



Machine Learning for Data Certification

Lukas Danev

11/08/2017

Big picture

CMS
experiment



Trigger



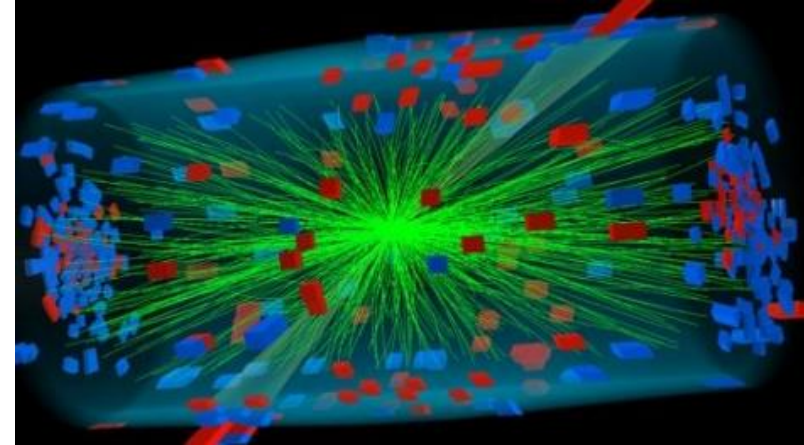
Data
reconstruction



Data
certification



Physics
analysis



What is Data Certification?

