CMS open data: Analysis examples for external users and for validation, thoughts on physics analysis

Achim Geiser, DESY Hamburg,

Irene Dutta, DESY Hamburg and IISER Pune,

Reminder: Open data release

Open data portal: http://opendata.cern.ch/about/CMS

Proposals for open data benchmark and validation analyses

(see also presentation at CMS collaboration week)

New: Some thoughts on physics with open data:

"next-to-minimum-bias" physics analyses

Reminder: Open data release

Overview of data model: see talk K. Lassila-Perini at collaboration week in Physics/Offline/PPD/Computing/Trigger plenary
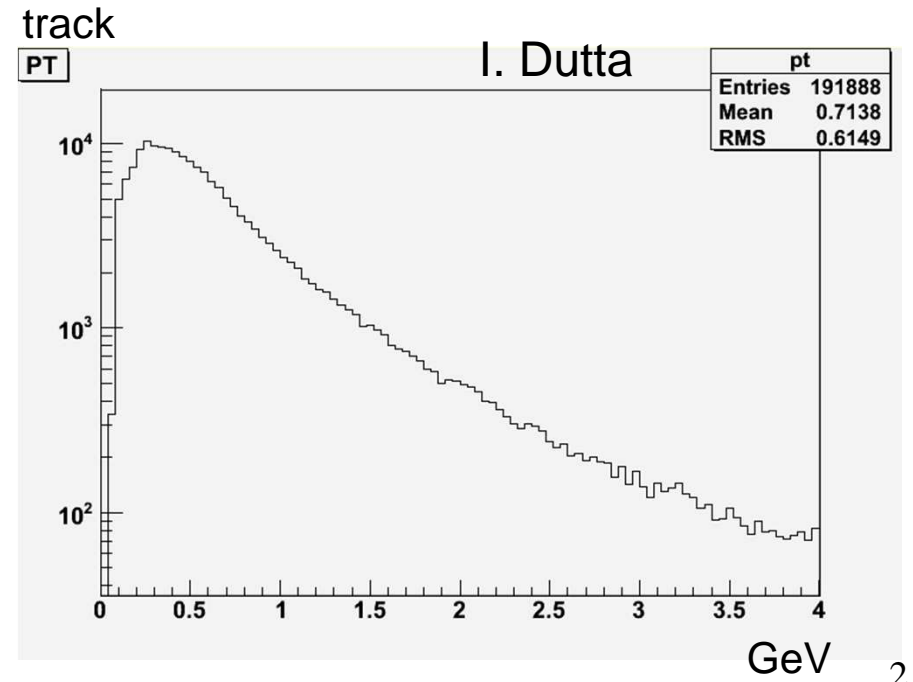
Open data release: http://opendata.cern.ch/about/CMS

- includes all primary data sets of run 2010B

- has basic documentation

- has tutorial to get started

- and it actually works!

    plot from 2010 AOD open data:

    (by Irene

    from scratch within ~ 2 days)

track

I. Dutta

pt	
Entries	191888
Mean	0.7138
RMS	0.6149

GeV

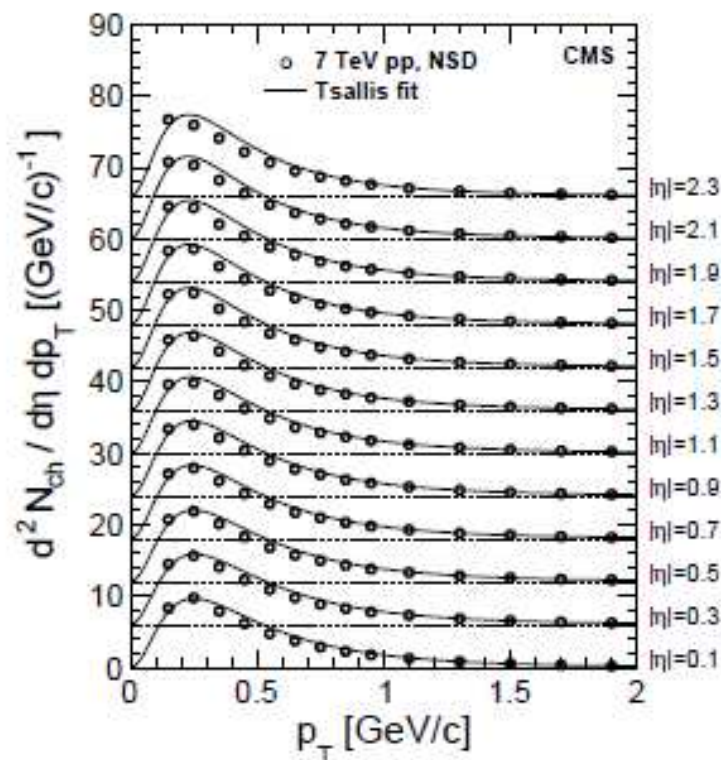# Proposal for open data benchmark analyses

