

Yüksek Enerji Fiziğinde Veri Analiz Yöntemleri

Sezen SEKMEN

Kyungpook National University

LHC CMS Deneyi

ssekmen@cern.ch

Veri analizi nedir?

Veri analizi bizi deney çıktısından deney sonucuna oluşturan işlemler bütünüdür.

Deney sonucu deneyden gelen ölçüm çıktısıdır, örneğin olay sayısı ya da fiziksel bir niceliğin ölçümü gibi.

Bu derste veri analizi kavramlarını **LHC analizi** üzerinden tanıyacağız.

LHC verisi **farklı fizik sorularına cevap aramak** amacıyla yüzlerce farklı şekilde analiz edilmektedir. Örneğin:

- Bildiğimiz **standart model (SM) parçacıklar ile ilgili ölçümler yapmak** (kütle, toplam tesir kesiti, diferansiyel tesir kesiti dağılımı, vb.)
- SM ölçümleri aracılığıyla veride **SM ile uyumsuzluk** aramak.
- Doğrudan **SM ötesi parçacıkları ya da SM ötesi fiziğin diğer izlerini** aramak.

Veri analizinin adımları (dersimizin içeriği)

- Sinyali tanımak ve arama yöntemini tasarlamak
- Tetikleyicileri tasarlayıp kullanılacak veriyi belirlemek
- Nesneleri yapılandırmak ve tanımlamak
- Sinyali ardalandan ayıran olay seçimi
- Ardalan kestirimi
- Sistemik belirsizlikler
- İstatistik analiz
- Deney sonuçları
- Yorumlama

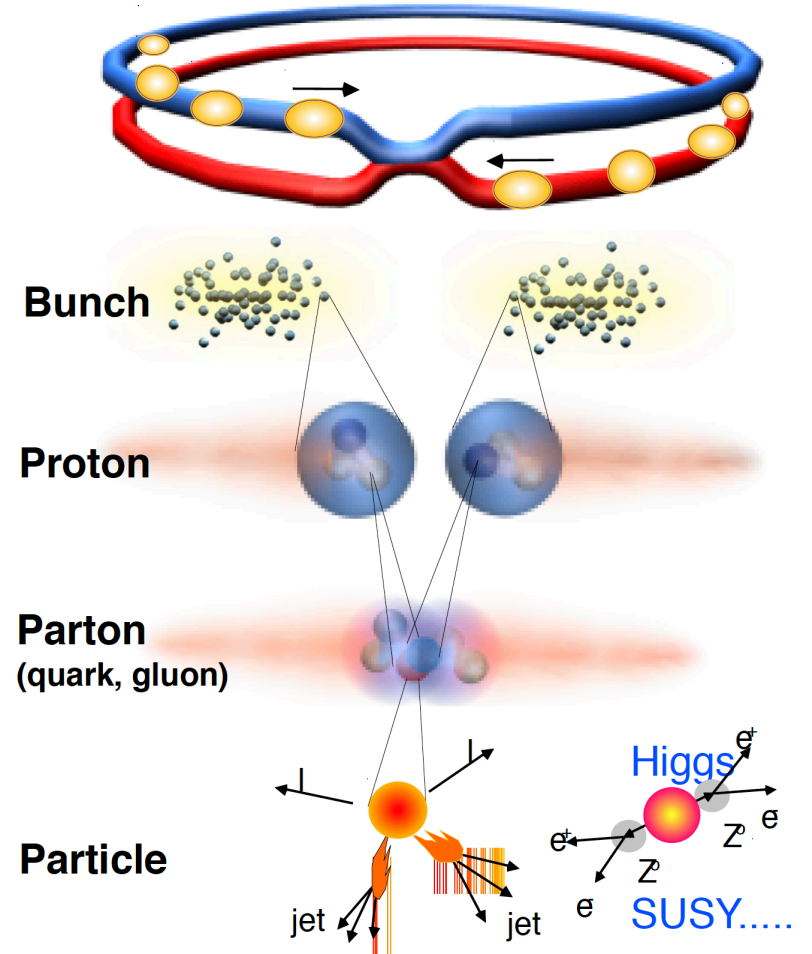
Çarpışma verisi nedir?

Çarpışma verisi parçacık çarpışma olaylarını içerir.

Olay (event), bir temel parçacık etkileşimine dahil olan tüm parçacıklara dair bilgilerin ve özelliklerin toplamıdır.

Analizde iki tür olay kullanırız:

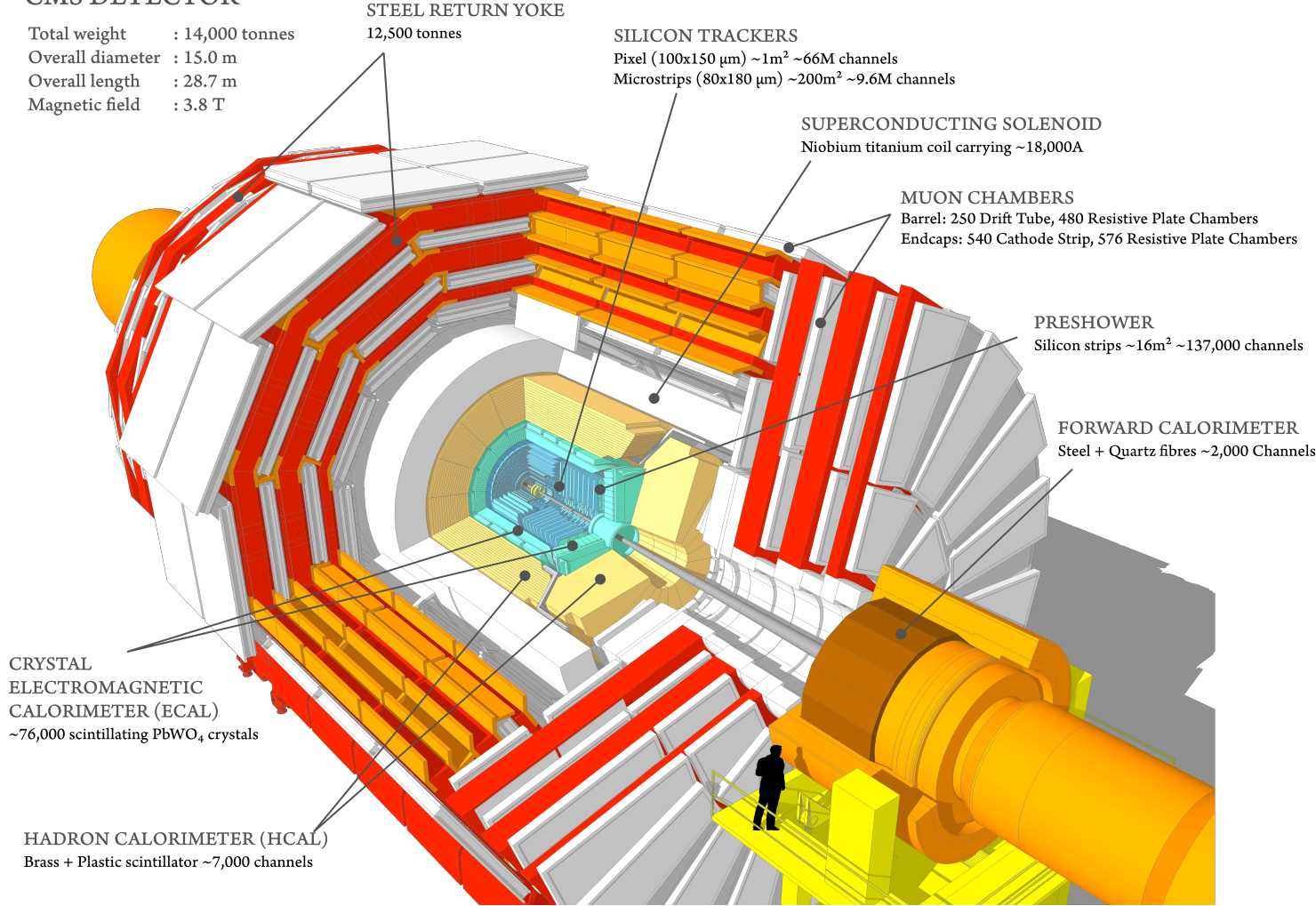
- **Gerçek çarpışma olayları:** Deneyde ortaya çıkmış ve algıç tarafından toplanmış olaylar.
- **Simülasyon çarpışma olayları:** Gerçek olayların bilgisayar modelleri. İki aşamada üretilir:
 - Fizik süreci modellemesi
 - Algıç etkisi modellemesi.



Algıç ile veri kaydetme

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T



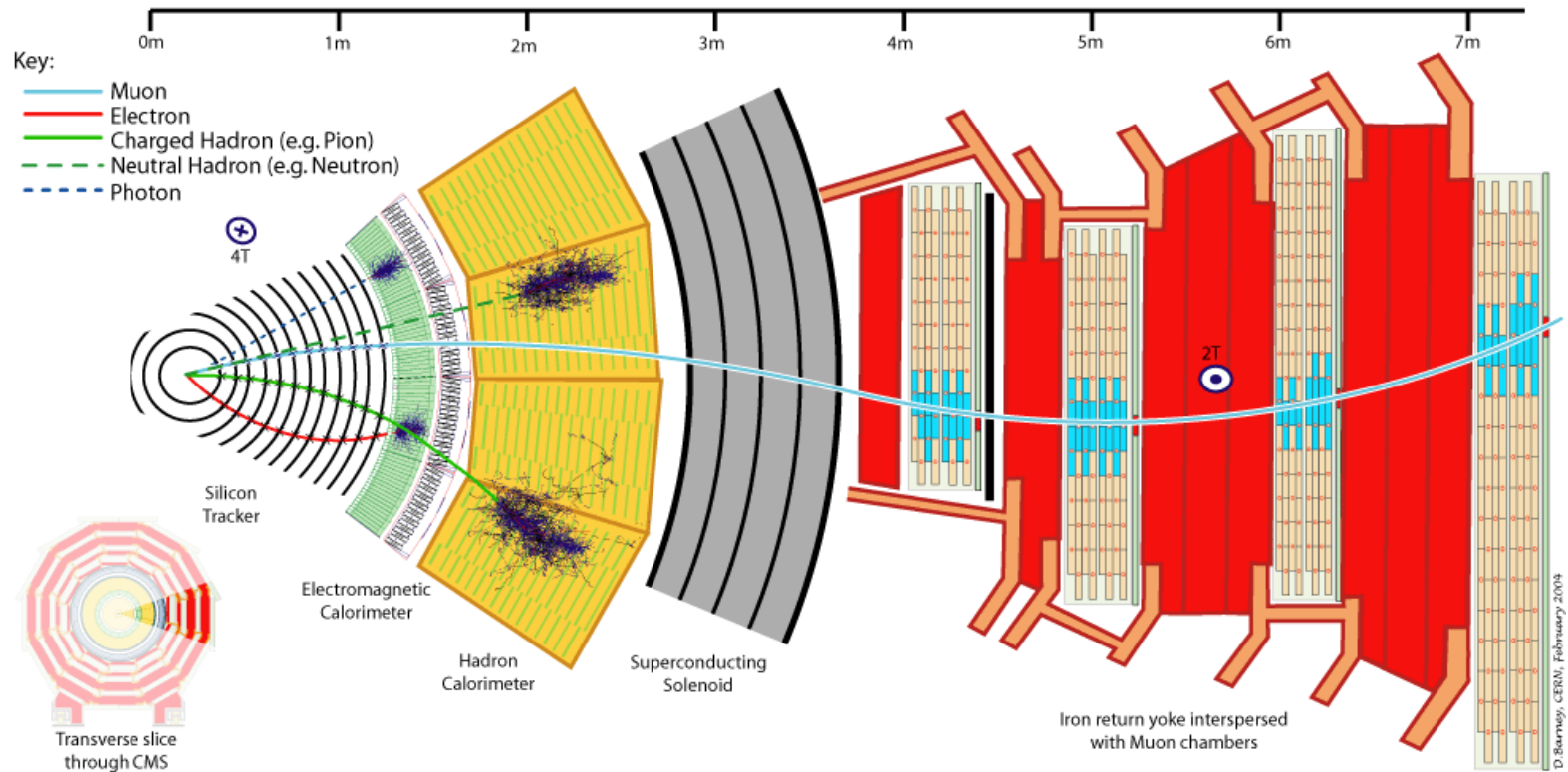
Çarpışmada oluşan parçacıkların algıç ile etkileşince enerji birikintileri kaydedilir.

Nesneler ve veri yapılandırma

Etkileşim sonrasında ortaya çıkan parçacıkların algıçta görüntülenen hallerine **nesneler** denir: **jet, b-kuark jet, e, μ , τ , γ , kayıp dikey enerji (MET),**

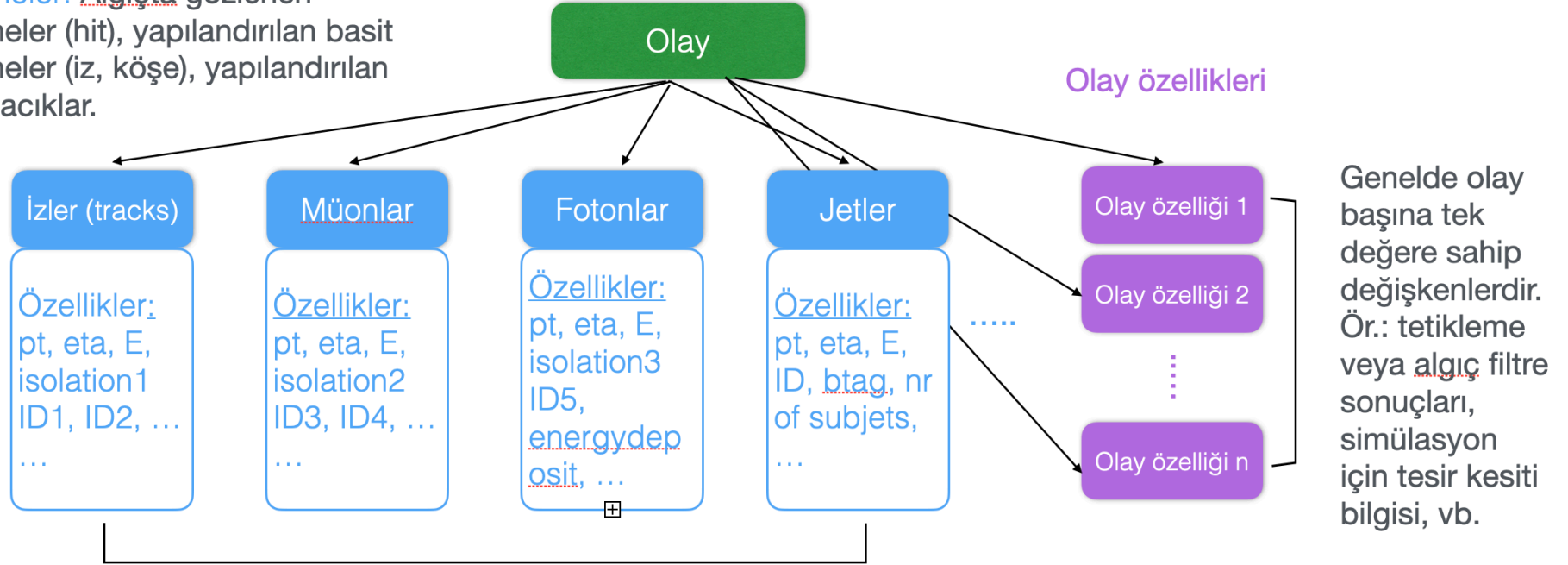
Nesneler alt algıçlardan toplanan bilgileri, ör. iz sürücü ve muon odasındaki vurmaları, kalorimetrelerdeki enerji birikintileri, **yapılandırarak** oluşturulur.