Data Quality Monitoring

Distinguishing good from bad data

Certified data → Good for physics analysis

Good ↔ the expected performance observed

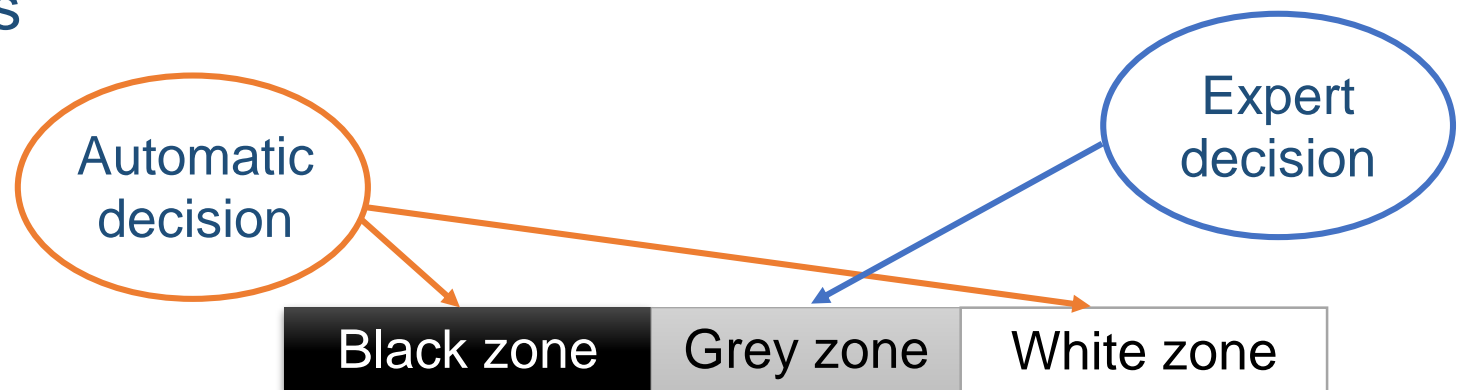
Automation of Certification

Aiming to aid human experts

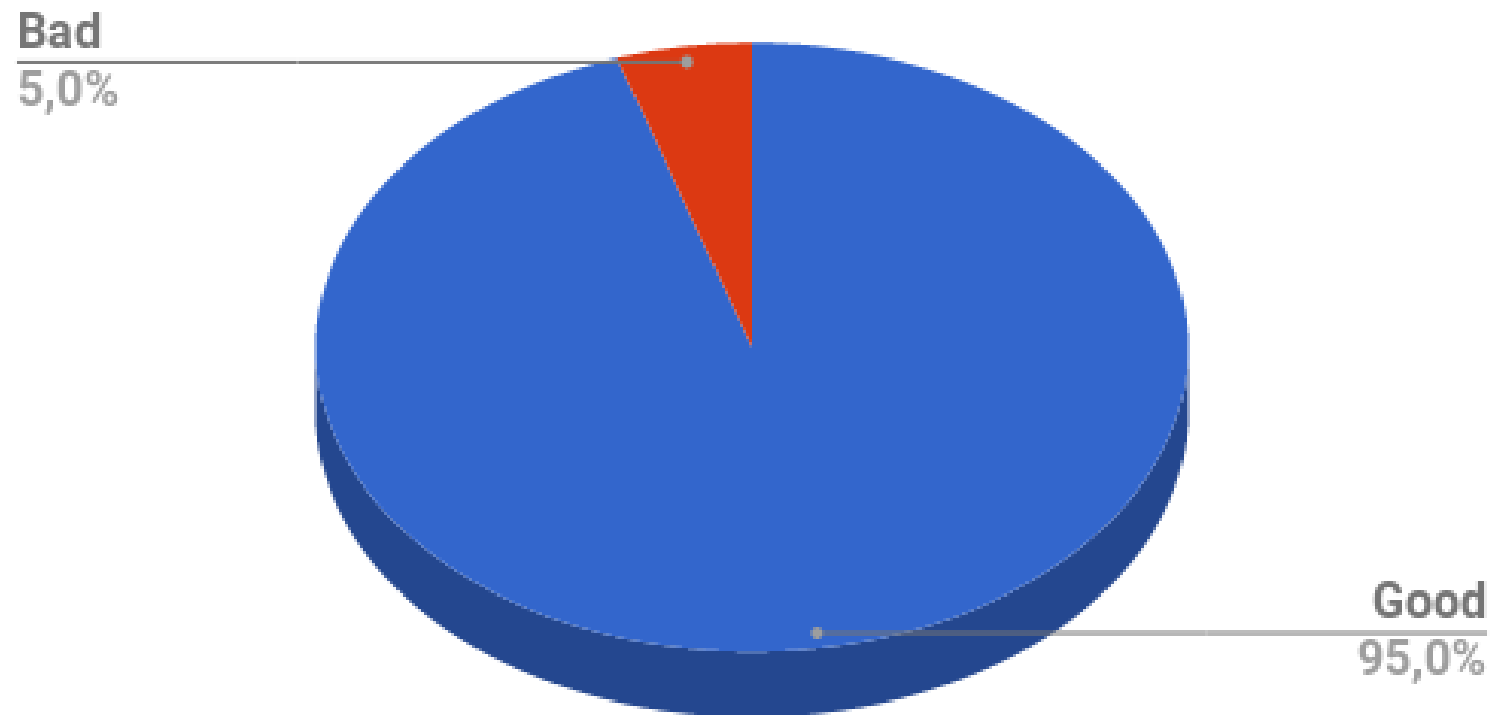
Expand the phase space of the monitored variable