Criteria:

- feasable on 7 TeV AOD data, with small luminosity

- possibility for comparison (later) with data at other center-of-mass energies

- not too complicated but nevertheless interesting physics objects

- basic studies feasable without MC

- published reference available

- so far at proposal (these slides) and first implementation stage (see slides I. Dutta)

  -> try to find out what works, and what not

Proposal for open data benchmark analyses

Tracking:

Phys.Rev.Lett. 105 (2010) 022002, CERN-PH-EP-2010-009, CMS-QCD-10-006

Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s}$ = 7 TeV

Figure 1: Differential yield of charged hadrons in the range $|\eta| < 2.4$ in 0.2-unit-wide bins of $|\eta|$ in NSD events. The solid curves represent fits of Eq. 1 to the data. The measurements with increasing η are successively shifted by six units along the vertical axis.

validation of MinBias data set

started ☺

challenges:

tracking acceptance without MC,

definition of "Minimum Bias"

4

Proposal for open data benchmark analyses

Muons and secondary vertices:

Eur.Phys.J. C71 (2011) 1575, CERN-PH-EP-2010-046, CMS-BPH-10-002

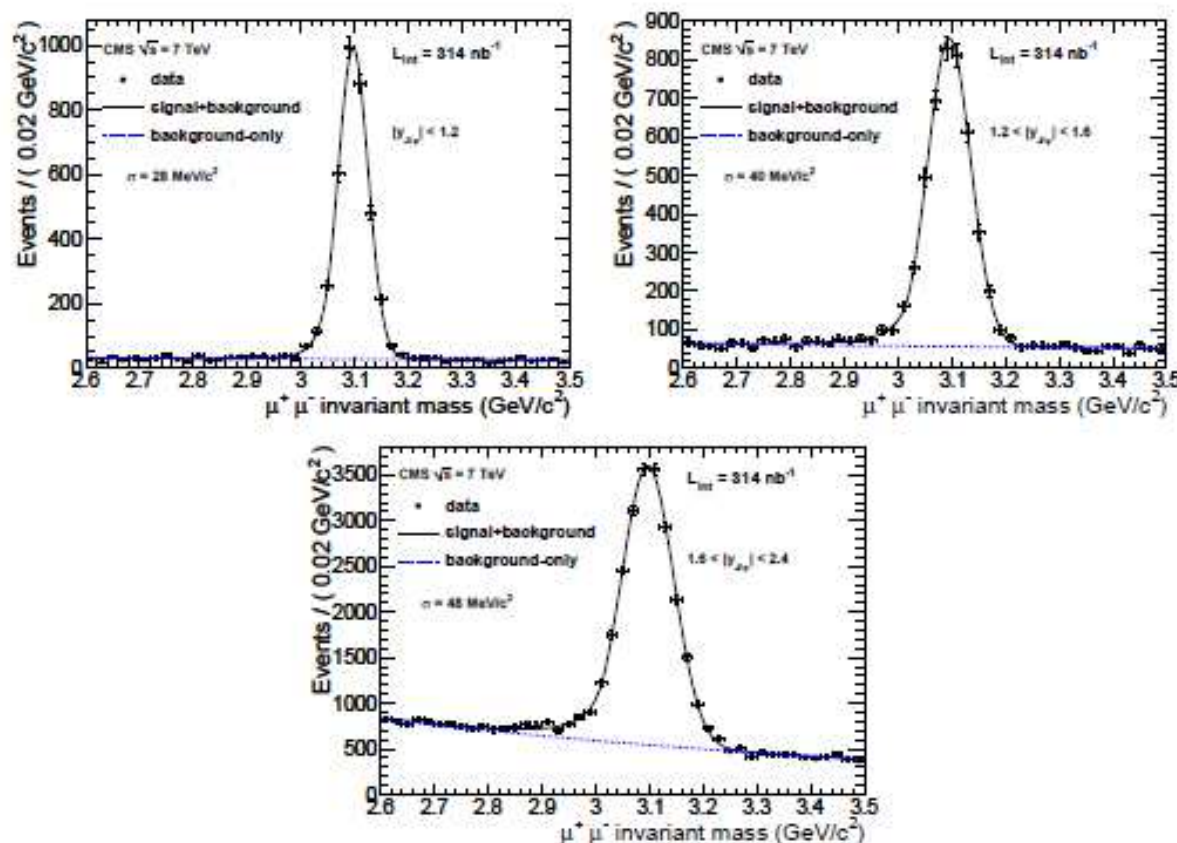Prompt and non-prompt J/ψ production in pp collisions at $\sqrt{s}$ = 7 TeV