Yapılandırma işlemi hem gerçek hem simülasyon verisine uygulanır.



Çarpışma olay verisi neye benzer?

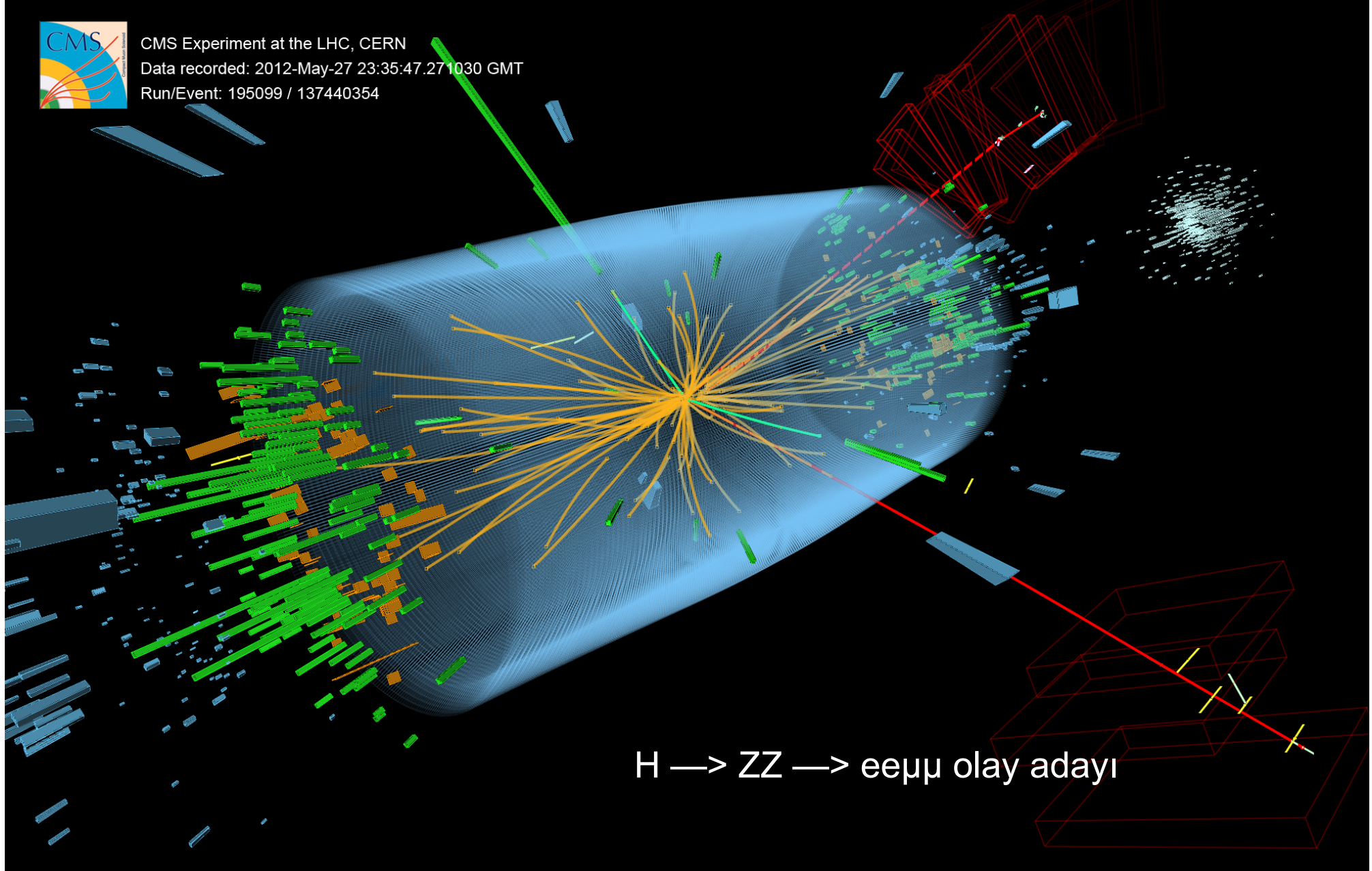
Nesneler: Algıda gözlenen nesneler (hit), yapılandırılan basit nesneler (iz, köşe), yapılandırılan parçacıklar.



Nesneler 4 vektör bilgisi + diğer özellik bilgilerini taşırlar.
Bir olayda her parçacıktan farklı miktarda olabilir.
Ör: Bir olayda 2 elektron, 1 muon, 0 foton ve 12 jet olabilir.



CMS Experiment at the LHC, CERN
Data recorded: 2012-May-27 23:35:47.271030 GMT
Run/Event: 195099 / 137440354

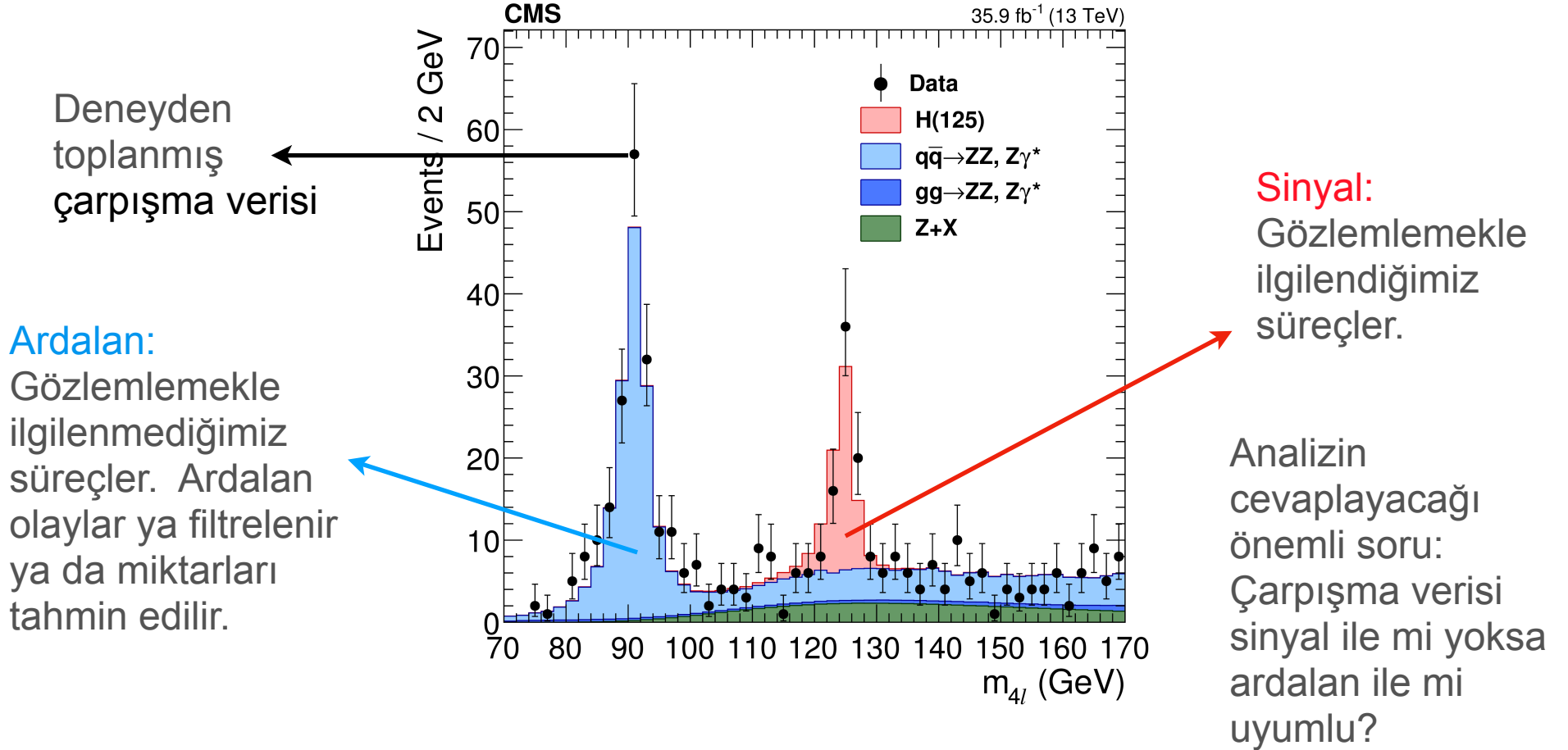


$H \rightarrow ZZ \rightarrow ee\mu\mu$ olay adayı

Pek çok olay: gerçek veri, sinyal ve ardalan

Bir tek olaya bakarak çıkarım yapamayız.

Pek çok olayı bir araya toplayıp, ayırıcı değişkenlerin dağılımına bakmalıyız:



Sinyali tanımak

Uğraşmaya değer mi?

- Fizik modelimizi bir **hesaplayıcı** içine yerleştirelim
- Deney süresince **kaç olay** bekliyoruz?
- Sinyal **hangi izi** bırakıyor? **Son durum parçacıkları** neler?
- Son durum parçacıkları algıçta **hangi nesnelere** denk geliyor?
- Benzer son durumlu **ardalanlar** neler?

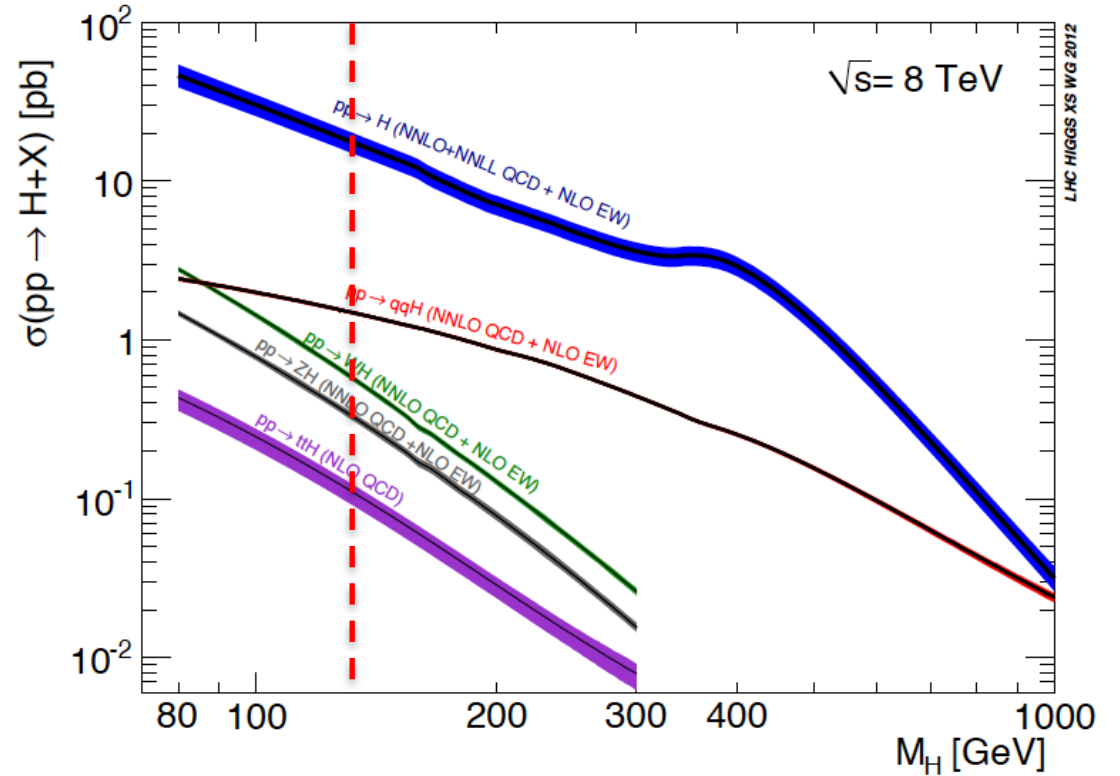
Sinyal ve ardalan için simülasyon olayları üretip inceleyelim:

- Olay üretip **algıç benzetimi** yapalım.
- Sinyali “kolayca” ardalandan nasıl ayırt ederim? Bunun için sinyal özelliklerini öne çıkaracak bir **olay seçimi** belirleyelim.
- Olay seçiminden sonra **sinyal hala “görülüyor” mu?**

Kaç olay bekliyoruz?: Tesir kesiti

Belli bir parçacık etkileşiminin hangi olasılıkla meydana geleceğini gösteren niceliğe **tesir kesiti** (σ) denir. Tesir kesiti etkileşimde yer alan parçacıkların cinsine, özelliklerine ve etkileşim enerjisine bağlıdır, ve alan birimleri ile ölçülür (**barn** ($10^{-28}m^2$), **picobarn**, **femtobarn**).

