Extraction of Muons and Validation of MC

for the First G-APD Cherenkov Telescope

Maximilian Nöthe   Jens Buß

technische universität dortmund

SFB 876 Providing Information by Resource-Constrained Data Analysis

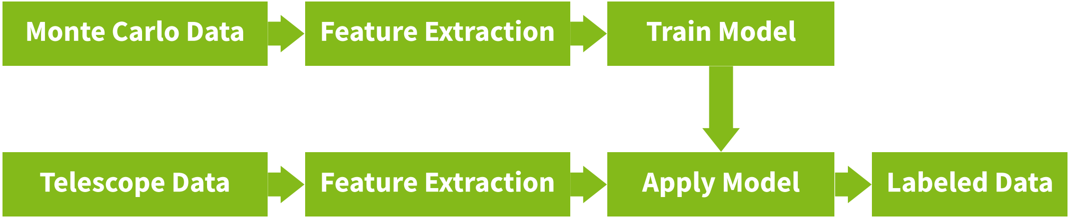
CTA Calibration Meeting 2015

Overview

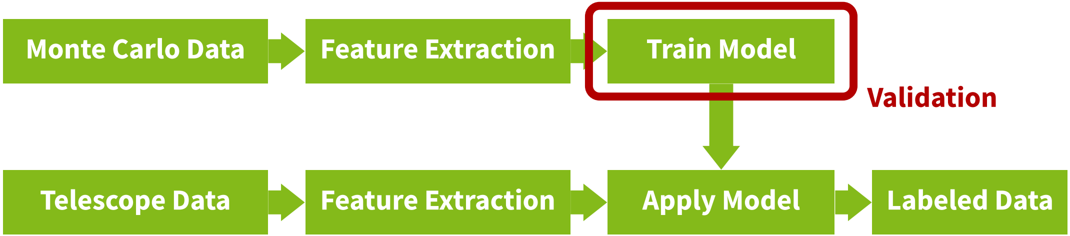
- Extraction of Muons
  - Training Data
  - CORSIKA MC
  - Feature Extraction with FACT-Tools
  - Machine Learning
- Validation of MC
  - Time Resolution
  - PSF
- Conclusion

[Miguel Claro]

Analysis Overview



Analysis Overview



Training Data

- → Supervised models need training data with known truth:
  - → Particle type (classification)
  - → Energy (regression)

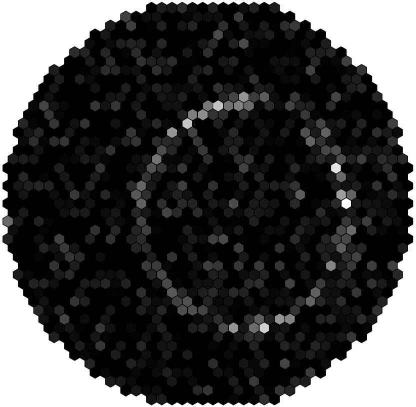
- → This can be hard for muon ring images
  - → Hadronic primary → muon ring image
  - → Muon primary → decay or air shower

$\Longrightarrow$ Signal in background MC and background in signal MC

Simulation with CORSIKA – Muons

To get a clean muon ring sample, we use the following settings:

Energy range	$10\,\mathrm{GeV}$ to $1\,\mathrm{TeV}$
Spectral index	-2.7
Starting altitude	$300\,\mathrm{m}$ above FACT
Zenith distance	$0°$
Maximum Impact Parameter	$1.75\,\mathrm{m}$ (mirror radius)
ViewCone	$2°$



