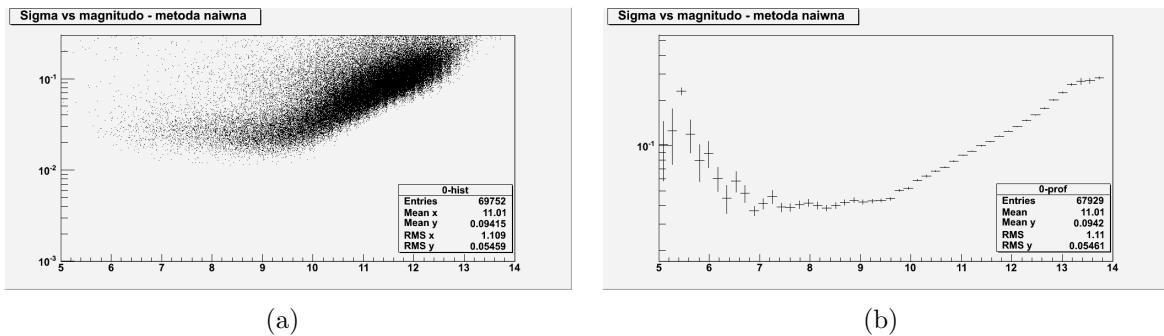
Na kolejnym rysunku zestawiono wykresy dla Metody I dla różnych wartości parametru (85%, 90% oraz 95%), co umożliwia wybranie najbardziej optymalnego. Wydaje się, że dobrym wyborem parametru jest odrzucanie 10% pomiarów, gdyż w newralicznych godzinach wycinana jest podobna ilość punktów, co przy wyrzucaniu 15% pomiarów, natomiast w godzinach, w których pomiary nie powinny być wycinane pozostaje tło rzędu 2-3%.



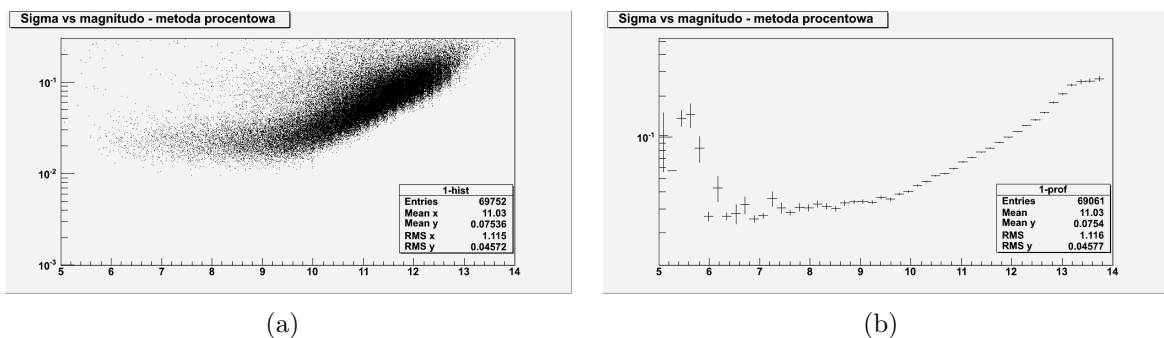
Rysunek 7: Porównanie rozkładów czasowych dla różnych wartości parametru w metodzie I