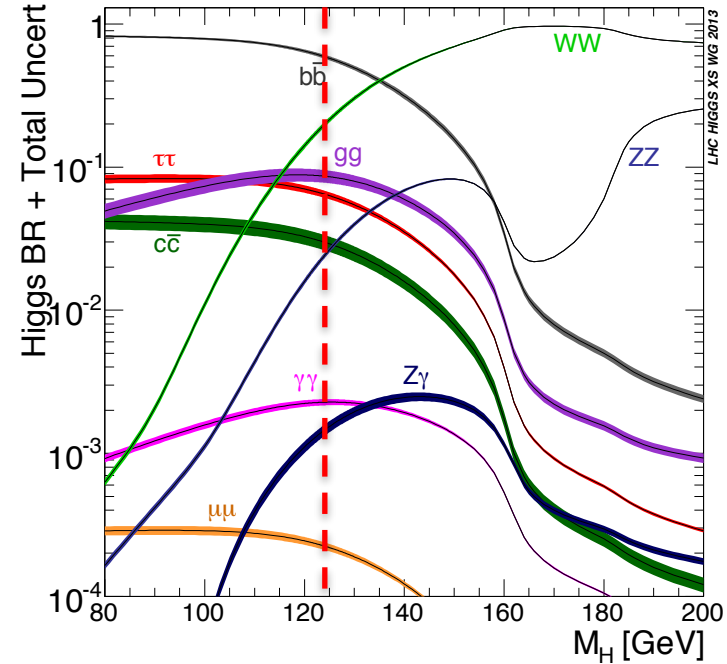
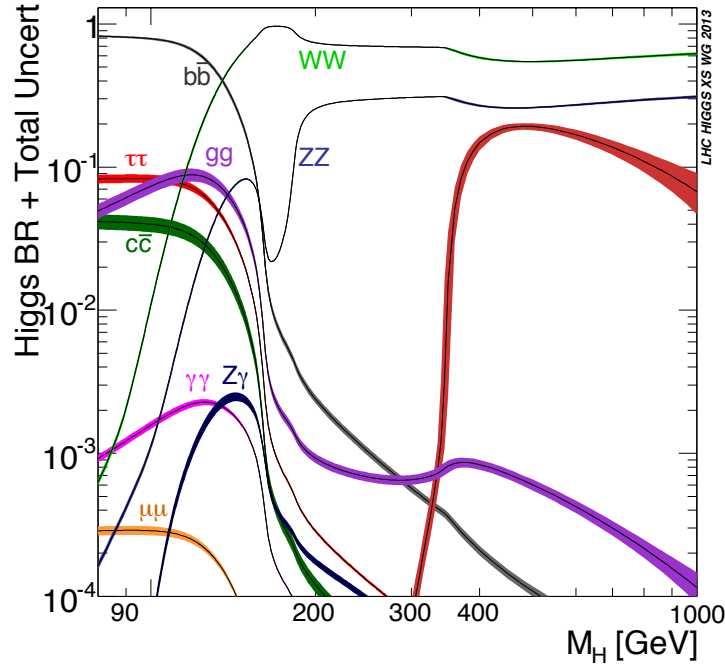
Ör: Farklı Higgs bosonu üretim mekanizmaları için Higgs tesir kesitlerinin Higgs kütlesine göre değişimi.



Kaç olay bekliyoruz?: Dallanma oranı

Pek çok parçacığın ömrü çok kısadır ve **kendinden hafif parçacıklara bozunur**.

Dallanma oranı (DO), bir parçacığın belli bir parçacık grubuna **bozunma olasılığıdır**. Tüm bozunma kanalları için DOların toplamı 1'dir.

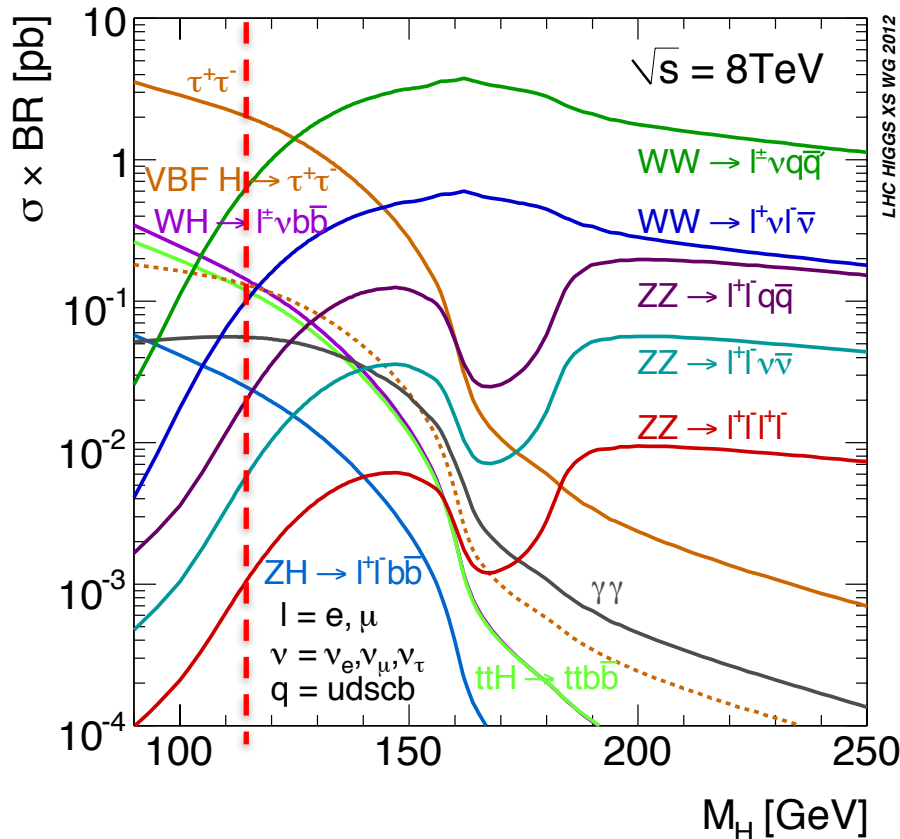


Ör: Higgs bosonunun farklı kanallardaki dallanma oranlarının Higgs kütleğine karşı değişimi.

Kaç olay bekliyoruz?: Etkin tesir kesiti

Belli bir üretim ve bozunma kanalındaki etkin tesir kesiti σ_e :

$$\sigma_e = \sigma \times \text{DO}$$



Ör: Higgs bosonunun farklı üretim ve bozunum kanallardaki etkin tesir kesitinin Higgs kütesine karşı değişimi.

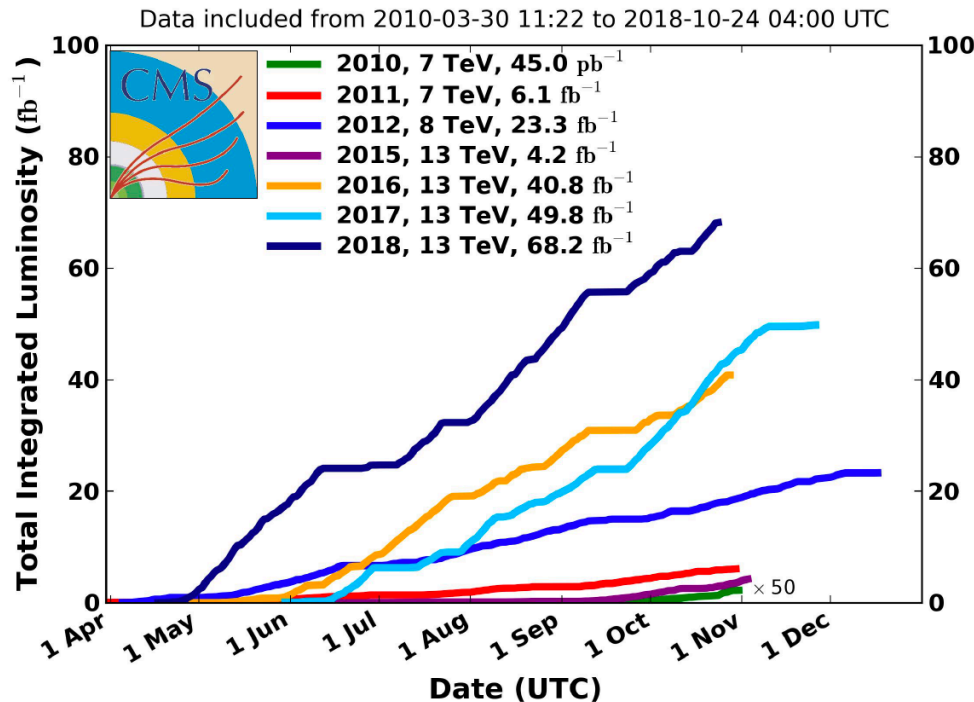
Yüksek $\sigma \times \text{DO}$ o üretim ve bozunma kanalında yüksek miktarda Higgs oluşacağına işaret eder.

Bu grafik hangi kanalda analiz yapılacağına karar vermede önemlidir.

Kaç olay bekliyoruz?: Işınlık ve toplam olay sayısı

- **Işınlık (luminosity, L)**, saniyede 1 cm^2 'de kaç pp çarpışması olduğunun ölçümüdür, $\text{cm}^{-2}\text{s}^{-1}$ birimi ile ölçülür. LHCnin ışınlığı 2012 yılı sonunda $10^{33} \text{ cm}^{-2}\text{s}^{-1}$ 'ye ulaşmıştı.
- **Toplam ışınlık** ise belli bir zaman sürecinde 1 cm^2 'de kaç pp çarpışması olduğunun ölçümüdür. fb^{-1} , pb^{-1} gibi birimlerle ölçülür.

CMS Integrated Luminosity, pp



Beklenen toplam olay sayısı:

$$N = \sigma_e \times L = \sigma \times BR \times L$$

şeklinde hesaplanır.

Tetikleyiciler (trigger)

Tetikleyiciler veri alımı anında en ilginç olayları seçip sonraki çözümleme için kaydeden filtrelerdir. Veri alımı anında kaydedilmeyen olaylar kaybolur.

Tetikleyiciler sonraki olay seçiminin kaba taslağı gibidir. Aradığımız son durumla tutarlı olayları seçeriz.

Peki tetikleme yaparken neleri kullanırız?

- **Nesneleri kullanırız** (jet, b-jet, e, μ , γ , kayıp dikey enerji)
 - Tetikleme nesneleri belirli tanım ölçütlerine uymalı (ör: izolasyon)
 - Tetikleme nesneleri üzerine eşikler koyarız (ör: $p_T > 100$, $\eta < 2.4$)
 - Tetikleme nesne sayılarına bakarız (ör: 2 μ , 4 jet)
- **Olay değişkenlerini kullanırız**: 2 jet değişmez kütlesi, dikey hadronik momentum, diğer karmaşık değişkenler (α_T , razor, vs.)

Nesne etiketlemek ve seçimi

Nesne seçimi: analizde kullanılacak nesnelere tanımlamak için nesne değişkenleri üzerine uygulanan bir dizi filtreden (cut) oluşur. Nesne seçiminde çoklukla birden fazla filtre kullanılır.

- Örneğin, bir elektronu tanımlamak için
 - Enerji, momentum, yön ve yük bilgisine
 - iz sürücü ve kalorimetrelerde bıraktığı izlerin ve enerji birikintilerinin kalitesi, sayısı ve yön uygunluğuna (etiketleme / identification)
 - olaydaki diğer nesnelere açısal çakışımına (izolasyon)bakarız.
- Analizde kullanılacak tüm nesnelere için benzer etiketleme ve seçim işlemleri uygulanır.
- Farklı amaçlara yönelik pek çok farklı nesne etiketlemesi ve seçimi uygulanabilir.

Olay seçimi

- **Olay seçimi:** olay değişkenleri üzerine uygulanan bir dizi filtreden (cut) oluşur. Olay seçiminde çoklukla **birden fazla filtre** kullanılır.
 - Ör: bir Z bosonu bulmak için önce olay başına 2 elektron ya da 2 müon iste, sonra onların zıt elektrik yüküne sahip olmalarını şart koş, sonra değişken kütlelerini hesapla ve değişken kütlenin Z kütlesine yakın olmasını (ör. 70-100 GeV arası) olmasını şart koş.
- **Seçim akışı (cutflow):** Bir seçime ulaşmak için kullanılan **seçim filtrelerinin tümüdür** (ayrıca seçimi geçen olayların sayısı ve seçim verimliliğidir)
 - **Seçim verimliliği:** Seçimi geçen olayların sayısı bölü tüm olayların sayısı.
- **Seçim bölgesi (region) / kategorisi:** Seçim filtrelerinin tümünün tanımladığı faz uzayı.

Sinyal bölgesi

Sinyal bölgesi: Beklenen sinyalin ardalana göre belirgin gözükeceği şekilde yapılmış olay seçimi. İyi bir sinyal bölgesi bulmak için şunları yapmalıyız:

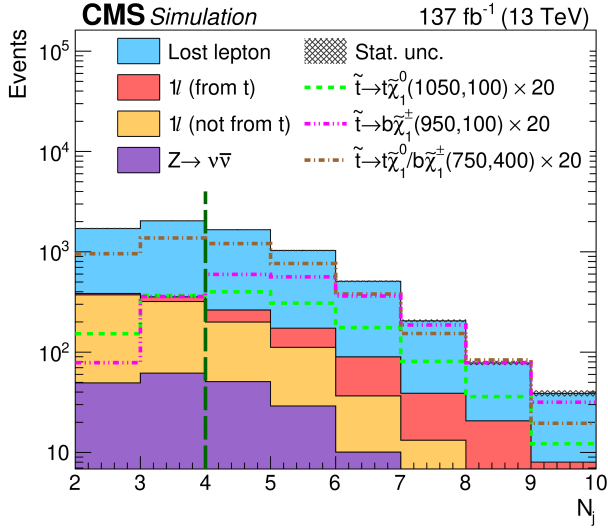
- Sinyali karakterize et:
 - Sinyale özgü özellikler neler? Çok jet? Ters yüklü çift lepton? Kinematik değişkenlerde tepeler, kuyruklar?
- Yeterli sinyal barındıran bir olay seçimi/sinyal bölgesi adayı bul.
- Sinyal bölgesindeki ardalanlara bak:
 - Sinyal ardalan üzerinde yeterli belirginlikte görünüyor mu?
 - Ardalanları belirleyecek yöntemler neler olabilir? (birazdan...)
- Sinyal bölgesini seçtiren tetikleyiciler neler? Bu tetikleyicilerin yeterince verimli olduğuna emin ol.
- Sinyal bölgesini tanımlayan nesnelere algıta düzgün yapılandırılabilirliğine emin ol (örneğin yüksek η 'ya sahip elektronları kullanamayız, çünkü iz sürücü kapsamı dışındalar).