validation of MuOnia data set

started ☺

challenges:

precise J/psi trigger and quality cuts

definition of "secondary vertex"

Figure 1: Opposite-sign dimuon invariant mass distributions in three J/ψ rapidity ranges, fitted with a Crystal Ball function plus an exponential (Section 5). The poorer dimuon mass resolution at forward rapidity is caused by the smaller lever arm of the muon tracks.

Proposal for open data benchmark analyses

Muons, electrons and missing ET:

JHEP 1101 (2011) 080, CERN-PH-EP-2010-050, CMS-EWK-10-002

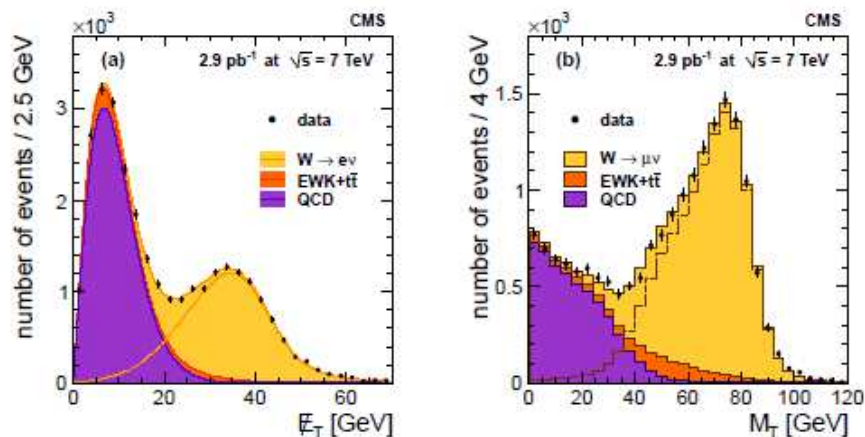Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV

Figure 1: The W signal distributions: (a) E_T distribution for the selected W → eν sample; (b) M_T distributions for the selected W → μν sample. The points represent the data. Superimposed are the results of the maximum likelihood fits for signal plus backgrounds, in yellow; all backgrounds, in orange; QCD backgrounds, in violet. The dashed lines represent the signal distributions.
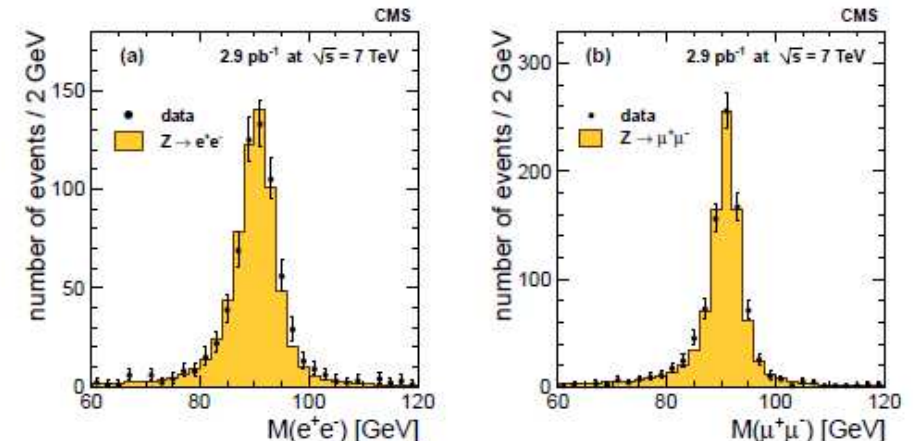
Figure 2: The Z signal distributions: (a) dielectron mass spectrum for the selected Z → e⁺e⁻ sample; (b) dimuon mass spectrum for the selected Z → μ⁺μ⁻ sample. The points represent the data and the histograms, the simulation. Backgrounds are negligible and are not represented in the plots.