Simulation with CORSIKA – Protons

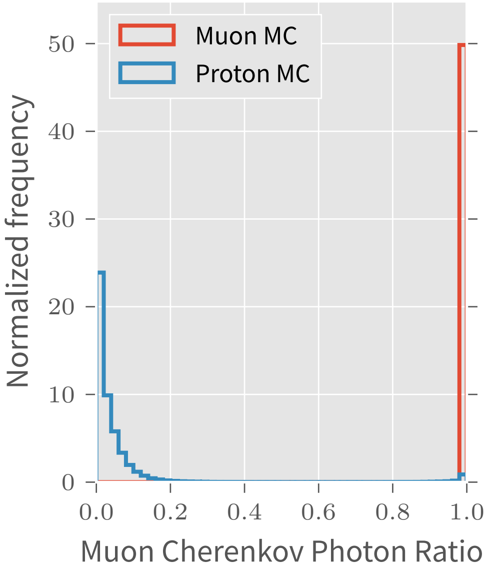
Energy range $100\,\mathrm{GeV}$ to $200\,\mathrm{TeV}$

Spectral index -2.7

Zenith distance $0° – 30°$

ViewCone $5°$ (mirror radius)

Discard events if more than $50\,\%$ of the Cherenkov photons were emitted by muons.



FACT-Tools

- → Joined forces with the Artificial Intelligence Group / Department for Computer Sciences
- → Extension of the `streams`-framework developed in Dortmund
- → Easy deployment (single `.jar`-file)
- → Rapid prototyping with Event-Viewer
- → Process is designed with `xml`

Current Calibration and feature extraction

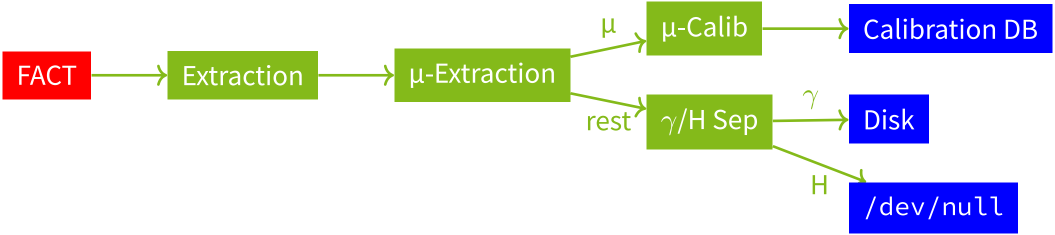
Planned Add application of trained model, proof of concept in:

Christian Bockermann et al. “Online Analysis of High-Volume Data Streams in Astroparticle Physics”. In: *Machine Learning: ECML 2015, Industrial Track*. Springer Berlin Heidelberg, 2015

`sfb876.tu-dortmund.de/FACT/`

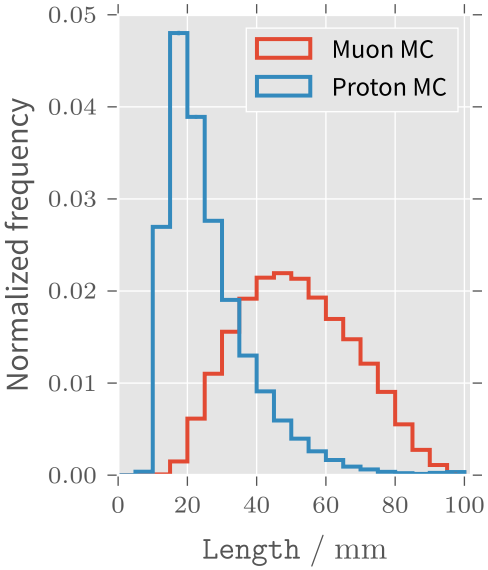
SFB 876 Providing Information
by Resource-Constrained Data Analysis