Örnek olay seçimi akışı

Bir CMS süpersimetri analizinden örnek bir seçim akışı tablosu:

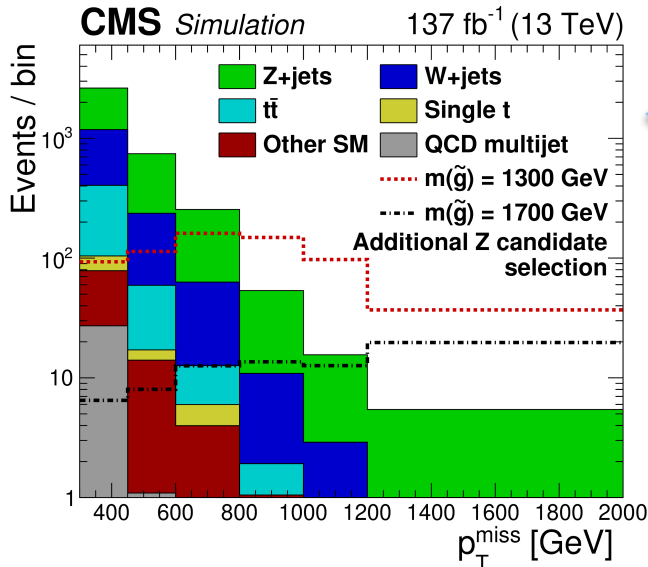
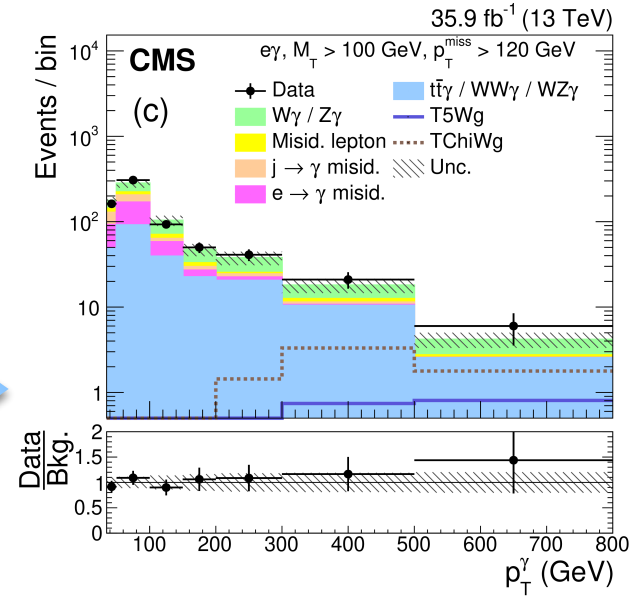
Selection	$pp \rightarrow \tilde{t}\tilde{t}, \tilde{t} \rightarrow t\tilde{\chi}_1^0$ $m_{\tilde{t}} = 950 \text{ GeV}$ $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$	$pp \rightarrow \tilde{b}\tilde{b}, \tilde{b} \rightarrow b\tilde{\chi}_1^0$ $m_{\tilde{b}} = 1000 \text{ GeV}$ $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$	$pp \rightarrow \tilde{q}\tilde{q}, \tilde{q} \rightarrow q\tilde{\chi}_1^0$ $m_{\tilde{q}} = 1400 \text{ GeV}$ $m_{\tilde{\chi}_1^0} = 200 \text{ GeV}$
$N_{\text{jet}} \geq 2$	99.9 ± 0.2	98.8 ± 0.5	99.1 ± 0.5
$H_T > 300 \text{ GeV}$	98.7 ± 0.4	98.3 ± 0.5	98.9 ± 0.6
$H_T^{\text{miss}} > 300 \text{ GeV}$	74.5 ± 1.2	79.6 ± 1.4	88.1 ± 1.4
$H_T^{\text{miss}} / H_T \leq 1$	73.6 ± 1.3	78.2 ± 1.4	86.8 ± 1.5
$N_{\text{muon}} = 0$	58.7 ± 1.4	77.9 ± 1.4	86.7 ± 1.5
$N_{\text{isolated tracks}}^{(\text{muon})} = 0$	58.2 ± 1.4	77.8 ± 1.4	86.7 ± 1.5
$N_{\text{electron}} = 0$	47.2 ± 1.4	77.5 ± 1.5	86.4 ± 1.5
$N_{\text{isolated tracks}}^{(\text{electron})} = 0$	46.4 ± 1.4	77.2 ± 1.5	86.2 ± 1.5
$N_{\text{isolated tracks}}^{(\text{hadron})} = 0$	45.5 ± 1.4	76.8 ± 1.5	85.6 ± 1.5
$N_{\text{photon}} = 0$	43.8 ± 1.4	75.2 ± 1.5	83.6 ± 1.6
$\Delta\phi_{H_T^{\text{miss}}, j_1} > 0.5$	43.6 ± 1.4	75.1 ± 1.5	83.5 ± 1.6
$\Delta\phi_{H_T^{\text{miss}}, j_2} > 0.5$	41.1 ± 1.4	70.6 ± 1.6	78.7 ± 1.7
$\Delta\phi_{H_T^{\text{miss}}, j_3} > 0.3$	39.8 ± 1.4	67.0 ± 1.6	74.4 ± 1.8
$\Delta\phi_{H_T^{\text{miss}}, j_4} > 0.3$	38.5 ± 1.4	64.5 ± 1.6	71.4 ± 1.9
Event quality filter	36.7 ± 1.4	61.4 ± 1.7	67.8 ± 1.9

Sinyal ayırıcı değişkenler: Basit değişkenler



Nesne sayısı, ör. jet sayısı, lepton sayısı, b-jet sayısı, vb.

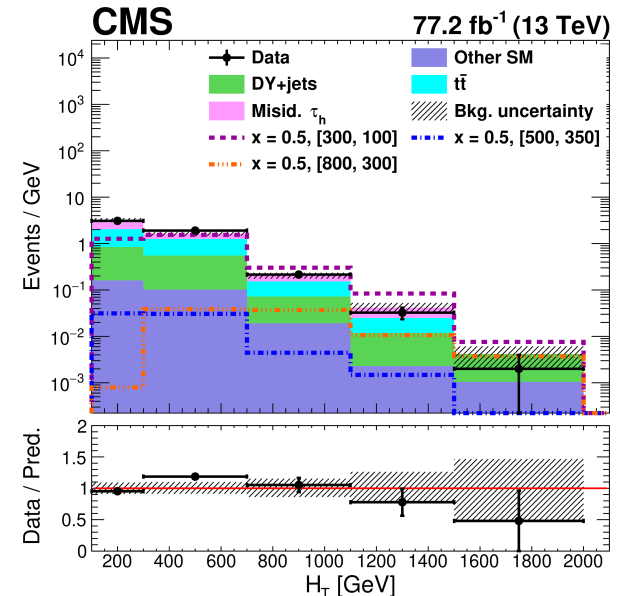
Nesne dik momentumu p_T , ör. 1. jet p_T , 2. muon p_T , foton p_T , vb.



Kayıp dik enerji / momentum E_T^{miss} / p_T^{miss} (diğer adıyla MET!)

Hadronik dik enerji.

$$H_T = \sum_i^{n \text{ jets}} p_T^{\text{jet}_i}$$

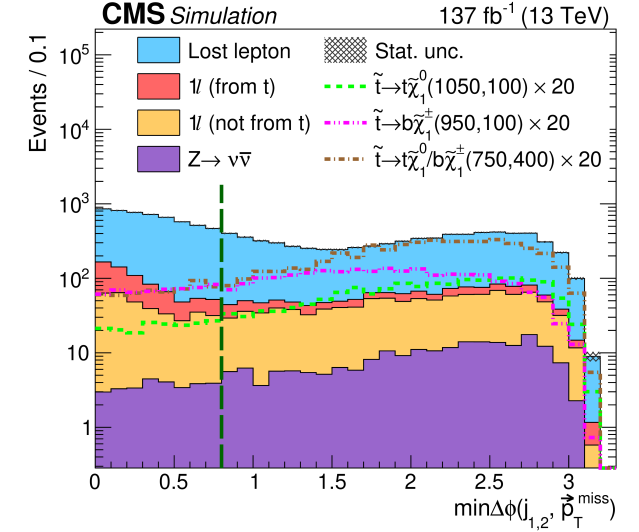
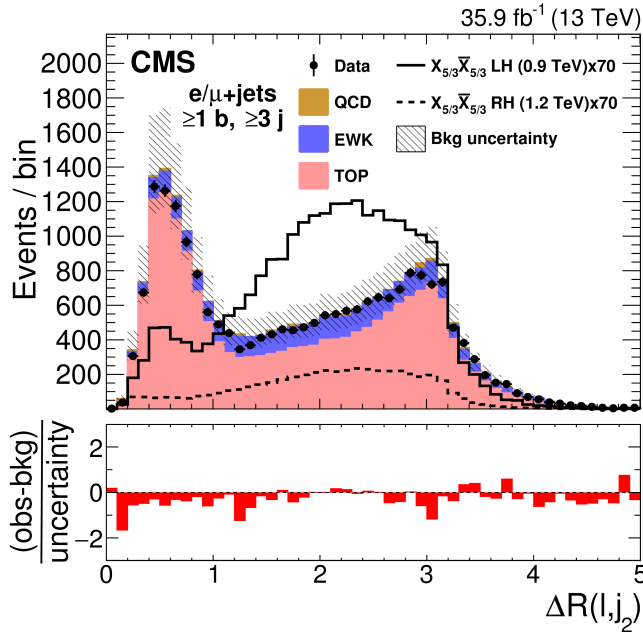


Sinyal ayırıcı değişkenler: Açısal değişkenler

İki nesne arasındaki ya da bir nesne ve MET arasındaki **azimutal açısal aralık $\Delta\phi$** ,

Olayda birden fazla aynı nesne olduğunda **minimum $\Delta\phi$** kullanılabilir.

Sahte METli olayları atmak için kullanılır.



İki nesne arasındaki **açısal aralık ΔR** :

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

Nesnelerin yaklaşıklığını ölçer. Örneğin iki nesnenin bir parçacıktan bozunup bozunmadığını anlamak için kullanılır.

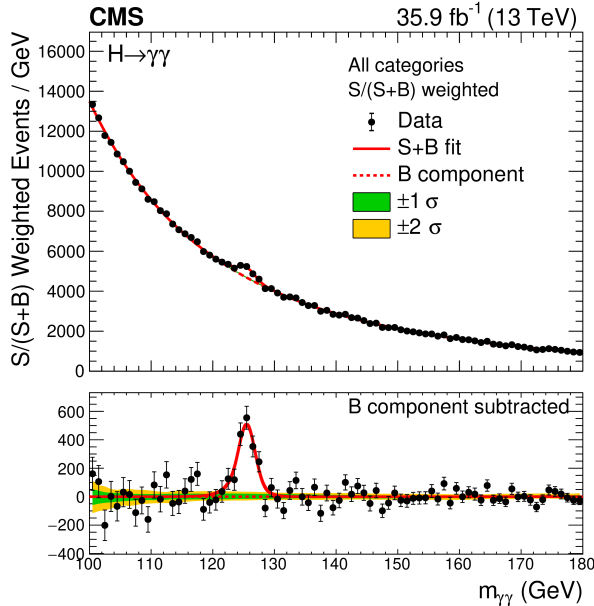
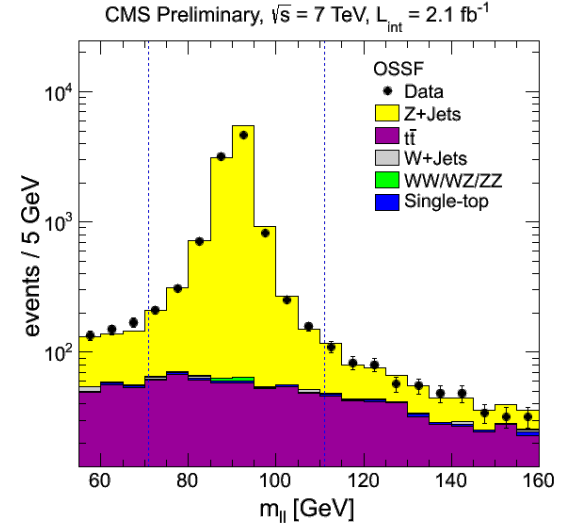
Sinyal ayırıcı değişkenler: Değişmez kütle

i tane son durum parçacığına bozunan bir parçacık için **değişmez kütle**

$$m = \sqrt{\left(\sum_i E^i\right)^2 - \left(\sum_i \vec{p}^i\right)^2}$$

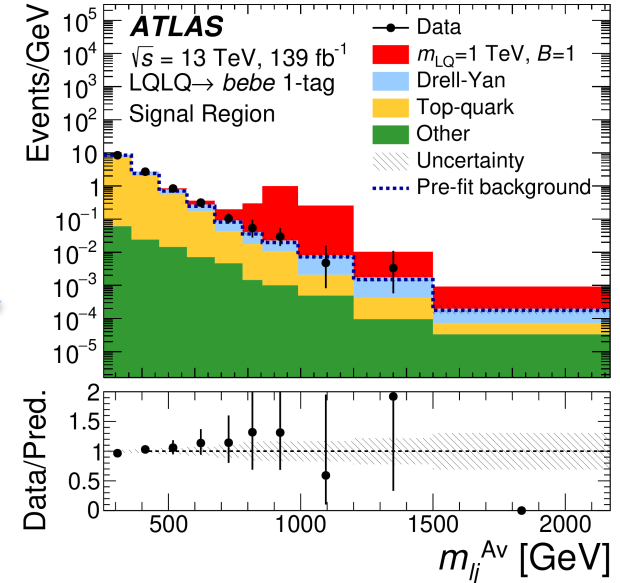
Ana parçacık için değişmez kütle ancak **tüm son durum parçacıklarının 4-momentumu bilinirse** hesaplanabilir.

Değişmez kütle, seçimde kütlesi bilinen parçacıklar (Z gibi), ya da görünür parçacıklara bozunduğu varsayılan yeni parçacıklar aranırken kullanılır.:



$m_{\gamma\gamma} = 125 \text{ GeV}$ 'de Higgs gözlemi.

Ancak leptokuark değişmez kütle dağılımında veriden farklılık gözlenmedi.