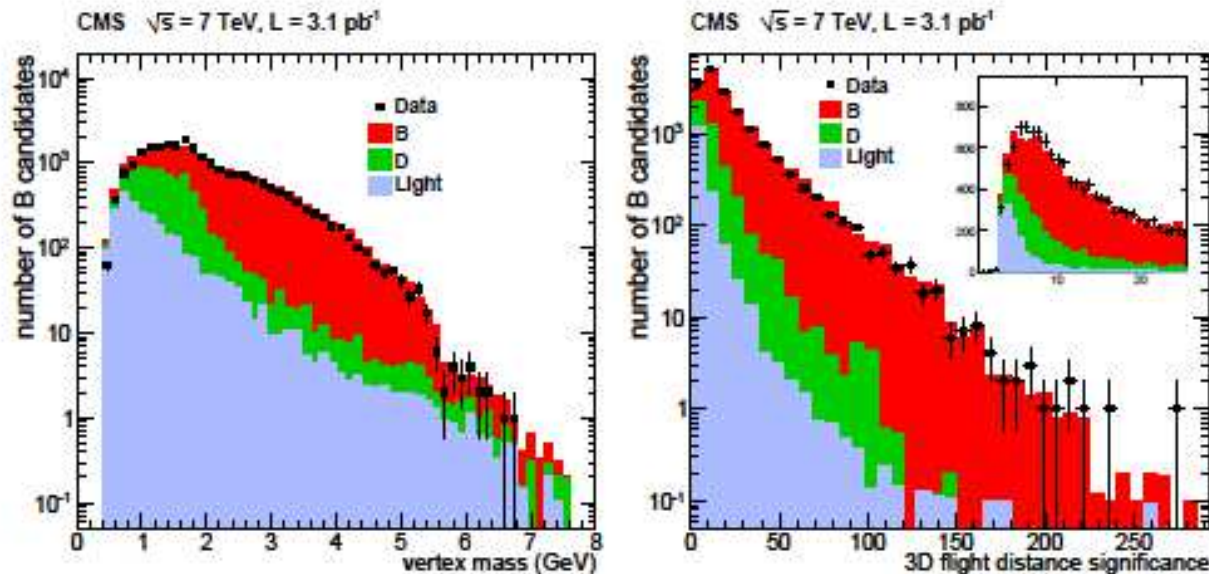validation of Mu and Electron data sets

Mu started ☺   (no results yet)

6

Proposal for open data benchmark analyses

Jets and B tagging:

JHEP 1103 (2011) 136, CERN-PH-EP-10-093, CMS-BPH-10-010

Measurement of BB Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s}$ = 7 TeV

not yet started

Figure 2: Properties of the reconstructed B candidates: vertex mass distribution (left) and flight distance significance distribution (right). The inset in the right plot shows a zoom of the flight distance significance distribution with narrower bins and linear scale. The data are shown by the solid points. The decomposition into the different sources, beauty, charm and light quarks, is shown for the PYTHIA Monte Carlo simulation. The simulated distributions are normalised to the total number of data events. All selection cuts apart from those on the shown quantities are applied.