A Possible Data Stream



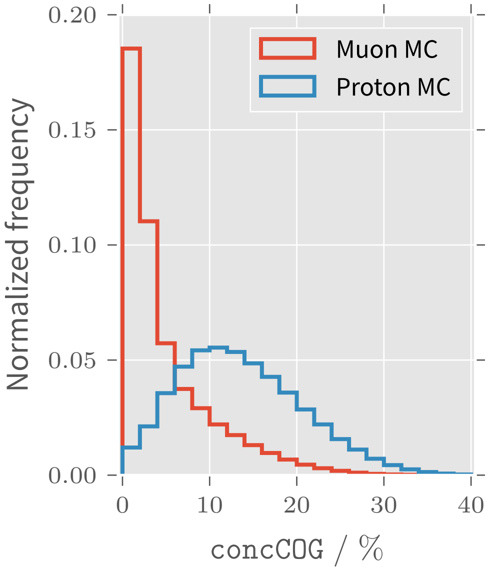
Feature Extraction

- Standard Hillas parameters also helpful for muon identification
- Timing features very discriminative
- Dedicated muon features improve the performance



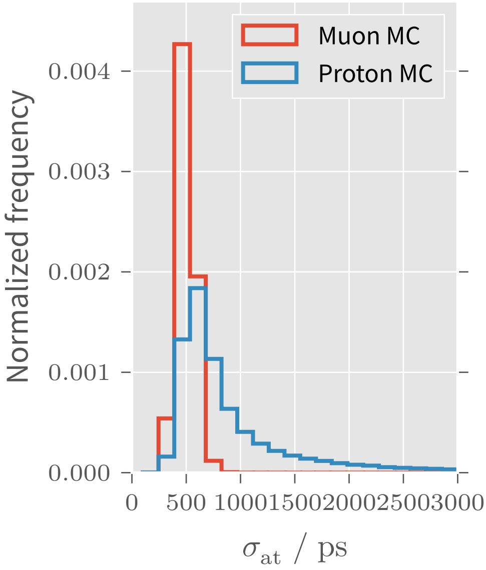
Feature Extraction

- → Standard Hillas parameters also helpful for muon identification
- → Timing features very discriminative
- → Dedicated muon features improve the performance



Feature Extraction

- → Standard Hillas parameters also helpful for muon identification
- → Timing features very discriminative
- → Dedicated muon features improve the performance



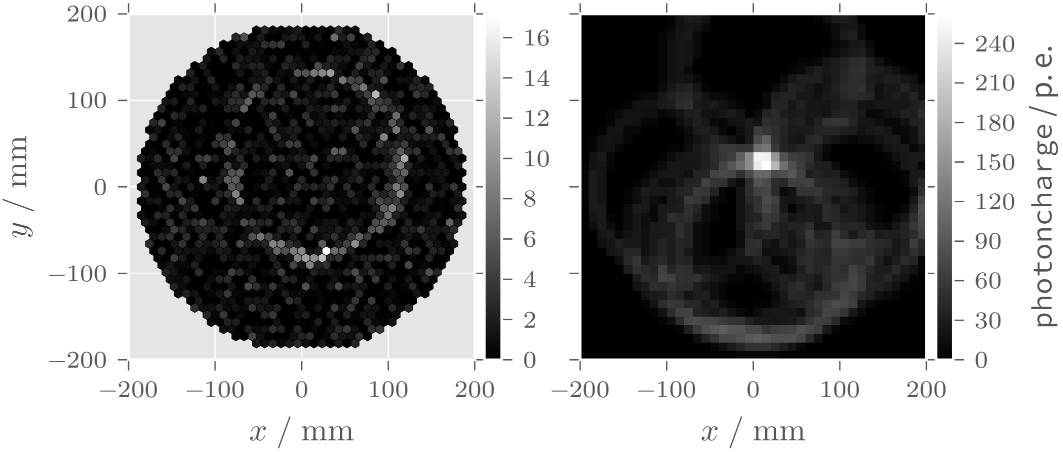
Hough Transform

- Shape recognition algorithm
- Transformation from image space into discrete parameter space