Veri analizinde makina öğrenimi

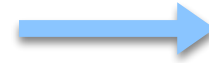
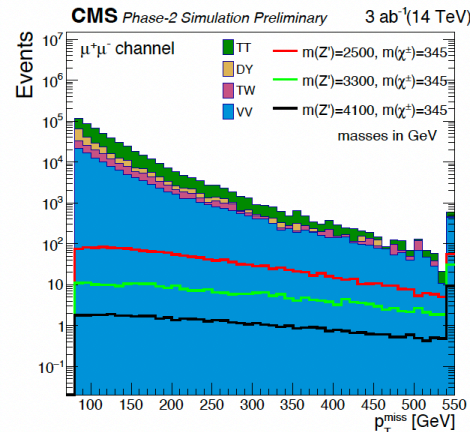
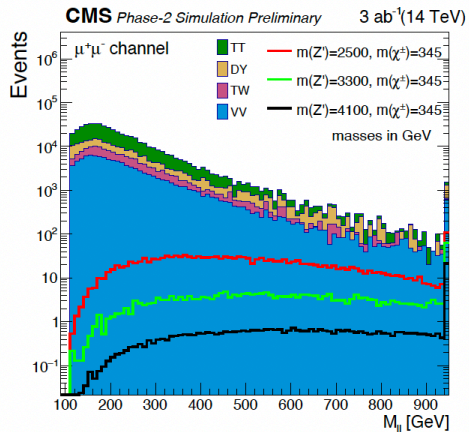
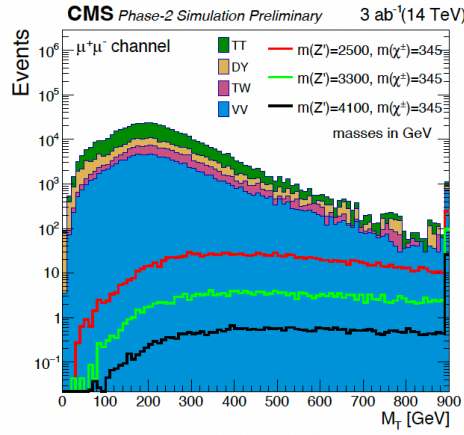
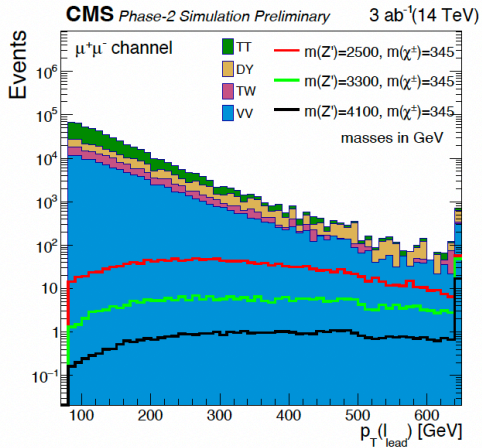
Makina öğrenimi yöntemleri, optimal olay seçimi elde etmede çok yararlıdır.

- Sınıflandırma MÖ yöntemleri olayları sinyal veya ardaan gibi çeşitli gruplara ayırmak için kullanılır.
 - Ayrıca nesnelere sınıflandırmak ve nesnelere için en iyi tanıma kriterlerini elde etmek için de kullanılır. Örnek: b-jetleri hafif jetlerden ayırmak.
- Yıllardır arttırılmış karar ağaçları gibi geleneksel yöntemler kullanılmıştır. Tekli üst kuark (single top) ve Higgs bozonu gibi keşifler bu yöntemlerle gerçekleşmiştir.
- Günümüzde sinir ağları, hatta derin sinir ağları, daha geniş GPU erişilebilirliği nedeniyle daha yaygın kullanılmaktadır.
- MÖ yöntemleri, özellikle büyük ardaanın altında gömülü küçük sinyallerin bulunduğu durumlar için çok faydalıdır. Basit seçimler yeterli duyarlılık sağlamadığında, en üst düzeyde duyarlılık çıkarmak için uygulanırlar.
- Şimdilerde MÖ yöntemleri veride genel anomalileri / sinyalleri arama amaçlı kullanılmaya başlanmıştır.
- Sınıflama yöntemleri haricinde regresyon yöntemleri de, enerji, kütle vb. değerlerin ölçümleri için yaygın olarak kullanılmaktadır.

Makina öğrenimi ile olay sınıflama

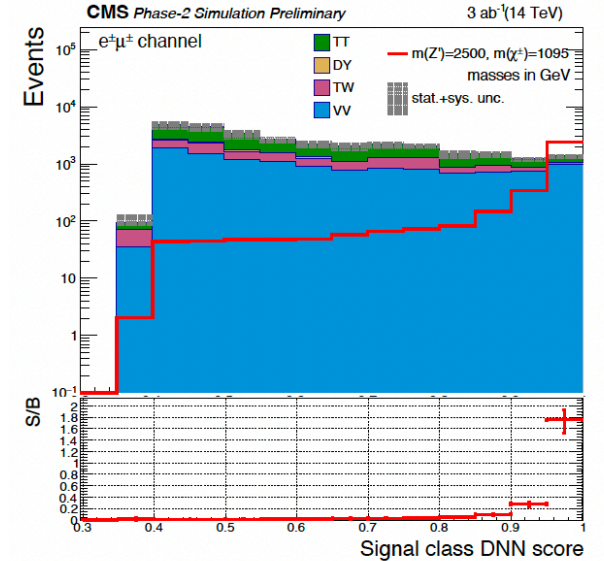
Makina öğreniminde sınıflama N kısıtlı ayırıcı güçte değişkeni birleştirip bir adet etkin ayırıcı güçte değişken hesaplayacak bir fonksiyon elde etmeyi kapsar:

$$ML_{\text{ayırıcı}} = f(x_1, x_2, \dots, x_N) \leftarrow \text{Görev } f \text{ fonksiyonunu bulmaktır.}$$



Soldaki değişkenler tek başlarına sinyal olaylarını ayırtmaya yetmez.

Bu değişkenler ve bir derin sinir ağı aracılığı ile birleştirilerek sağdaki güçlü ayırıcı oluşur.

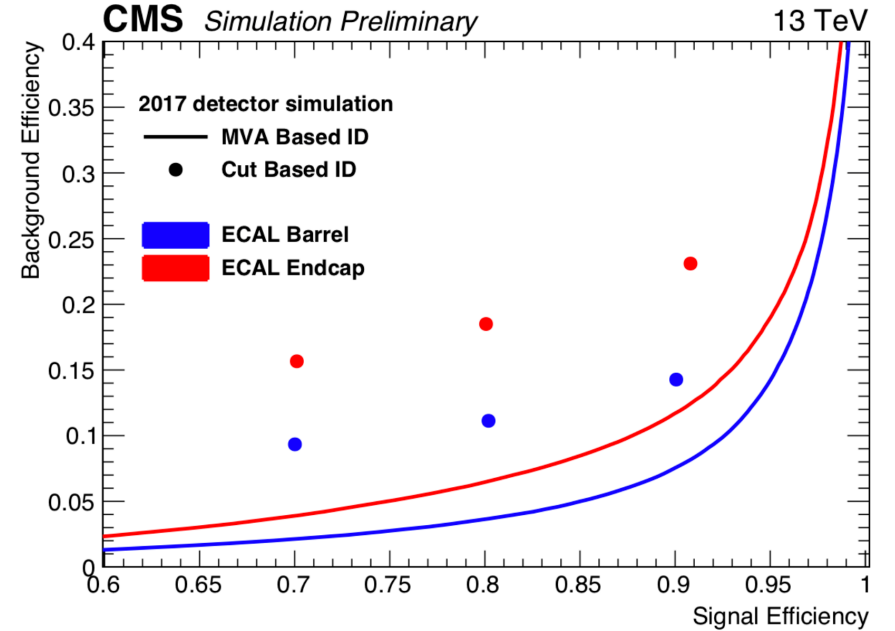


Makina öğrenimi ile nesne tanımlama

MO yöntemleri, nesne tanımlama ve sınıflandırma için de kullanılır.

Basit değişken temelli seçimlerle tanımlamaya göre daha iyi performans sağlarlar.

Daha iyi sınıflandırılmış nesnelere, daha iyi sinyal karakterizasyonuna, optimizasyona ve ölçümlere olanak tanır.



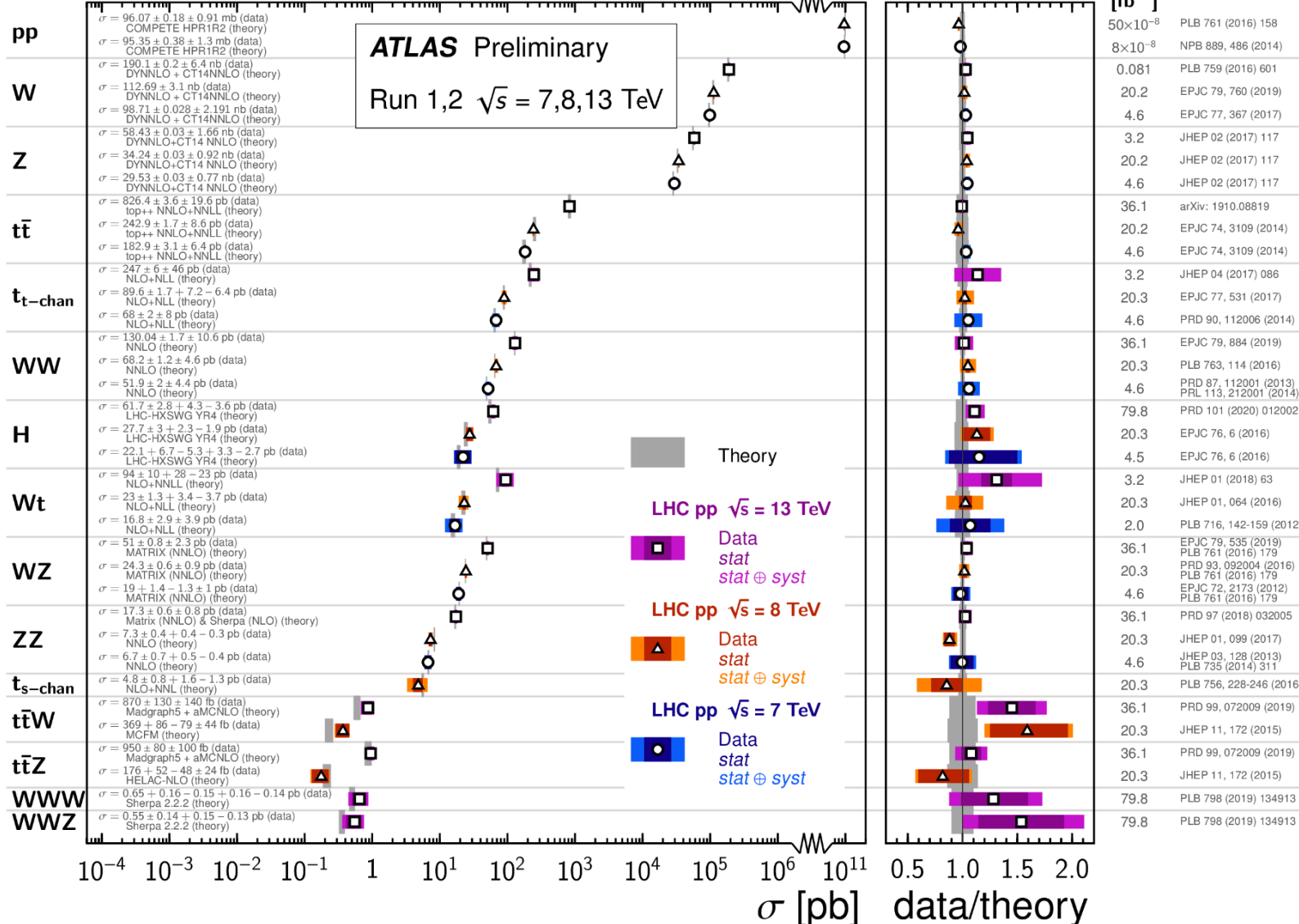
SM ardalanlar

Standard Model Total Production Cross Section Measurements

Status:
May 2020

$\int \mathcal{L} dt$
[fb⁻¹]

Reference



SM süreçler özellikle yeni fizik arayışlarında ardalan oluşturur.

Ardalan kestirimi

Sinyal seçiminden sonra sinyal bölgesinde kalan indirgenemez ardalanı kestirmemiz gerek. Bu iş için farklı yöntemler kullanılır:



Monte Carlo benzetim beklentilerini kullan:

- Kuram ve algıç hakkındaki tüm bilgimizi içerir.
- MC olayların hangi fiziği içerdiğini biliriz.
- Kabalıktan kesinliğe giden uzun ama ısrarlı bir yol.

Veri güdümlü kestirim yöntemleri bul:

Ortak ilke: **kontrol bölgeleri kullan**

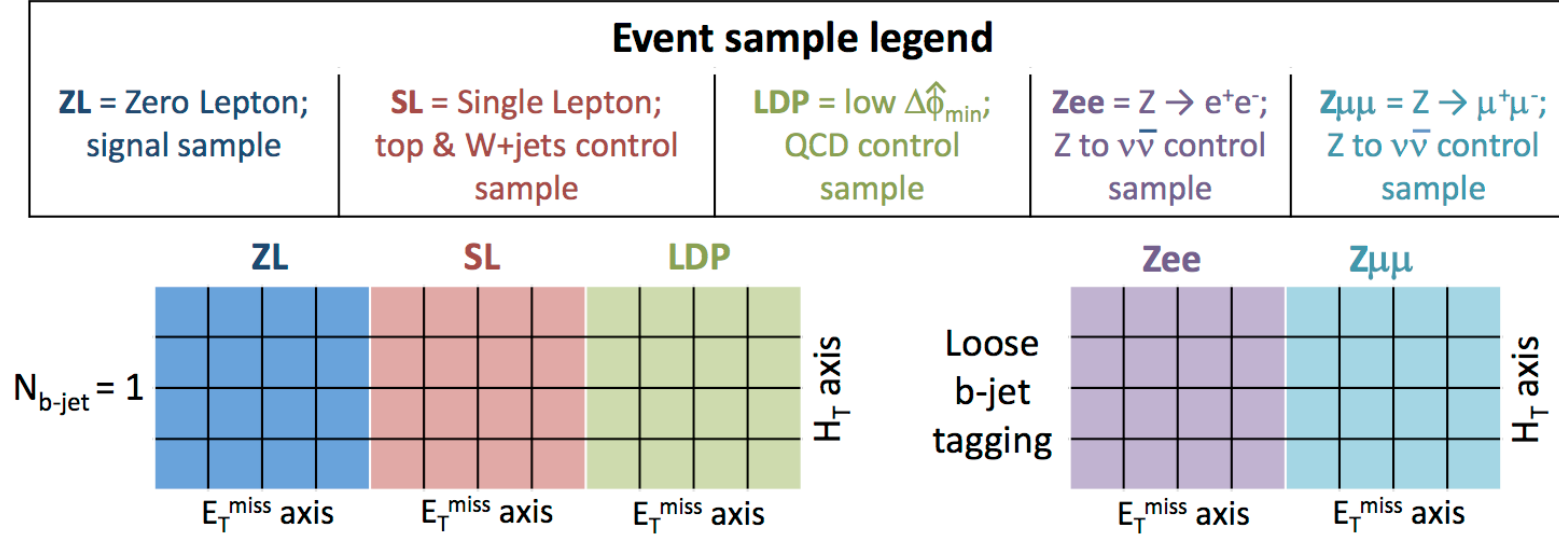
- **Kontrol bölgesi:** Ardalanın hakim olduğu, sinyalin göz ardı edilebilir olduğu bir olay seçimi
- Kontrol bölgesinden ardalanına dair bilgi edin ve bu bilgiyi sinyal bölgesine ulaştır.



Veri ve simülasyon birlikte çalışır:

- Veri, simülasyon üreteçleri ince ayarlama kullanılır.
- Simülasyondaki kinematik dağılımlar veri güdümlü yöntemlerde kullanılır.