Proposal for open data benchmark analyses

Jets:

Phys.Rev.Lett. 107 (2011) 132001, CERN-PH-EP-2011-053, CMS-QCD-10-011

Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s}$ = 7 TeV
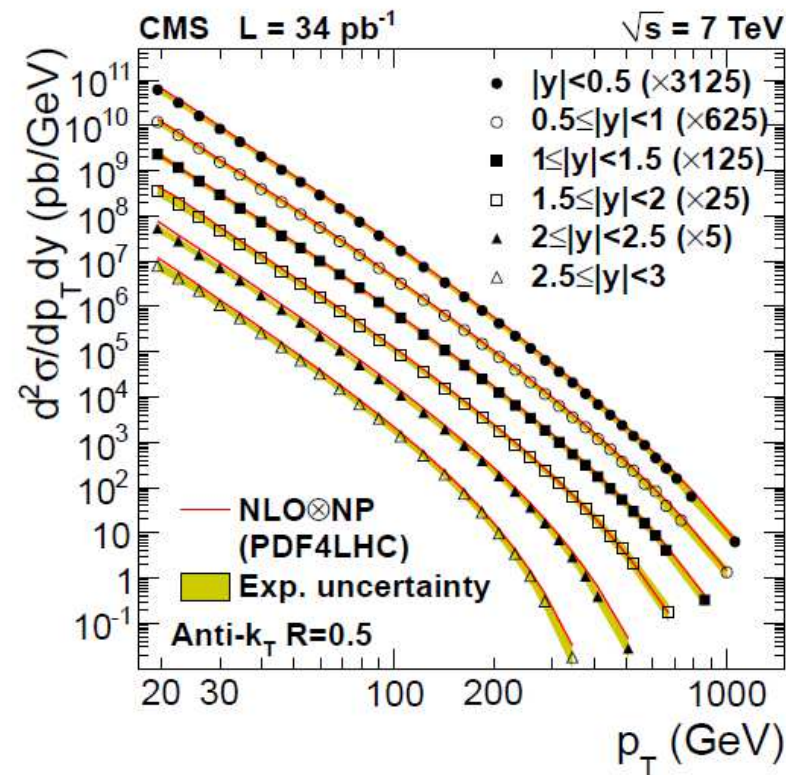

Figure 1: Fully-corrected inclusive jet differential cross sections as a function of p_T for six different rapidity intervals, scaled by the factors shown in the legend for easier viewing. The next-to-leading-order (NLO) theoretical predictions, corrected for non-perturbative (NP) effects via multiplicative factors, are superimposed.