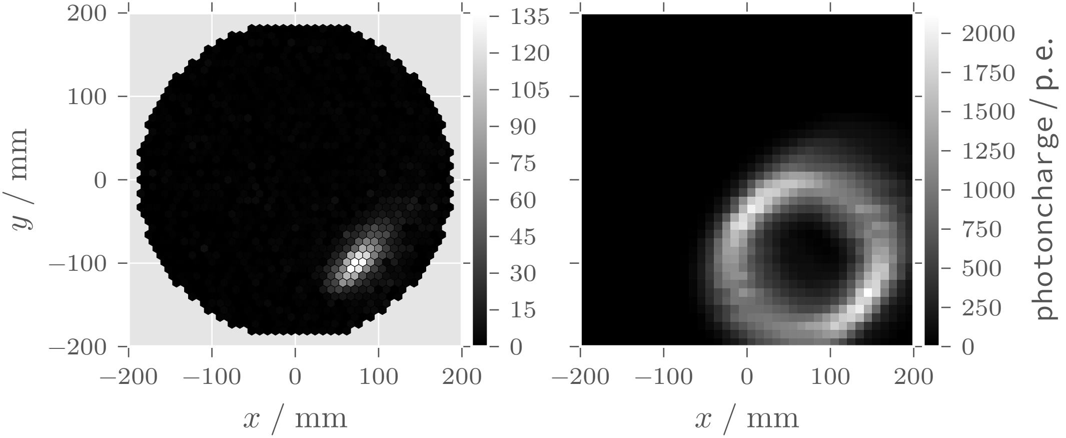
$$\text{Photons}(\text{pixel}) \to \text{Photons}(R, x_0, y_0) \tag{1}$$

- Pixels “vote” for rings
- Strong features can be calculated from shape of parameter space
- Computationally expensive

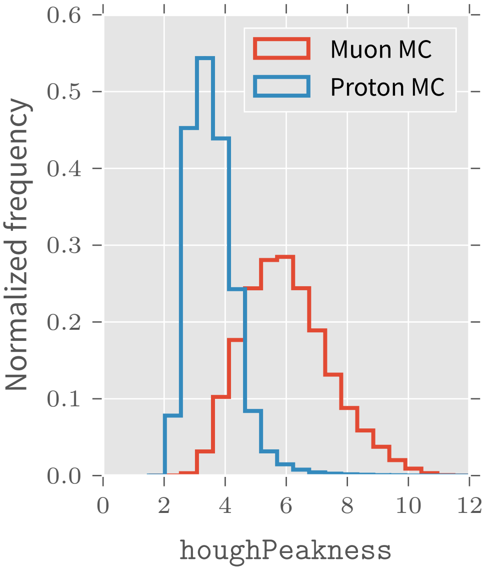
Hough Transform: Muon Ring Image

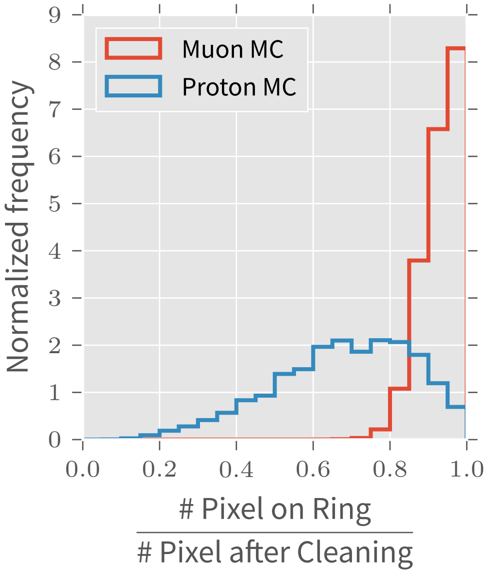


Hough Transform: Proton Shower



Hough Transform: Features





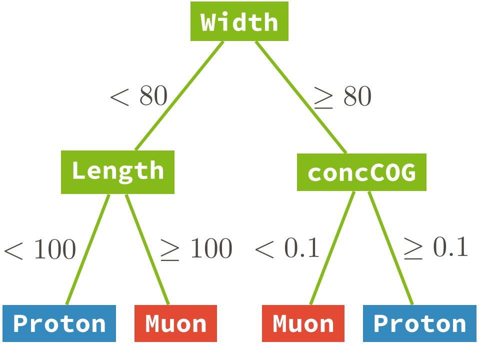
Machine Learning: Decision Trees

Training

- → Needs *labeled* data
- → Pick the most discriminative feature
- → Split at the best possible cut
- → Repeat until only pure data sets are left or a given *maximum depth* is reached