### 3.2.2 Wyniki zsumowane ze wszystkich kamer

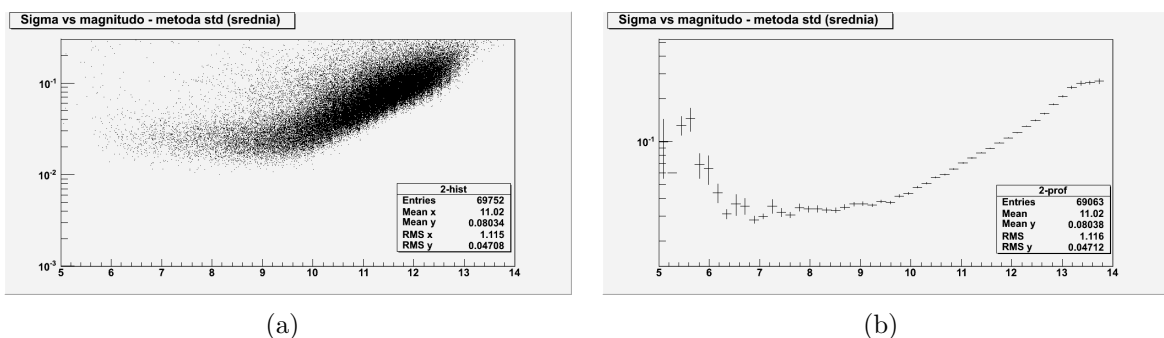
Poniżej przedstawiono wykresy zależności STD od magnitudo oraz profile. Wykresy są w skali logarytmicznej, aby łatwiej można było dostrzec ich szczegóły. Wykres przedstawiono w zakresie STD od  $10^{-6}$  do 0.3, co spowodowało wyrzucenie sporadycznie pojawiających się wyższych wartości, jednak dało lepszy wgląd w istotną część wykresu. Wykresy pokazują zależność dla danych bez obróbki, oraz dla danych poddanych ulepszeniu trzema metodami (o współczynnikach 90% oraz 1), opisanymi wyżej.



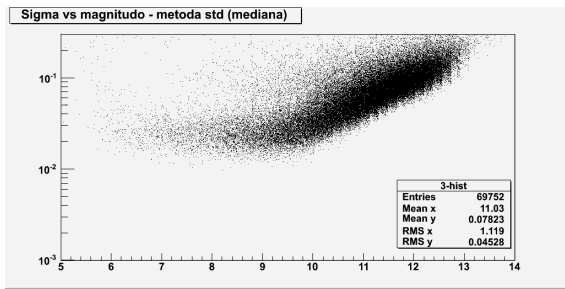
Rysunek 8: Zależność STD vs magnitudo oraz profil - bez odrzucania



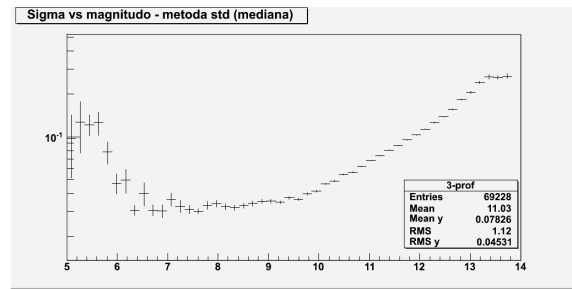
Rysunek 9: Zależność STD vs magnitudo oraz profil - metoda I



Rysunek 10: Zależność STD vs magnitudo oraz profil - metoda II

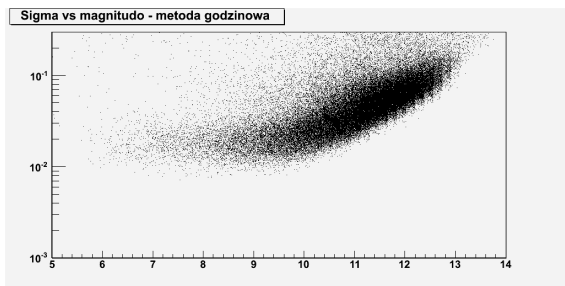


(a)

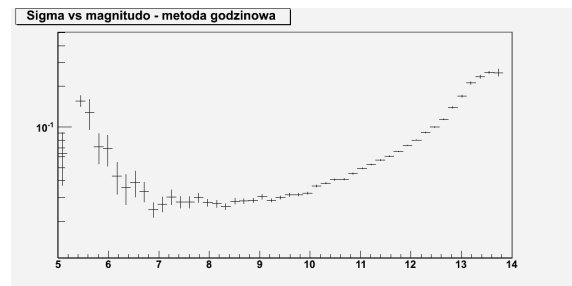


(b)

Rysunek 11: Zależność STD vs magnitudo oraz profil - metoda III



(a)

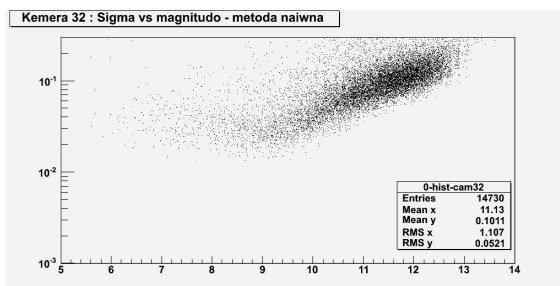


(b)