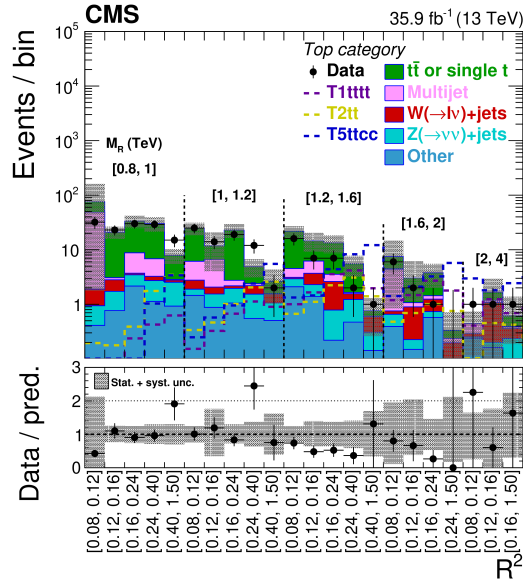
Ardalan kestirimi: Kontrol bölgeleri ve simülasyon oranları



- Sinyal bölgelerindeki **bazı seçim kriterleri ters çevrilerek** kontrol bölgeleri elde edilir.
- Kontrol bölgesindeki her i aralığındaki gerçek veri ardalan miktarı bulunur.
- Bu miktar simülasyon ile hesaplanan sinyal bölgesi / kontrol bölgesi ardalan beklentisi oranı ile çarpılır.

$$N_{\text{veri, ardalan}}^{\text{SB},i} = \frac{N_{\text{MC, ardalan}}^{\text{SB},i}}{N_{\text{MC, ardalan}}^{\text{KB},i}} N_{\text{veri, ardalan}}^{\text{KB},i}$$

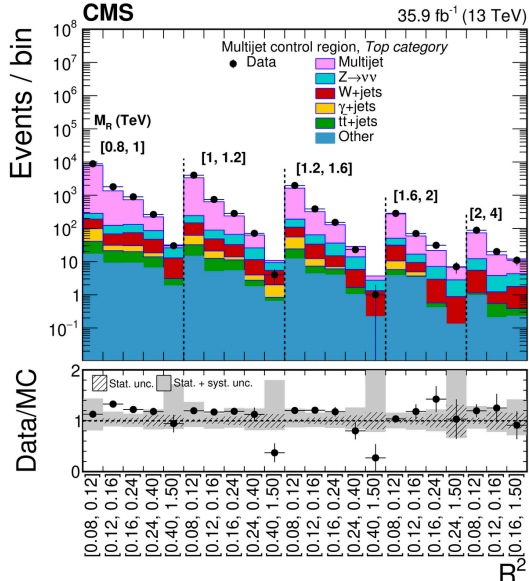
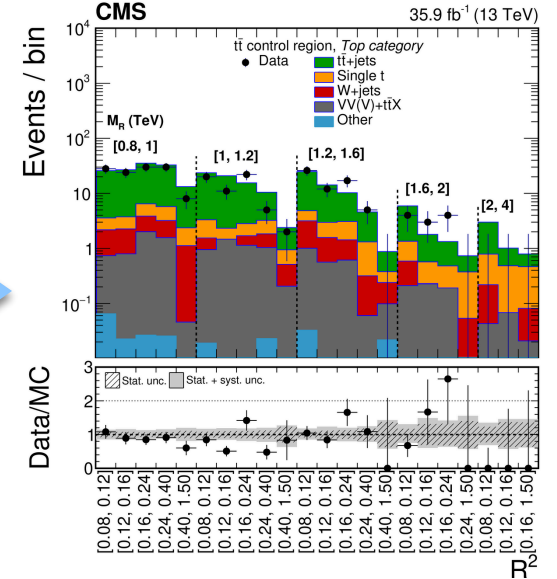
Ardalan kestirimi: Kontrol bölgeleri ve simülasyon oranları



0 lepton sinyal bölgesi.

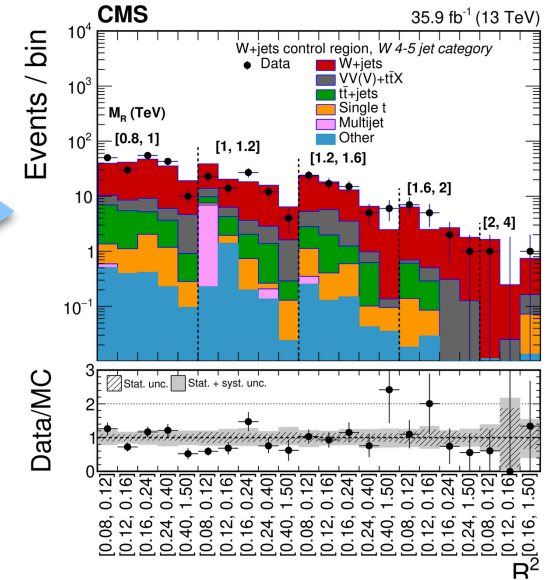
tt+jets için tek lepton + b jet kontrol bölgesi

Signal and control regions in the CMS SUSY analysis with razor variables.



W+jets için single lepton + 0b jet kontrol bölgesi

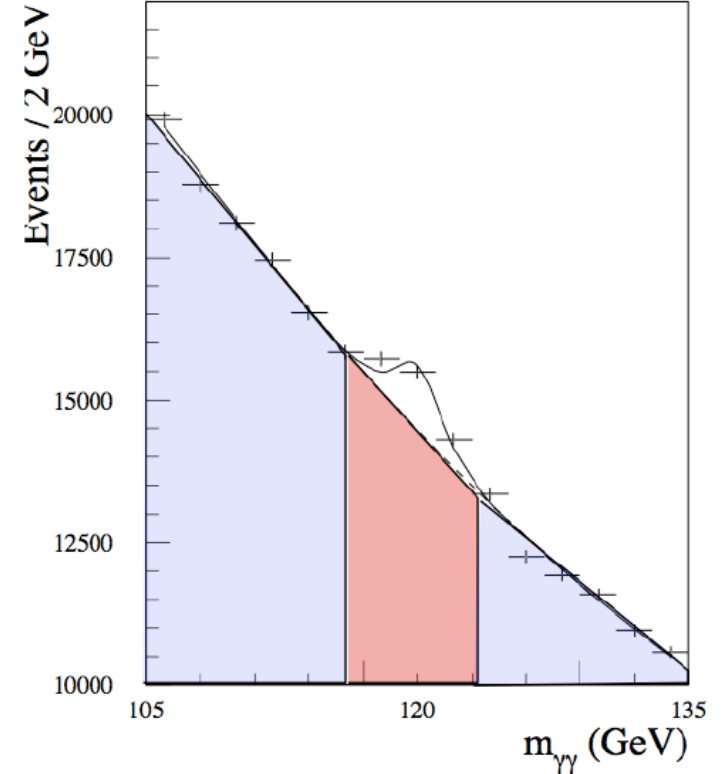
QCD için ters çevrilmiş $\Delta\phi_{min}$ kontrol bölgesi.



Ardalan kestirimi: Yan aralık yöntemi

Rezonans ararken kullanılır: Ardalan düz olacak ve sinyal ardaan üzerinde tepe yapacak.

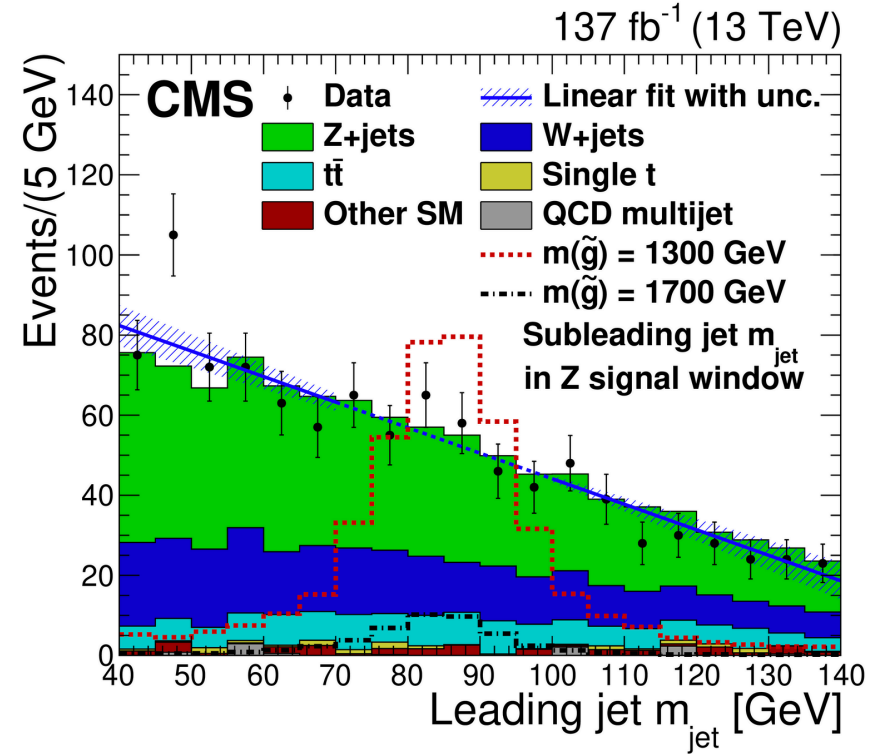
- Sinyal bölgesinin iki yanında sinyalsiz yanaralık bölgeleri belirle.
- Ardalanın şeklini yanaralıklardan bul (polinom, üstel, vs.)
- Yanaralıklarda bulunduğun ardalanı sinyal bölgesine uzat.
- Ya sinyal tepesinin altındaki ardalanı say ya da veri dağılımını sinyal + ardalan fonksiyonuna oturt ve ardalan fonksiyonunun parametrelerini bul.



Ardalan kestirimi: Yan aralık yöntemi

Rezonans ararken kullanılır: Ardalan düz olacak ve sinyal ardalın üzerinde tepe yapacak.

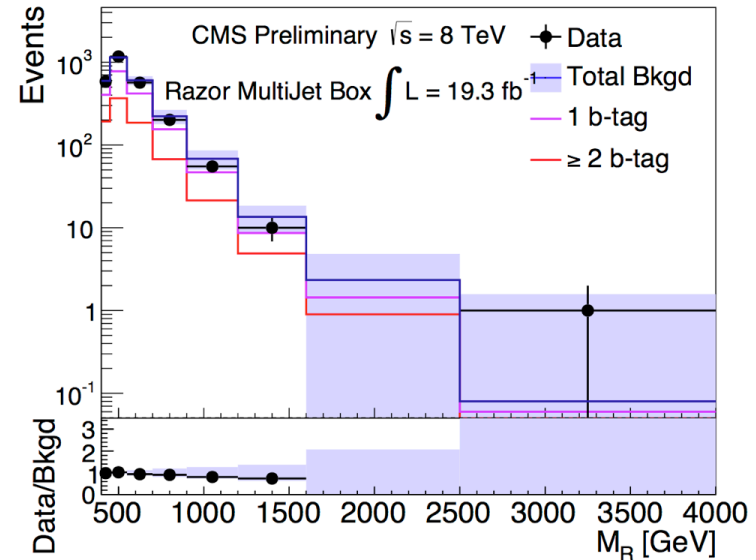
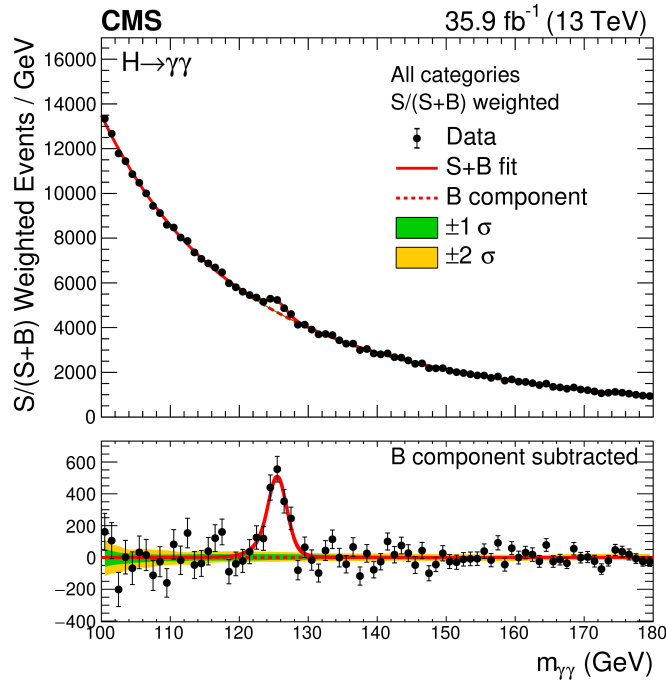
- Sinyal bölgesinin iki yanında sinyalsiz yanaralık bölgeleri belirle.
- Ardalanın şeklini yanaralıklardan bul (polinom, üstel, vs.)
- Yanaralıklarda bulunduğun ardalanı sinyal bölgesine uzat.
- Ya sinyal tepesinin altındaki ardalanı say ya da veri dağılımını sinyal + ardalan fonksiyonuna oturt ve ardalan fonksiyonunun parametrelerini bul.



Ardalan kestirimi: Analitik fonksiyona oturtma

Bazen ardaan analitik bir fonksiyon ile tanımlanabilir. Bu durumda:

- Ardalanın baskın olduğu kontrol bölgesinde veriye analitik fonksiyona oturtup analitik fonksiyonun ardaanı iyi anlattığına emin olunur. Fonksiyonun parametreleri bulunur.
- Ya oturtulan eğri sinyal bölgesine uzatılır, ya da sinyal bölgesinde ardaan + sinyal için bir fonksiyon bularak veriye oturtulur.



Sistemik hatalar

Sistemik hatalar deney / hesap düzeneğine ait kusurlardan kaynaklanan hatanın ölçüme etkisidir. Şansa bağlı hata değildir.