Application

- → Unknown event goes through the tree
- → Leaf determines classification



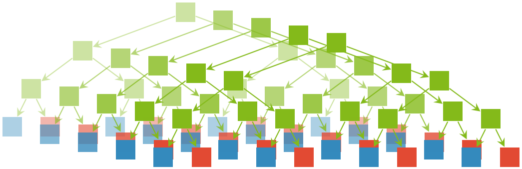
Machine Learning: Random Forests

Problems with single DTs

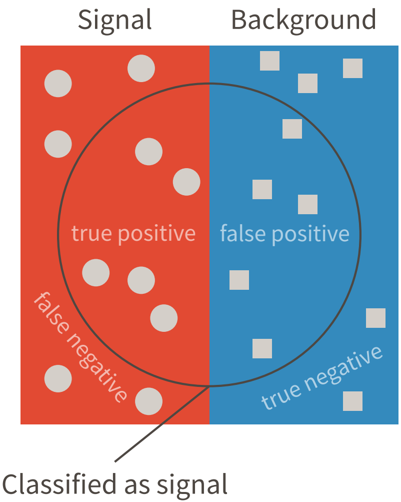
- → Tend to overtraining
- → Rather large reaction on small fluctuations in training data

Solution

- → Train many trees on slightly different training data (*Bagging*)
- → Consider only a random subset of features at each node
- → *Confidence:* proportion of the trees that voted for target class



Performance Evaluation



Precision

$\mathrm{tp} \,/\, (\mathrm{tp} + \mathrm{fp})$

→ also known as purity

Recall

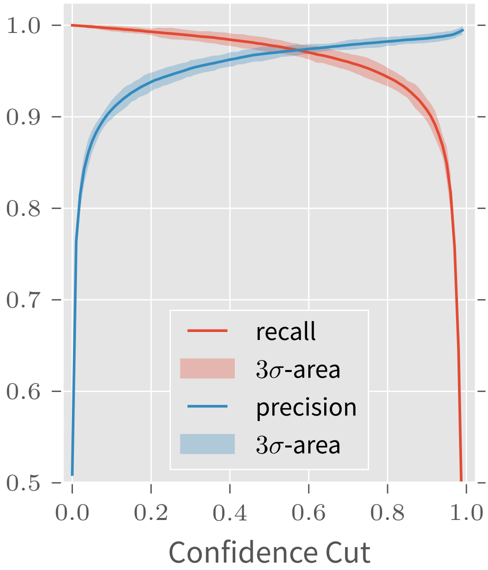
$\mathrm{tp} \,/\, (\mathrm{tp} + \mathrm{fn})$

→ also known as efficiency or true positive rate

False positive rate

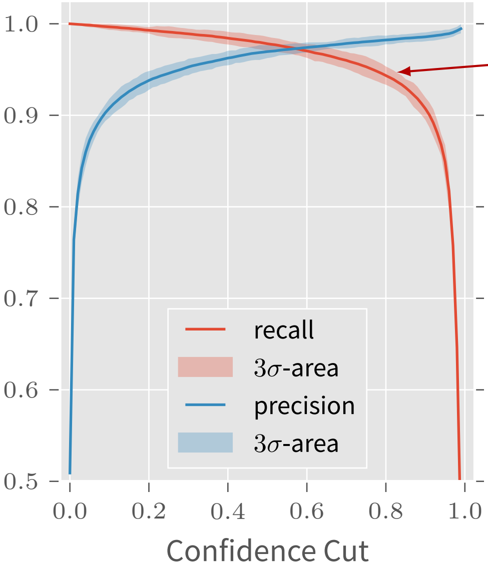
$\mathrm{fp} \,/\, (\mathrm{tn} + \mathrm{fp})$

Choosing a Confidence Cut



- → All events with `confidence` $>$ `cut` are classified as signal
- → Classification is trade-off between precision and recall
- → Which is more important for your task?