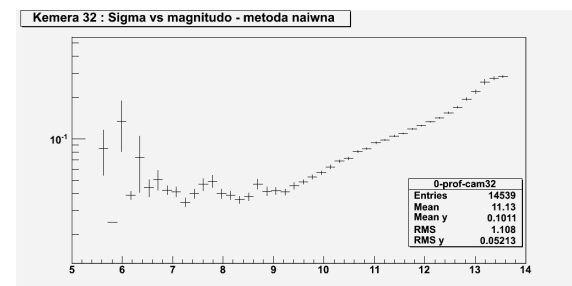
Rysunek 12: Zależność STD vs magnitudo oraz profil - metoda IV

### 3.2.3 Wyniki z podziałem na kamery

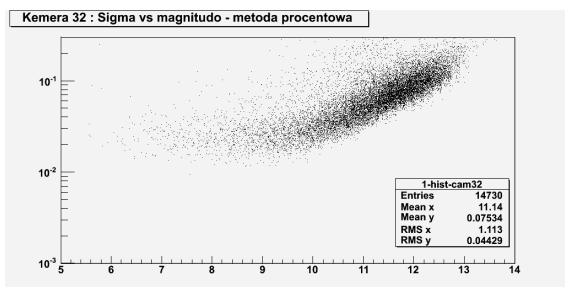
W tej części zaprezentowano wykresy dla poszczególnych kamer, jednak tylko dla danych bez obróbki oraz po zastosowaniu metody I (90%).



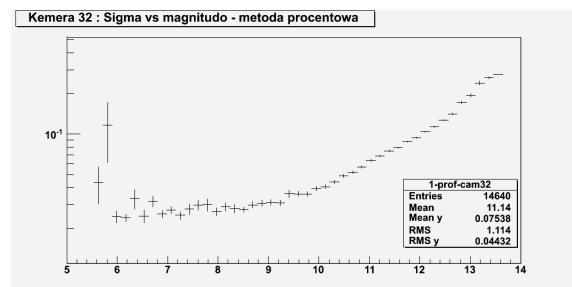
(a)



(b)



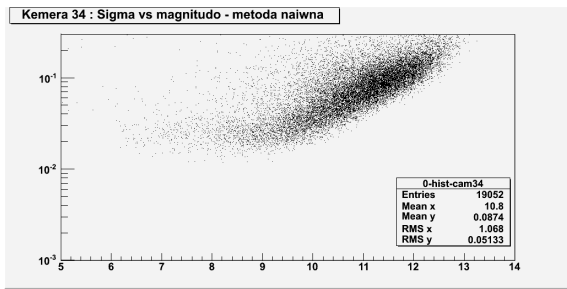
(c)



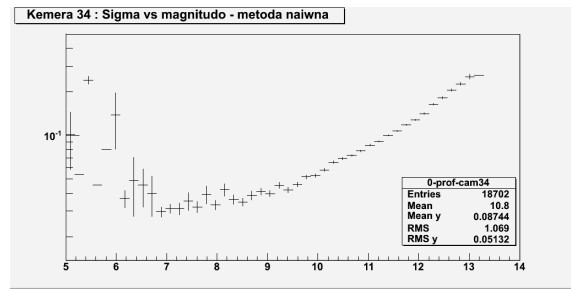
(d)

Rysunek 13: Zależność STD vs magnitudo oraz profil - kamera 32

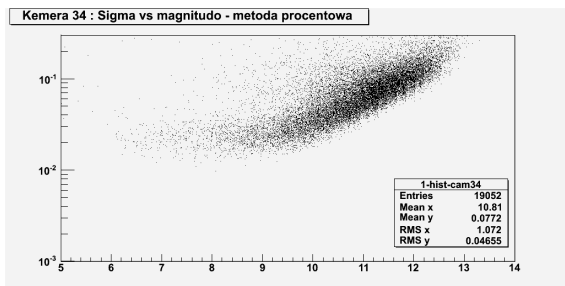




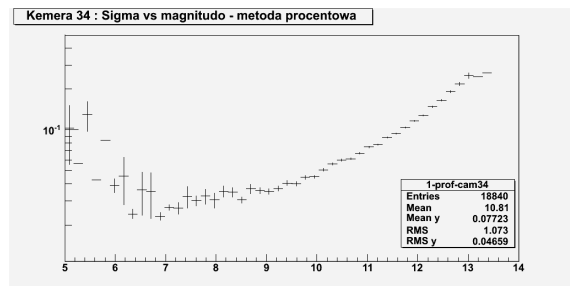
(a)



