Two Machine Learning techniques for Model Independent New Physics searches at the LHC

LLR Seminar

Fabricio Jiménez

20th of April, 2020

Laboratoire Leprince-Ringuet

Former support

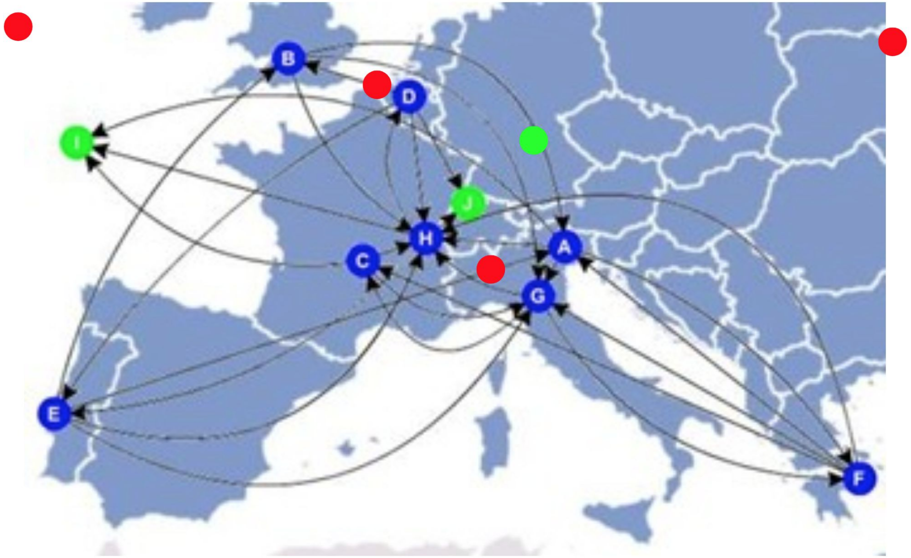
Supervisor: Prof. Julien Donini

AMVA4NewPhysics

Research sponsored by

Advanced Multivariate Analysis for New Physics
(2015 - 2019)

- 10 students across **European nodes**
- **Academic** & **industrial** partners
- Secondments
  - **University of Padova** (1.5 months)
  - **University of California** (2 months)
  - **The MathWorks, Inc.** (3 months)
  - **CERN** (2 months)



Outline

1. Searching for New Physics at the LHC
    - The Standard Model, the LHC and ATLAS
    - Model Independent searches for New Physics
    - Monitoring generic physics channels

2. Methods for Model Independent searches for New Physics
    - A Semi-supervised approach for anomaly detection
    - Gaussian Processes for resonance searches

The Standard Model (SM)

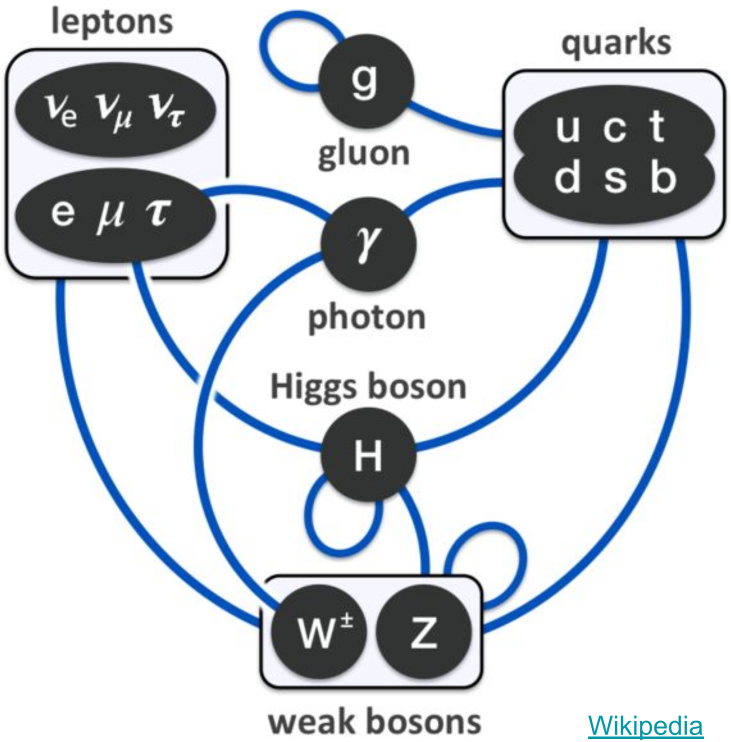
Successful theory of fundamental particles + interactions