Analizlerde görebileceğimiz sistemik hatalar:

Deneysel (algıç kaynaklı):

- Işınlık hesabı belirsizliği
- Tetikleme verimliliği hataları
- Jet enerji ölçeği, jet energy çözünürlüğü
- Lepton, foton, b-jet, W-jet, top-jet, vs. verimliliklerindeki hata/belirsizlik

Kuramsal (hesabı yeterli kesinlikte yapamamak kaynaklı):

- Tesir kesiti ve dallanma oranı hesaplarındaki hata / belirsizlik
- Kinematik dağılımlardaki hesap hataları
- Parton dağılım fonksiyonlarındaki hata / belirsizlik

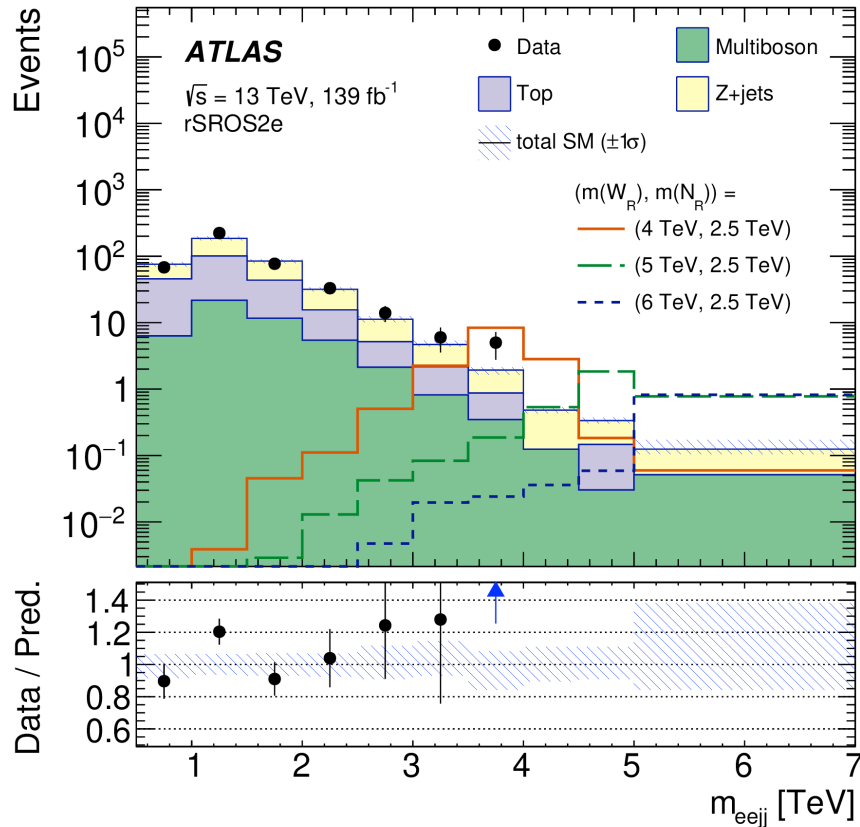
Özellikle analizde simülasyon kullandığımız zaman algıç kaynaklı sistemik hataları MC'ya yansıtmak gerekir.

Analiz sonuçları

Bir **deneyin sonucu** o deneyde **gözlenen niceliklerdir**:

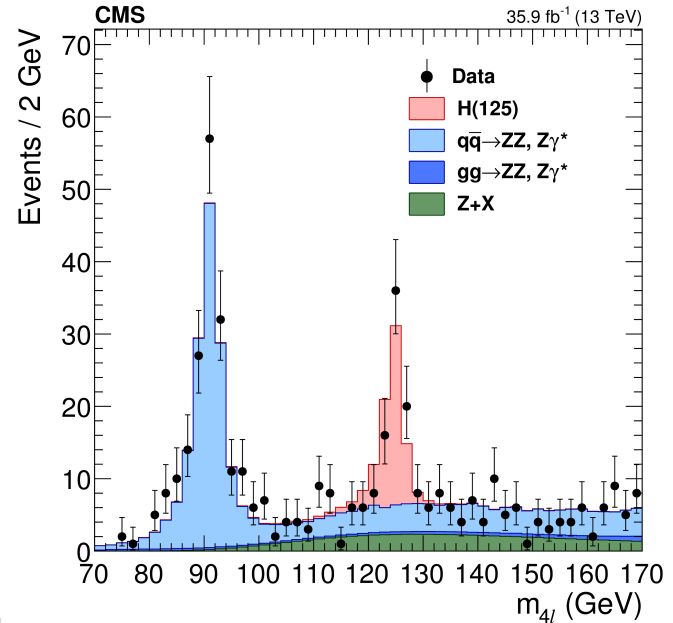
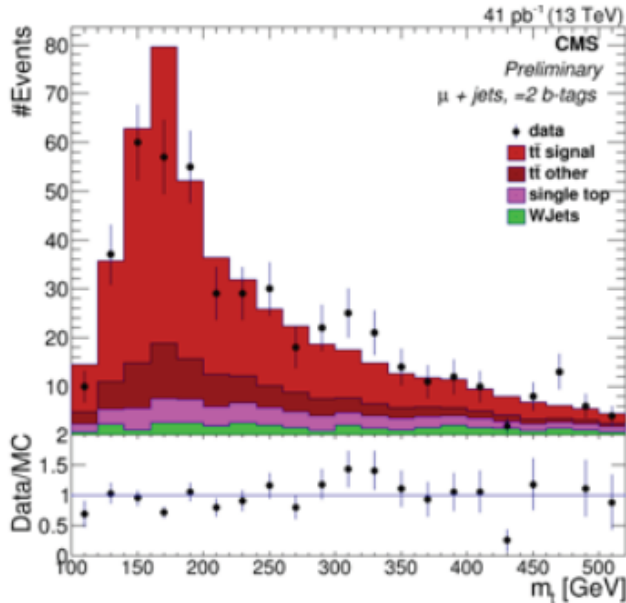
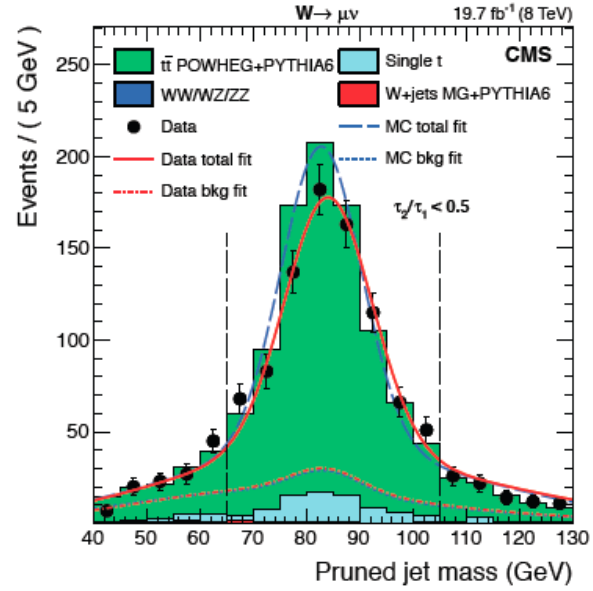
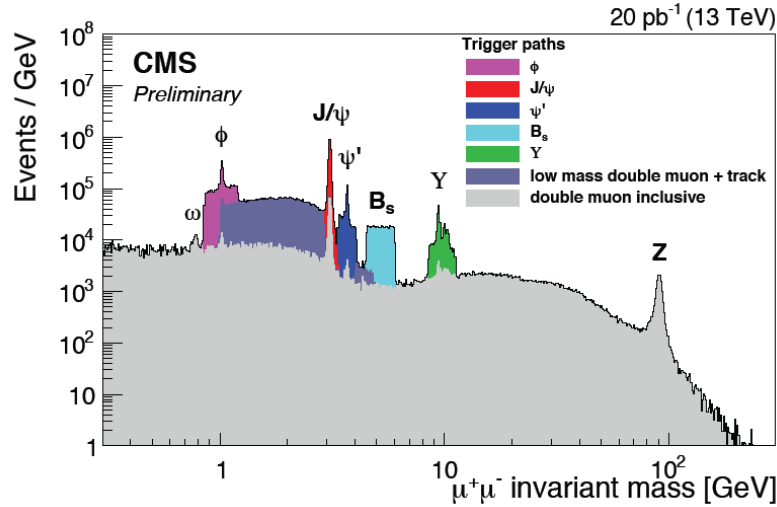
- Olay sayıları (beklenen ardalardan sapma var mı?)
- Kinematik dağılımlar
- Kinematik uçların varlığı
- Kütleler
- Tesir kesitleri, dallanma oranları
- Genlik
- Yük asimetrisi
- Spin bağıntıları
- vs.

Analiz sonuçları: Olay sayımı



- Analizde **sinyale en duyarlı değişken(ler)** belirlenir ve bölmelere (bin) ayrılır.
- Gözlenen olay sayısı, kestirim ile tahmin edilmiş aralan sayısı ve sistematik hataların binlerdeki dağılımları gösterilir.
- Bir arama analizinde, gözlenen verileri ve beklenen aralan tahminini karşılaştırarak uyum veya uyumsuzluk olup olmadığını inceleriz. Bunu istatistiksel bir analizle nicelendiririz.
- Birkaç sinyal noktasından beklenen Monte Carlo olay dağılımlarını göstermek gelenekseldir, ancak bunlar analiz sonucunun bir parçası değildir.

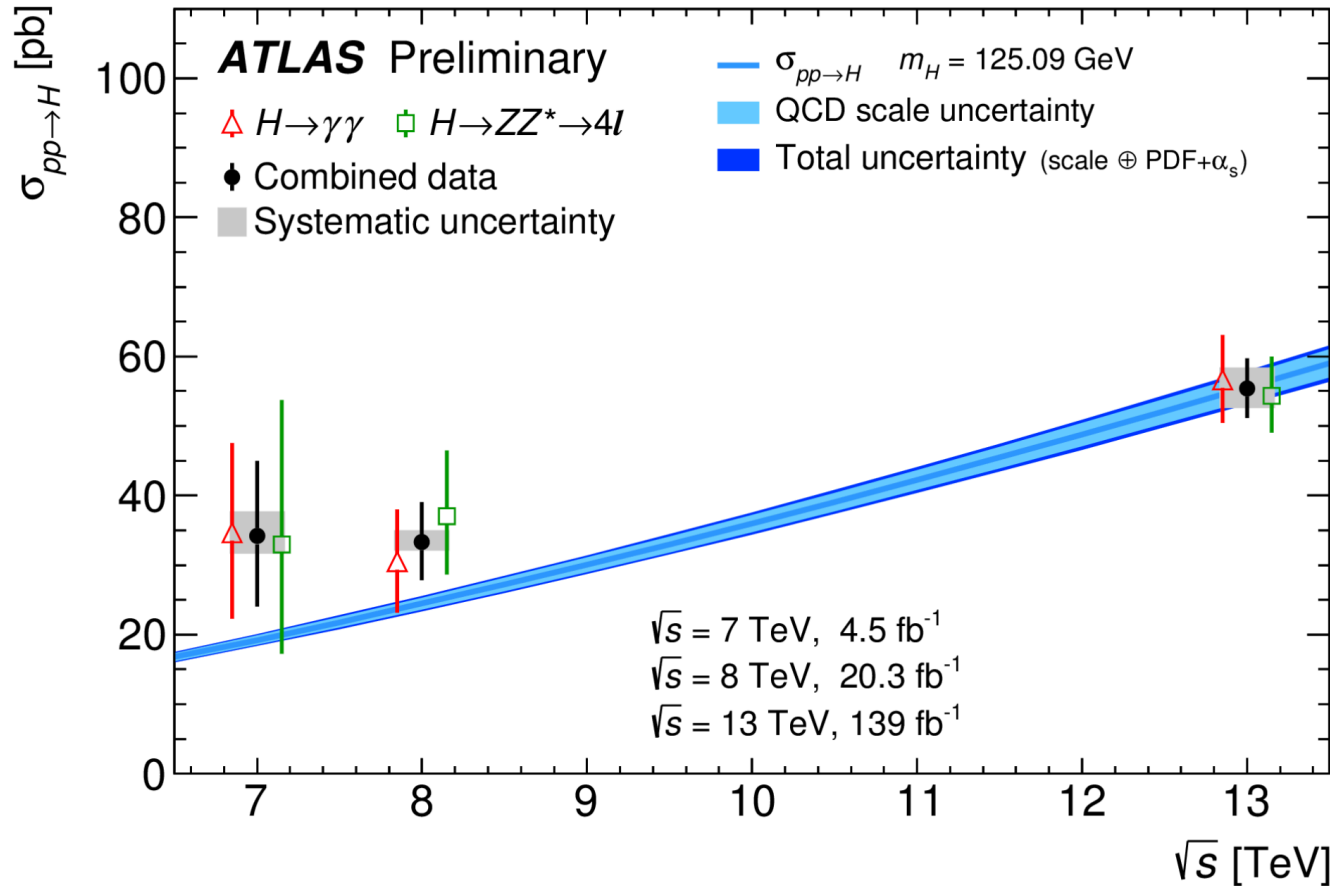
Analiz sonuçları: Kütle



Analiz sonuçları: Tesir kesiti

Tesir kesiti olayları sayıp, ardalan beklentisini çıkarıp toplam ışınlığa ve sinyal verimliliğine bölerek bulunur.

$$\sigma_{sinyal} = \frac{N_{gozlenen} - N_{ardalan}}{\int \mathcal{L} dt \cdot \epsilon_{sinyal}}$$

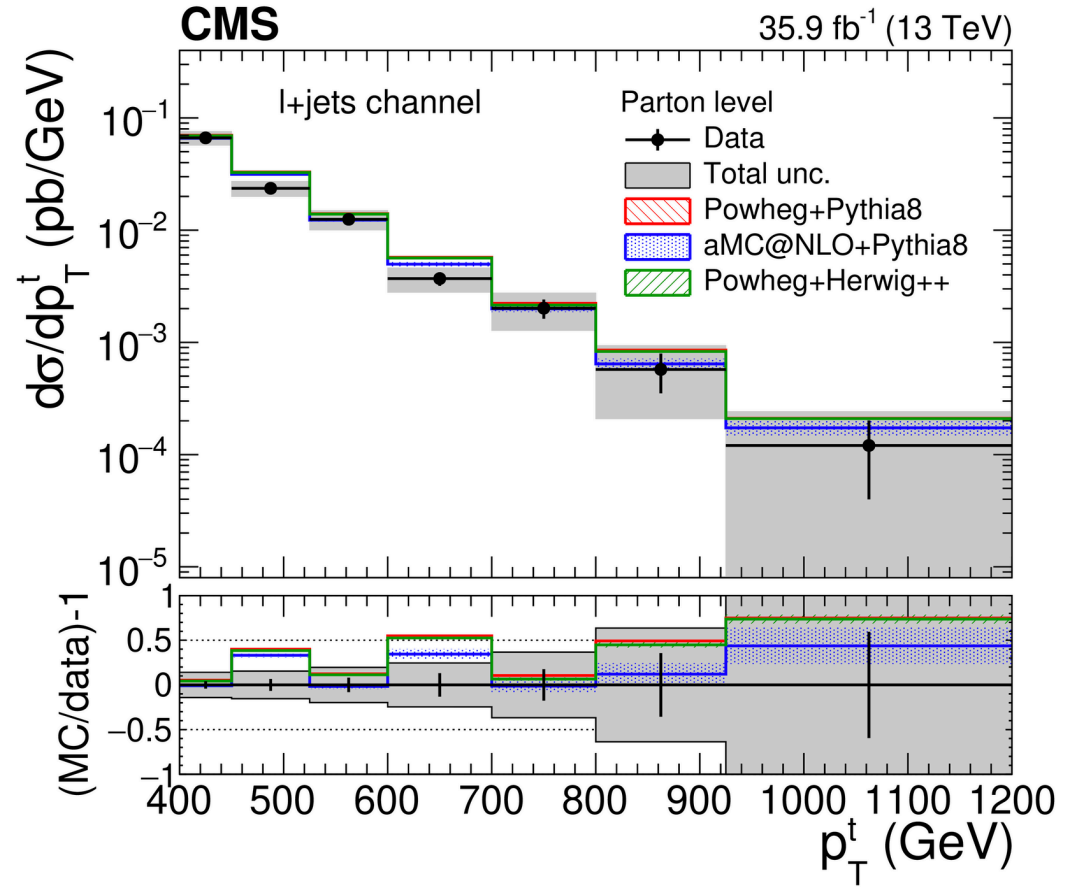


Higgs oluşum toplam tesir kesitinin farklı enerjilerde ölçümü ve SM beklentisiyle karşılaştırılması.