(b)

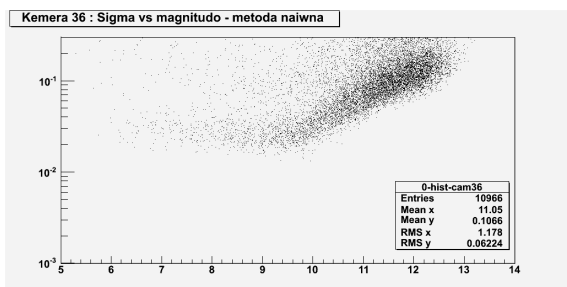


(c)

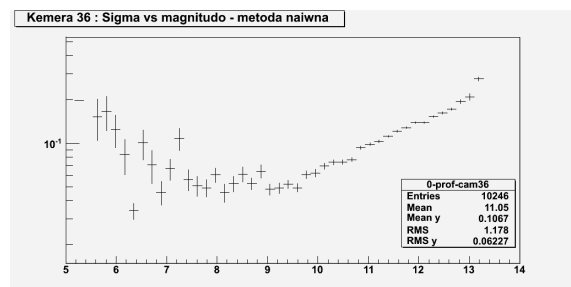


(d)

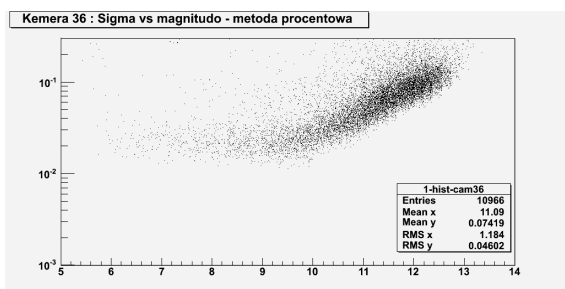
Rysunek 14: Zależność STD vs magnitudo oraz profil - kamera 34



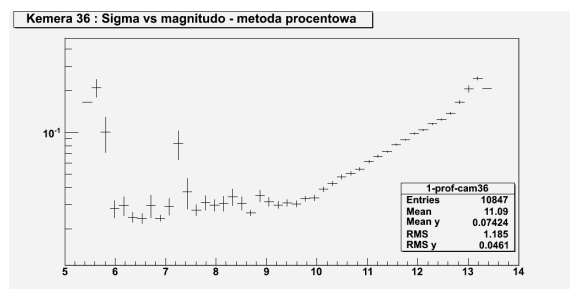
(a)



(b)