Choosing a Confidence Cut

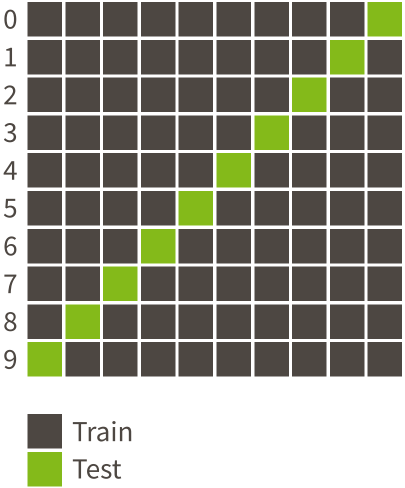


Where do the uncertainties come from?

- → All events with `confidence` > `cut` are classified as signal
- → Classification is trade-off between precision and recall
- → Which is more important for your task?

Cross Validation

- Performance values without uncertainties are meaningless
- Split training data in N subsamples
- Train on $N-1$, evaluate performance on the other part
- Calculate mean and standard deviation of performance values
- Large uncertainties implicate overfitting

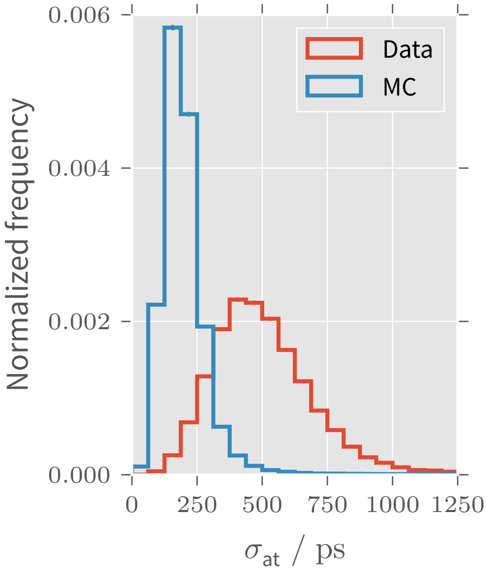


An Independent Channel

- → Extract Muons from telescope data and proton Monte-Carlo
- → Compare attribute distributions
- → Muons especially interesting for
  - → Time resolution
  - → PDE
  - → PSF

Timing

- Standard deviation of arrival times indicates time resolution
- Dependent on reconstruction and NSB levels
- Rather huge mismatch at FACT in the beginning