Analiz sonuçları: Diferansiyel tesir kesiti

Tesir kesitinin bir değişken karşısındaki değişimi ölçülebilir.

Sağda tepe kuark - karşı tepe kuark oluşum sürecinin tepe kuark dikey momentumuna göre tesir kesit değişiminin ölçümü görülmektedir.



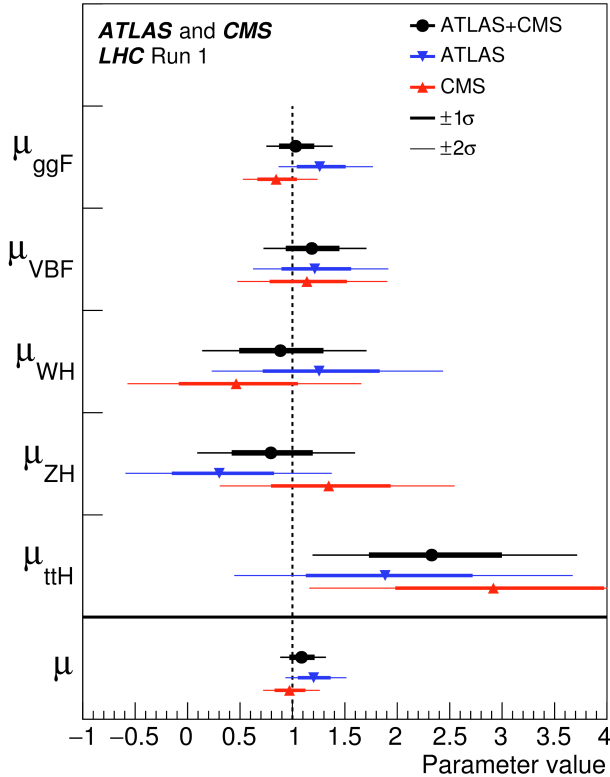
Analiz sonuçları: Sinyal gücü

Bir analiz kanalı için **sinyal gücü**:

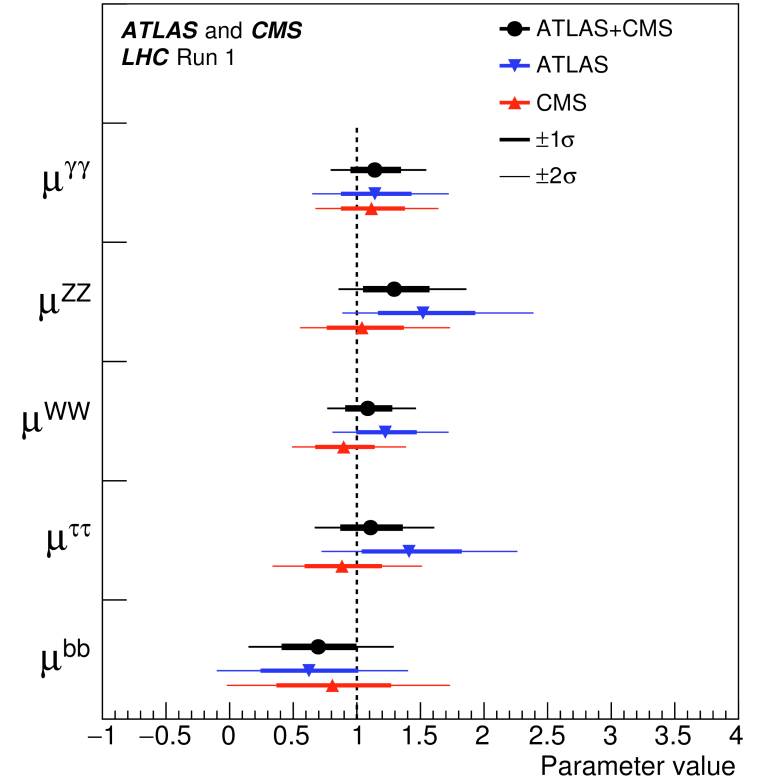
$$\mu = \frac{\text{Süreç için gözlenen olay sayısı}}{\text{Süreç için beklenen SM olay sayısı}}$$

$\mu = 0$: Sinyal yok
 $\mu = 1$: Sinyal SM ile uyumlu.
 $\mu > 1$: Sinyal SMden farklı
 (SM ötesi fiziğe işaret)

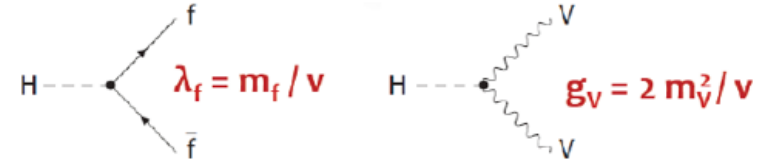
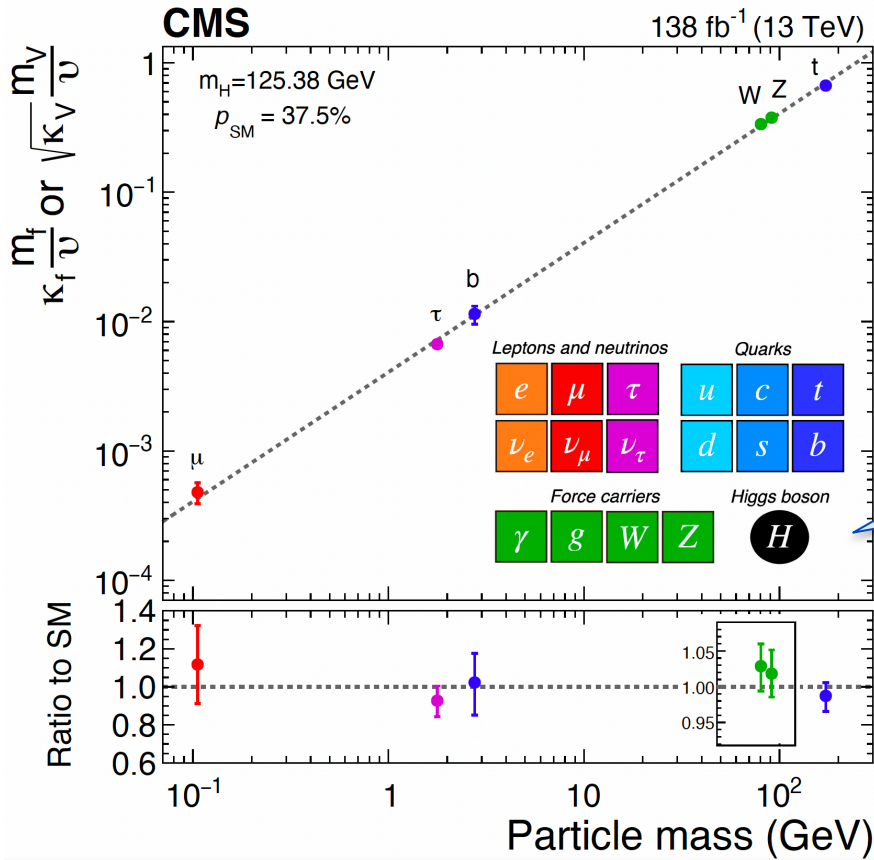
Higgs
tesir
kesiti



Higgs
dallanma
oranı



Analiz sonuçları: Bağlaşımlar



v : vakum beklenti değeri

Parçacıklar Higgs bosonu ile ne kadar çok etkileşirlerse o kadar çok kütle kazanırlar.

SM parçacıkların Higgs alanı ile etkileşim gücüne karşı SM parçacıkların kütleleri.

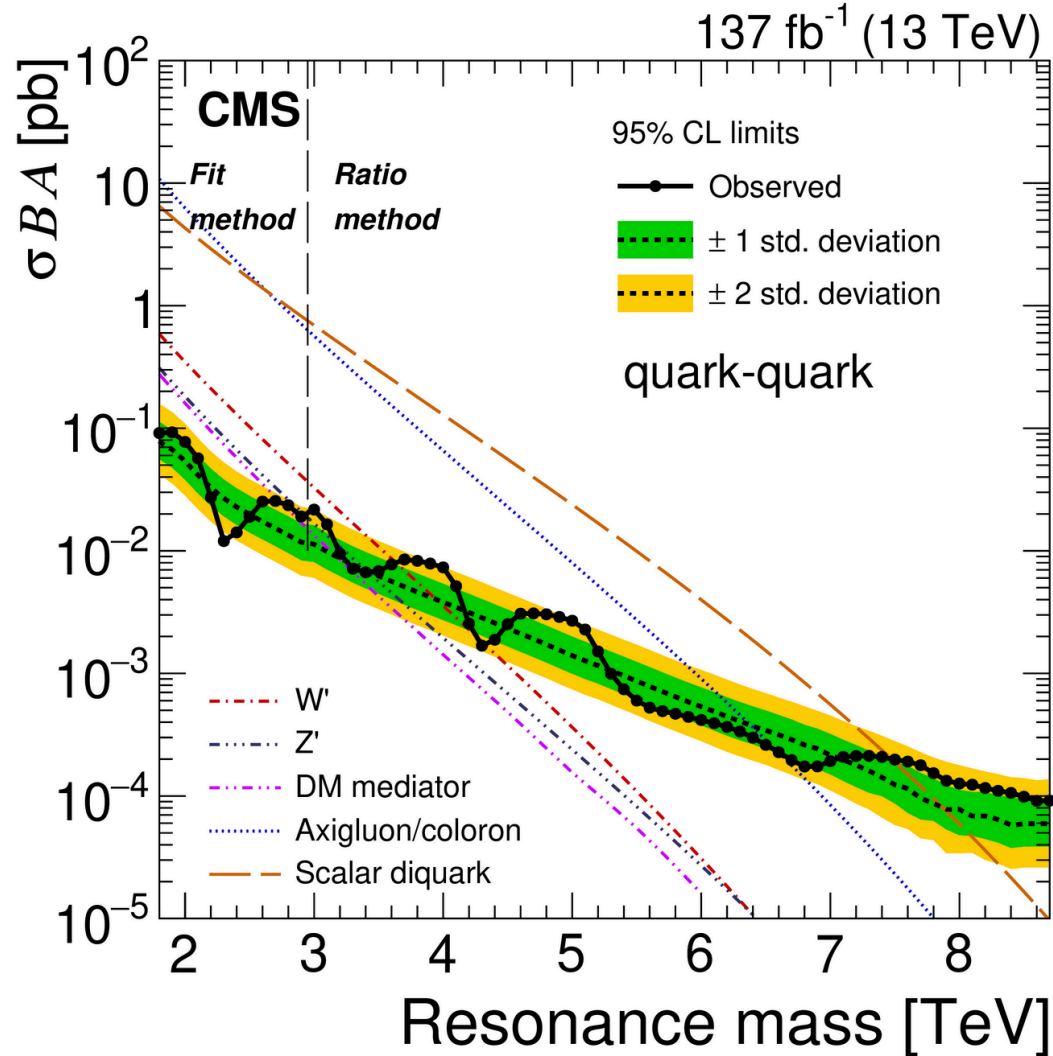
Renkli noktalar etkileşim ölçümünü ve hatayı, kesikli çizgi de SM beklentisini gösteriyor. **Veri SM ile uyumlu.**

Bu ölçümde görülebilecek **bir sapma yeni fiziğe işaret olurdu!**

Yorumlama (interpretation) ve istatistik analiz

- **Yorumlama** deney sonucunun bir kuramsal modelin beklentileri ile karşılaştırılmasına nedir.
- **DİKKAT!** Yorumlama, deney sonucu ile aynı şey değildir!
- Deney sonucunu yorumlamak için istatistik model ve likelihood kullanırız.
 - Bir analizin **istatistik modeli** o analizin tam matematiksel anlatımıdır.
 - **Olasılık yoğunluğu** $p(x|\theta)$ gözlenen değişkenler x 'i fizik model parametreleri θ ile ilişkilendirir.
 - **Likelihood** $L(\theta) = p(X_0|\theta)$ olasılık yoğunluğu $p(x|\theta)$ 'in **gözlemlenen değerler** X_0 kullanılarak hesaplanması ile bulunur.
 - Likelihood **her ciddi yorumlama çalışmasının başlangıç noktasıdır.**
- θ parametrelerini istatistiksel yöntemlerle hesaplarız ve kuramsal modelin geçerliliğini test ederiz.
- Bir deneysel sonuç **birden fazla kuramsal model** ile yorumlanabilir. Analizin olabildiğince çok kuramla yorumlanabilmesi için **gerekli bilgiler tüm camia ile paylaşılmalıdır.**

Tek serbest parametrelili model için yorumlama



Özet

Bu derste **veri analizi kavramını ve aşamalarını** inceledik.

Veri analizi bizi **deney çıktısından deney sonucuna ulaştırır.**

Anlamlı deney sonuçları elde etmek doğru veri analizine bağlıdır.

Doğru ve verimli bir analiz için aşağıdakiler gereklidir:

- Sinyalin ayırd edici özelliklerini iyi tanımak
- İyi tasarlanmış tetikleyiciler ve nesnelere çalışmak
- Sinyali en çok koruyan ve ardalanı en çok azaltan bir olay seçimi
- Yüksek kesinlikli bir ardalın tespit stratejisi
- Sistemik hataları eksiksiz belirlemek ve hesaplamak
- Doğru bir istatistiksel analiz.

Analiz sonrasında analizin farklı kuramlara etkisini görmek için **yorumlama çalışmaları yapılır, ve kuramcılarca yapılabilmesi için ek bilgiler sağlanır.**