Supervised learning

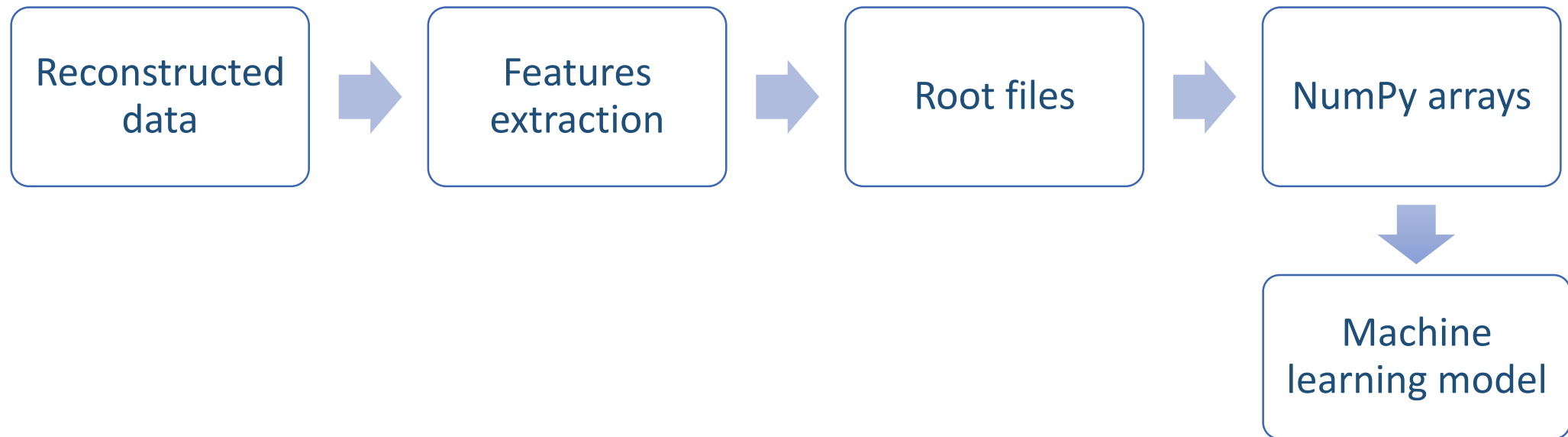
Semi-supervised techniques



Challenges



Data preprocessing



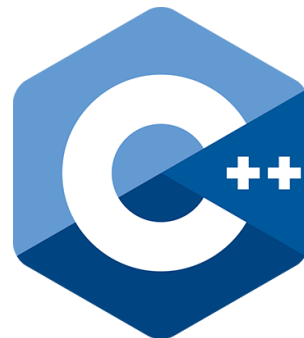
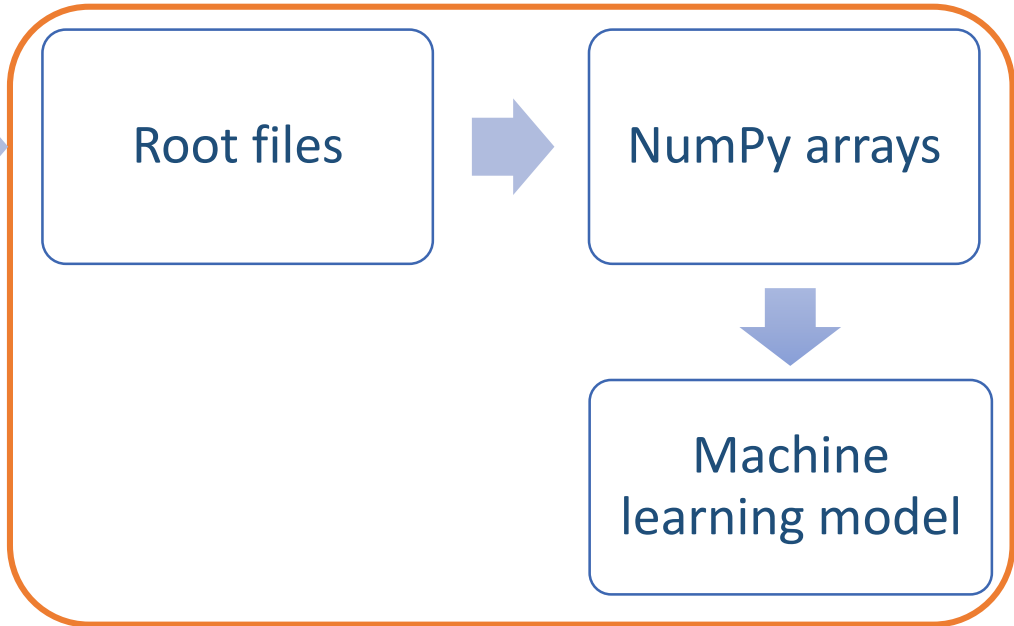
Data preprocessing