Cross sections
-> need MC
-> not feasable?
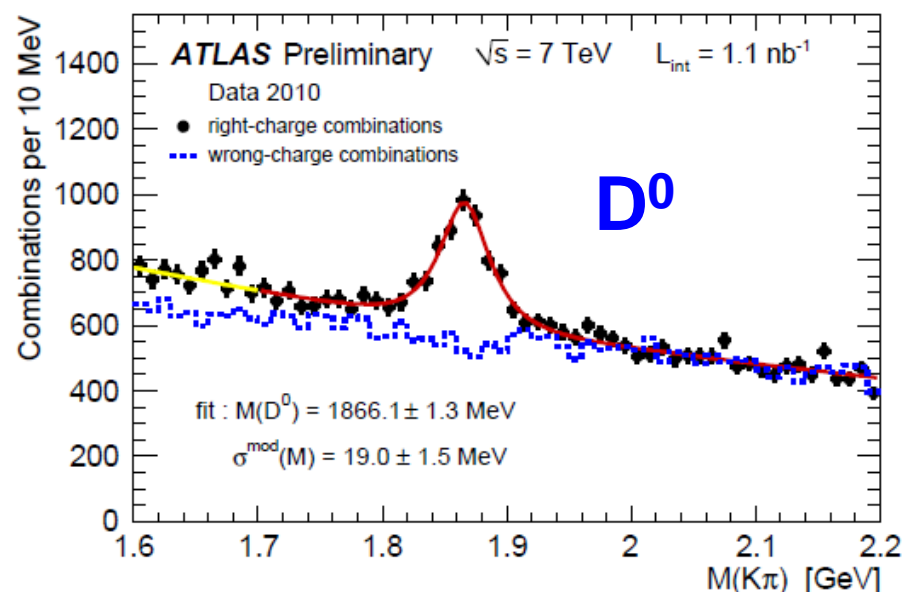
validation of Jet
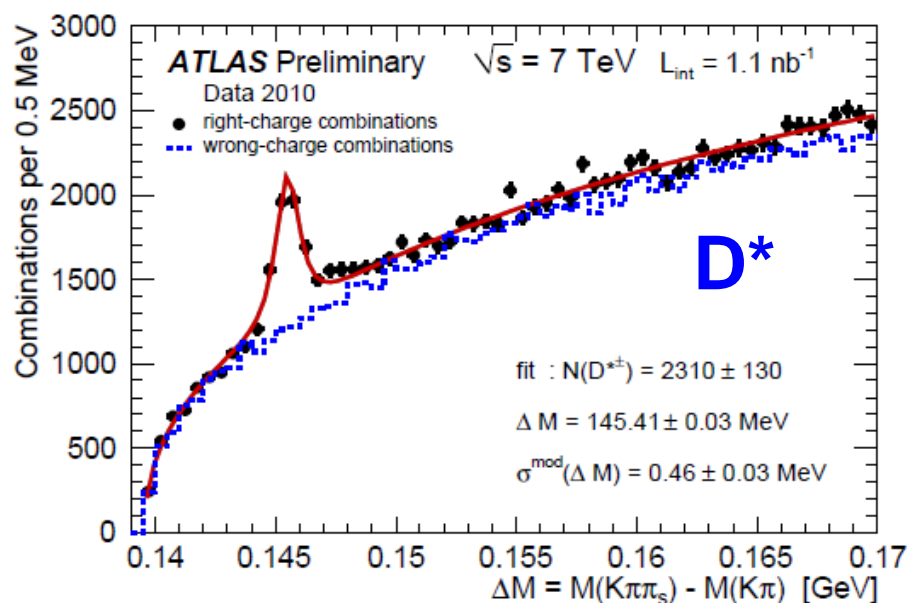data set

not yet started

8

Proposal for example real physics analysis on CMS open data

D mesons:

ATLAS preliminary (no corresponding CMS measurements so far)

Example for "next-to-minimum bias" project (see further slides)

not started yet

Concept of "next-to-minimum-bias" data

Standard "Minimum bias" data: data triggered by forward energy deposits on both sides of detector: "non-single-diffractive" cross sections.

New(?) concept (my terminology):

"Next-to-minimum-bias" (NMB) data:

Interactions with at least one (2?,3?) tracks from common vertex in central rapidity region, $|\eta|<2.5$, covering ~40-50% of the relevant η range (more forward part could be covered by LHCb).

Purely tracking-based analysis (do not use calorimeter information).

Practical implementation:

- Use any trigger on well-defined final state, with pileup.

- Remove the vertex corresponding to the triggered final state from the event.

- Treat the remaining pileup vertices as "next-to-minimum bias" (NMB) data set.

Considerations on "next-to-minimum-bias" data

Possible "next-to-minimum-bias" statistics/year (10^7 s) for 200 Hz trigger rate and pileup of 50:

N = 10^7 s x 200 Hz x 50 ~ 10^{11} events/year on tape (for free!)

Total pp cross section is σ~90 mb, fraction with activity in central region a~50%(?)

-> equivalent luminosity L ~ N/(aσ) ~ 2 pb^{-1}/year

Actual situation for 2nd half of 2010 AOD data: (typical pileup ~2.5 – 1 -> 1.5)

N = 5 x 10^6 s (?) x 40 Hz (?) x 1.5 ~ 3 x 10^8 NMB events,   L ~ 6 nb^{-1} (tbc)

(For comparison: ZEUS HERA II data have ~ 5 x 10^7 "useful" inclusive DIS events)

Enough luminosity for many interesting minimum bias-like QCD studies !

Check on 2010 open access data !

Conclusions

Open data release of part of 2010 data is available on http://opendata.cern.ch/about/CMS and works

Work on benchmark analyses on AOD for external users and for validation started (see slides I. Dutta)

Interesting possibilities e.g. for "next-to-minimum bias" tracking-based physics studies (see slides I. Dutta for very first look)