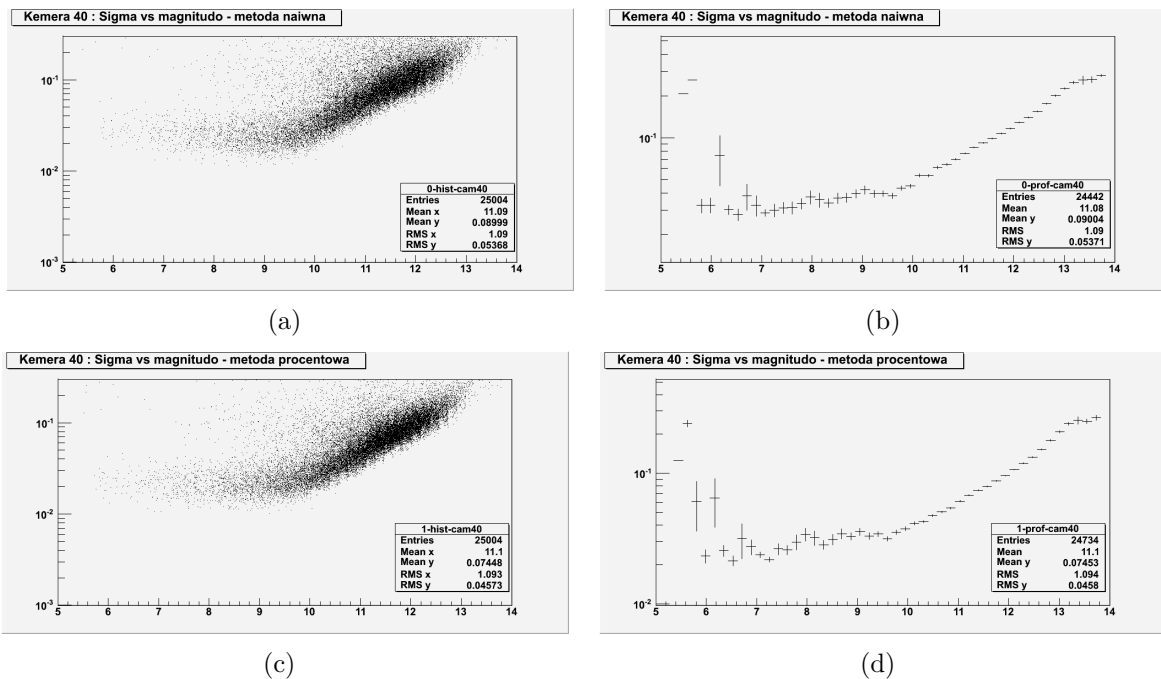


(c)



(d)

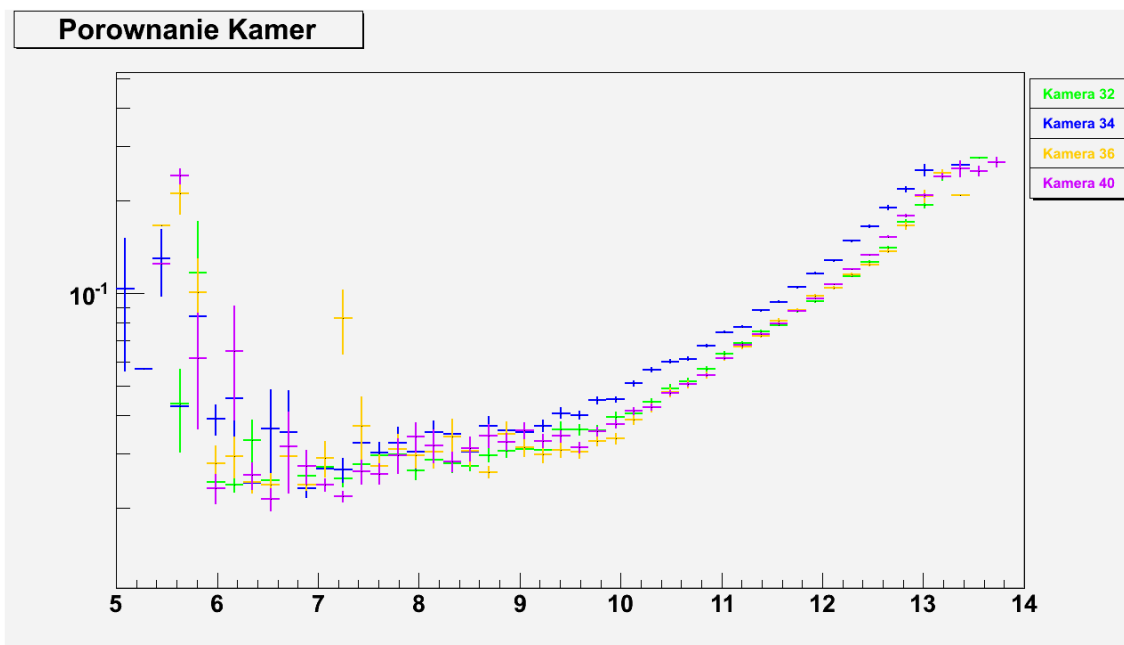
Rysunek 15: Zależność STD vs magnitudo oraz profil - kamera 36



Rysunek 16: Zależność STD vs magnitudo oraz profil - kamera 40

### 3.2.4 Porównanie pracy poszczególnych kamer

Poniżej zebrane zostały profile dla wszystkich kamer otrzymane przy pomocy Metody I, w celu porównania ich pracy.