Reconstructed data



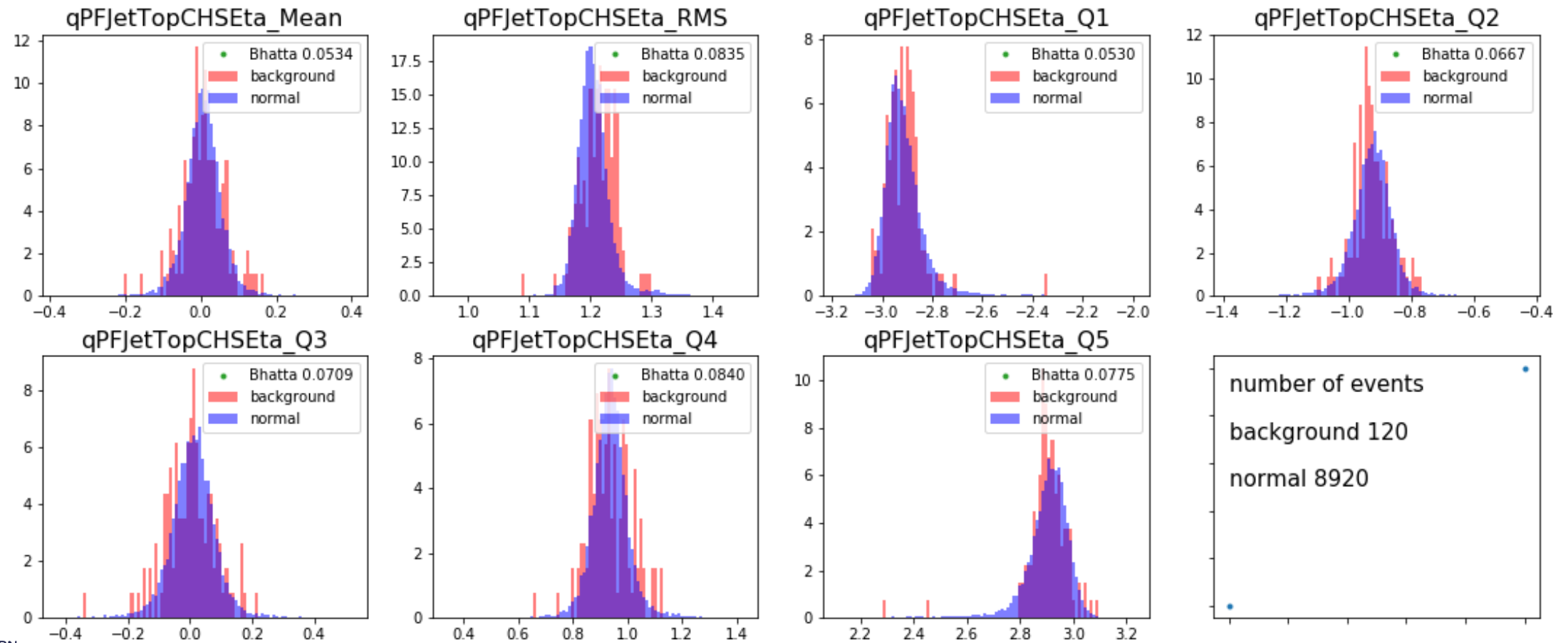
Features extraction



Feature examinations

Probably bad feature

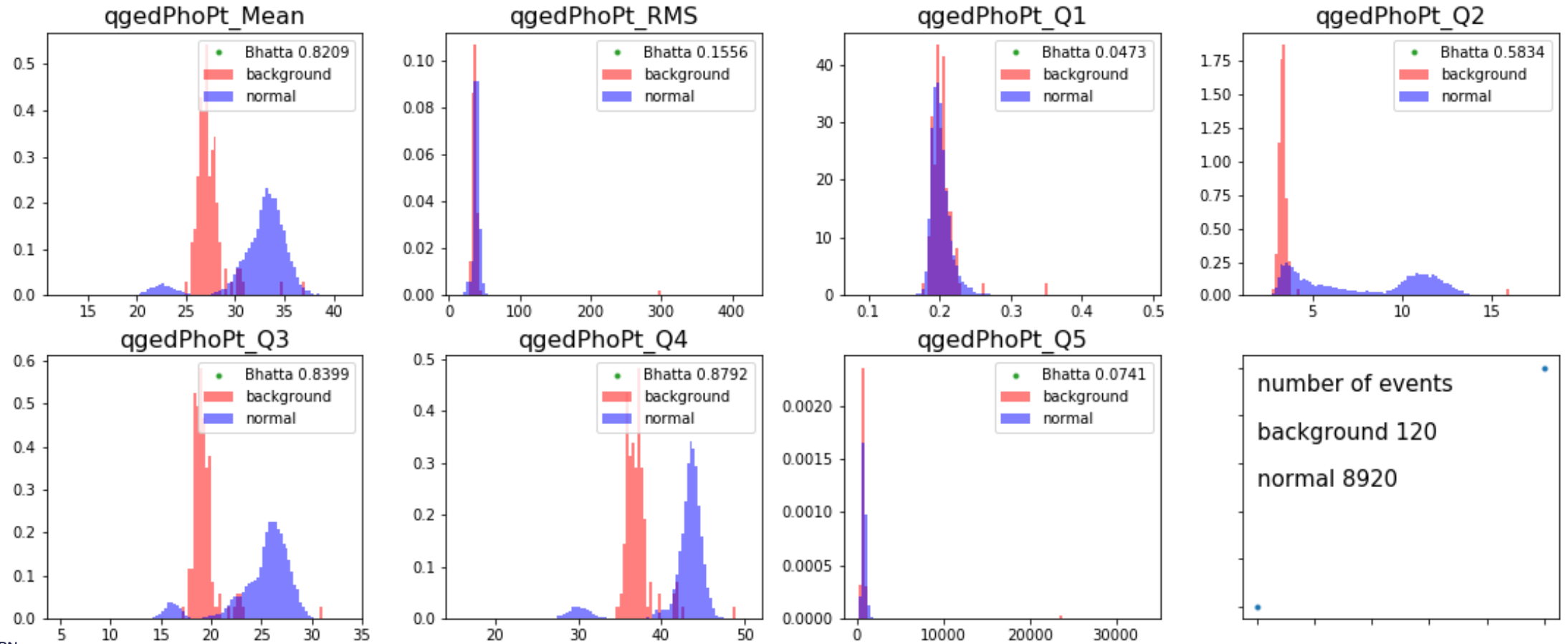
qPFJetTopCHSEta



Feature examinations

Probably good feature

qgedPhoPt



Acknowledgements

Supervisors

Giovanni Franzoni

Federico De Guio

Filip Široký

Sverre Herland

Yandex team

Acknowledgements

Supervisors

Giovanni Franzoni

Federico De Guio

Filip Široký

Sverre Herland

Yandex team



QUESTIONS?

Lukas Danev

lukas.danev@cern.ch

Feature examinations

Probably good feature without normalization

qgedPhoPt