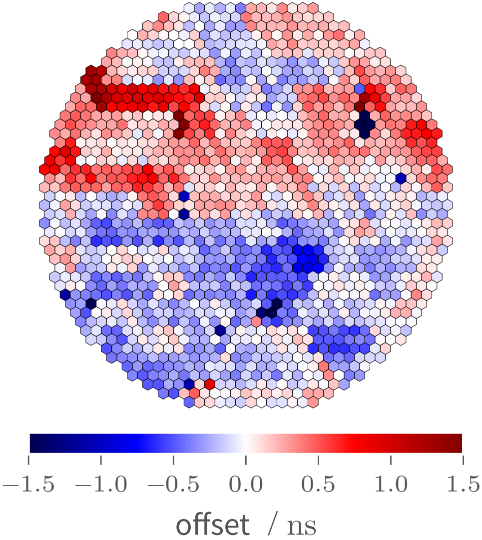


Timing

First idea Fixed offsets between channels

Measurement LED-Lightpulser Calibration runs

Result Better time resolution on data, but not close to MC

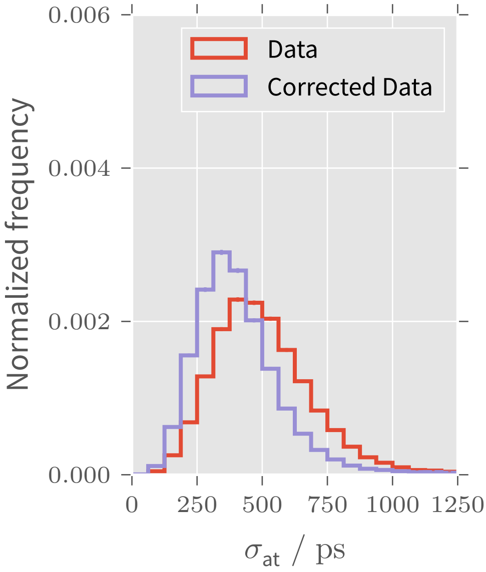


Timing

First idea Fixed offsets between channels

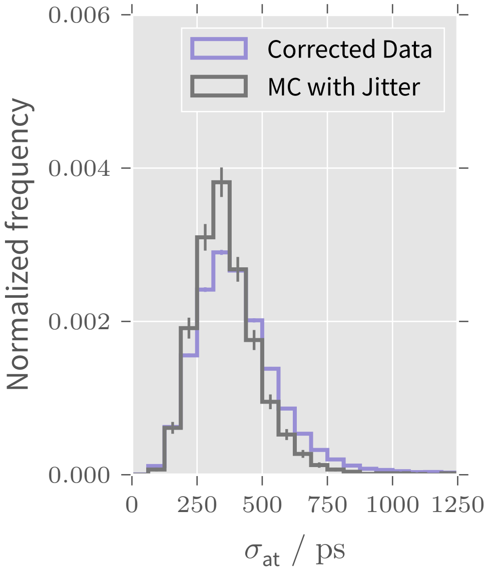
Measurement LED-Lightpulser Calibration runs

Result Better time resolution on data, but not close to MC



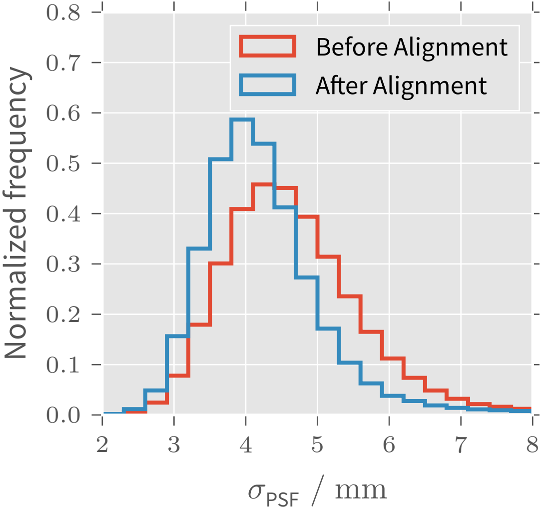
Timing

- Add additional time smearing to MC
- Possible explanations:
  - G-APD time jitter
  - Scattering in the Winston cones
  - Electronics
  - Residual of DRS time calibration
  - …



Point-Spread-Function

- → Mirror alignment in May 2014
- → Gaussian fit to radial light distribution
- → σ of the gaussian indicates PSF



Sebastian Müller et al. "FACT – Novel mirror alignment using Bokeh and enhancement of the VERITAS SCCAN alignment method". In: *Proceedings of the 34th ICRC*. 976. 2015

Conclusion

- → Machine Learning methods well suited for muon extraction
- → `streams`-framework allows online application of such methods
- → Pipe out muons for calibration from the live data stream
- → Important cross check for MC