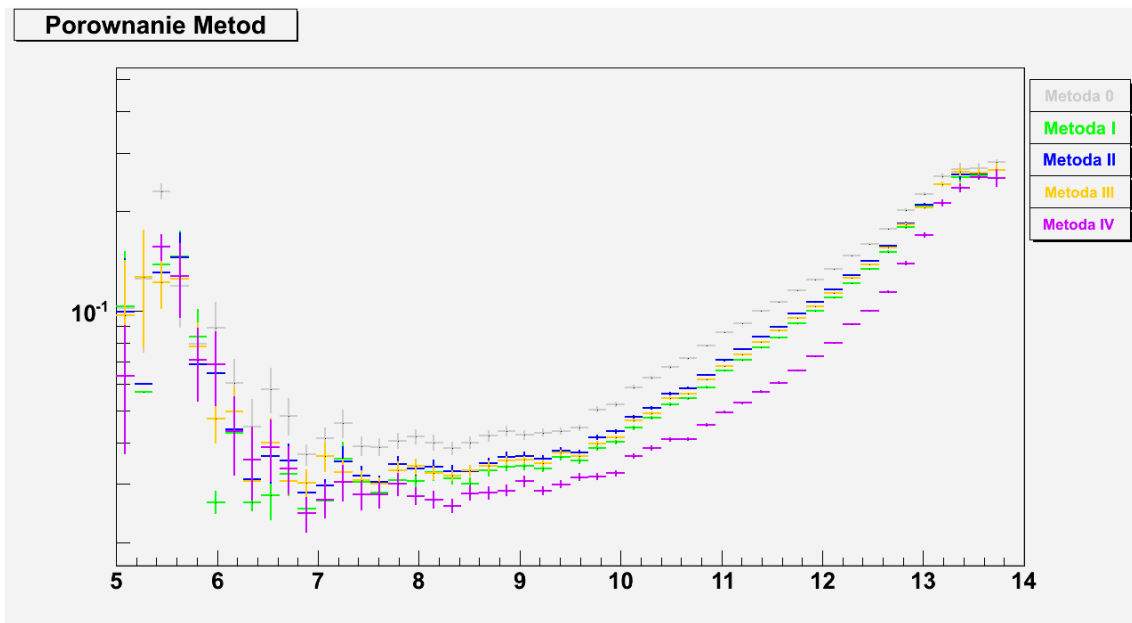


Rysunek 17: Profile otrzymane z Metody I dla wszystkich kamer

Na wykresach widać, że dla małych wartości magnitudo istnieją spore odchylenia od monotonicznego wzrostu, jednak później wzrost jest dosyć stabilny, zachwiany jedynie dla skrajnie wysokich wartości magnitudo.

## 4 Wnioski

### 4.1 Wybór algorytmu



Rysunek 18: Profile otrzymane dla różnych metod dla danych z wszystkich kamer

Na wykresie wyraźnie widać, iż najniższe wartości  $\sigma$  vs magnitudo daje Metoda IV. Trudno jednak porównywać ją z pozostałymi, gdyż jest to metoda oparta na zupełnie innych zasadach (odrzućanie danych następuje globalnie a nie na poziomie poszczególnych pomiarów), a dodatkowo jest bardzo restrykcyjna w porównaniu z nimi i powoduje odcięcie aż około 30% danych. Jest ona również zależna od metody, z której przy pomocy analizy rozkładu godzinowego odrzucanych pomiarów dobierane są odpowiednie cięcia czasowe.

Spośród pozostałych metod, najlepsza okazała się Metoda I - wynika to głównie z relatywnie często pojawiających się rozkładów, które posiadają duże wartości STD co czyni dwie pozostałe Metody praktycznie bezradnymi. Wyraźnie wyższe STD dla Metod I-III wynikają zapewne z obecności poprawek, które działają również w przypadku odcięcia tylko "właściwego" ogona.

Dalsza optymalizacja Metody I poprzez porównywanie rozkładów godzinowych odrzucanych pomiarów doprowadziła do wniosku, że najbardziej optymalnym parametrem ilości odrzucanych pomiarów jest 10%. W przypadku Metody IV trudno natomiast dobrać parametr, który mógłby pozwolić na jej optymalizację.

### 4.2 Wybór kamery

Porównanie profili sporządzonych dla poszczególnych kamer wskazuje, iż najlepiej pracującą kamerą jest kamera numer 32. W porównaniu z innymi kamerami posiada ona najmniejsze rozbieżności dla małych wartości magnitudo, a także nie posiada "zachwiań" dla skrajnie wysokich wartości - jej profil najbardziej przypomina funkcję monotoniczną. Najbardziej wartościami od pozostałych kamer odbiega natomiast kamera 34.