Describes three of the four forces in nature

- Electromagnetism
- Weak interactions
- Strong interactions

Electroweak

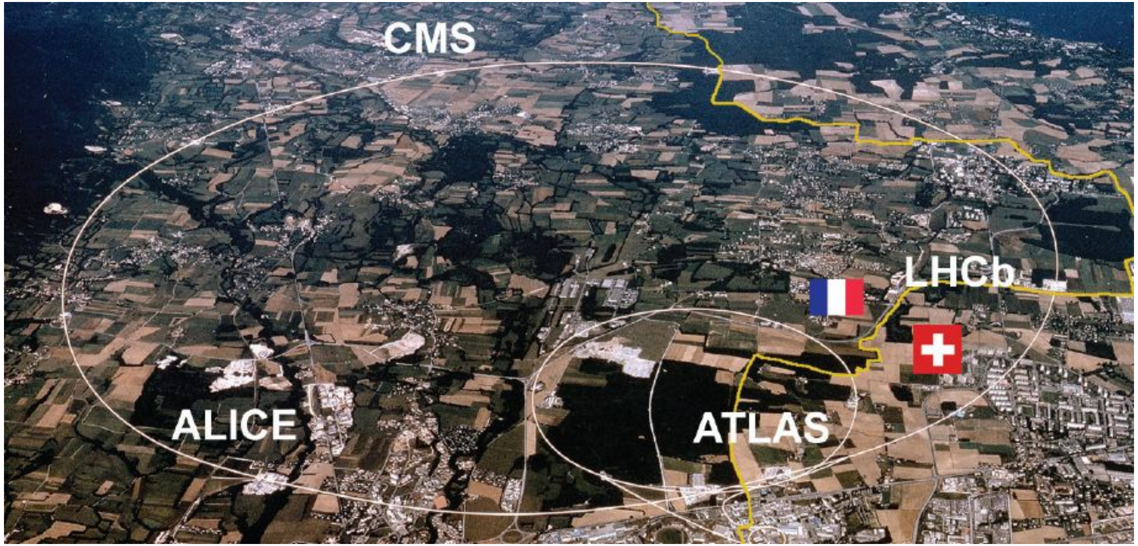
Experimentally tested for decades



Wikipedia

The Large Hadron Collider (LHC)

The **LHC** is the most powerful particle collider



IEEE Spectrum

- 27 km (main) ring
- 100 m underground
- Proton beams in opposite directions
- Collisions at an energy of 13 TeV every 25 ns

Probe the SM and search for New Physics

What is New Physics (NP)?

The SM is not complete: no gravity, no dark matter, matter-antimatter asymmetry,...

New Physics → phenomena beyond the SM

Theoretical extensions of the SM

- New symmetries:
  - Between fermions and bosons (SUSY), left/right symmetry of weak sector,...
- Extra dimensions:
  - Warped Extra Dimensions, Large Extra Dimensions (ADD),...
- Compositeness
- Extended Higgs sector:
  - Two Higgs doublets,...

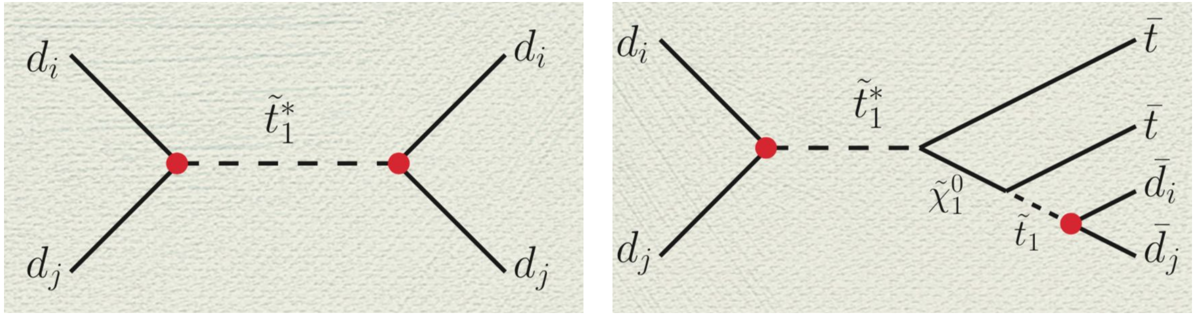
Example: RPV-MSSM

R-Parity Violating - Minimal Supersymmetric Standard Model

Conventional SUSY, R-parity is introduced:

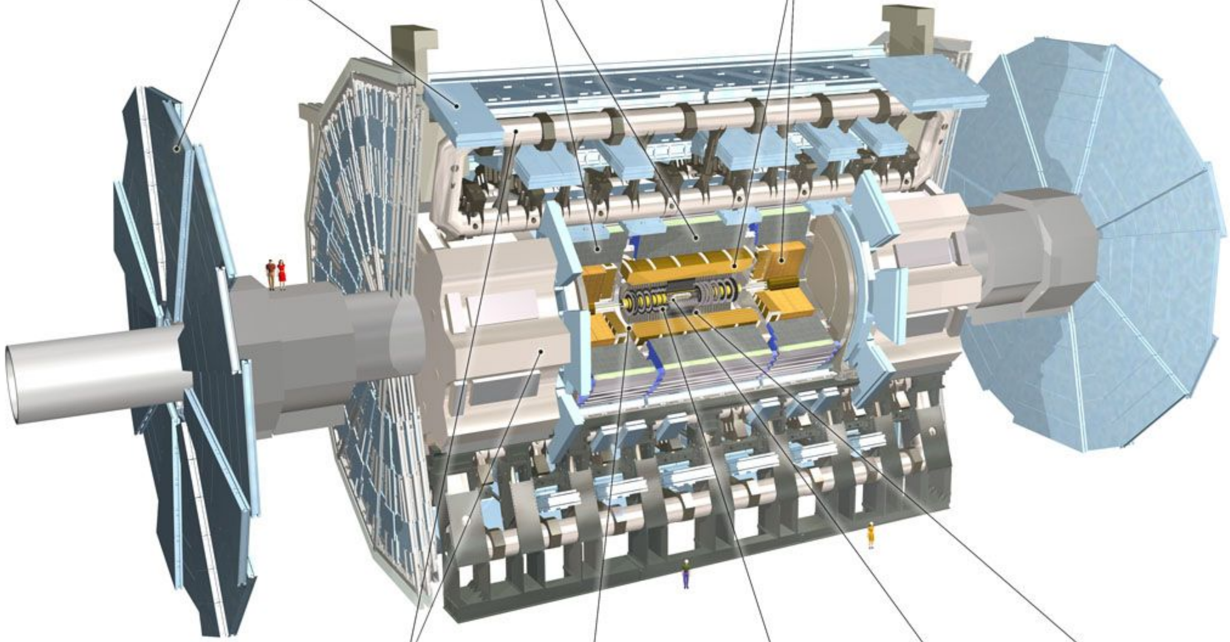
$$R = (-1)^{3B+L+2s}$$

More generally, R need not be imposed



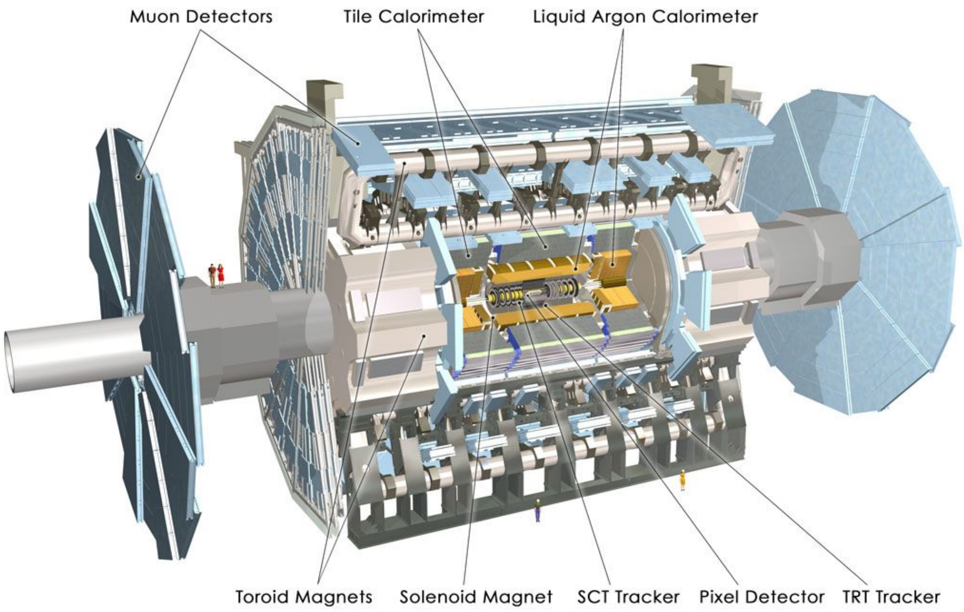
Figures from https://indico.in2p3.fr/event/14083/contributions/16644/

A Toroidal LHC ApparatuS (ATLAS)

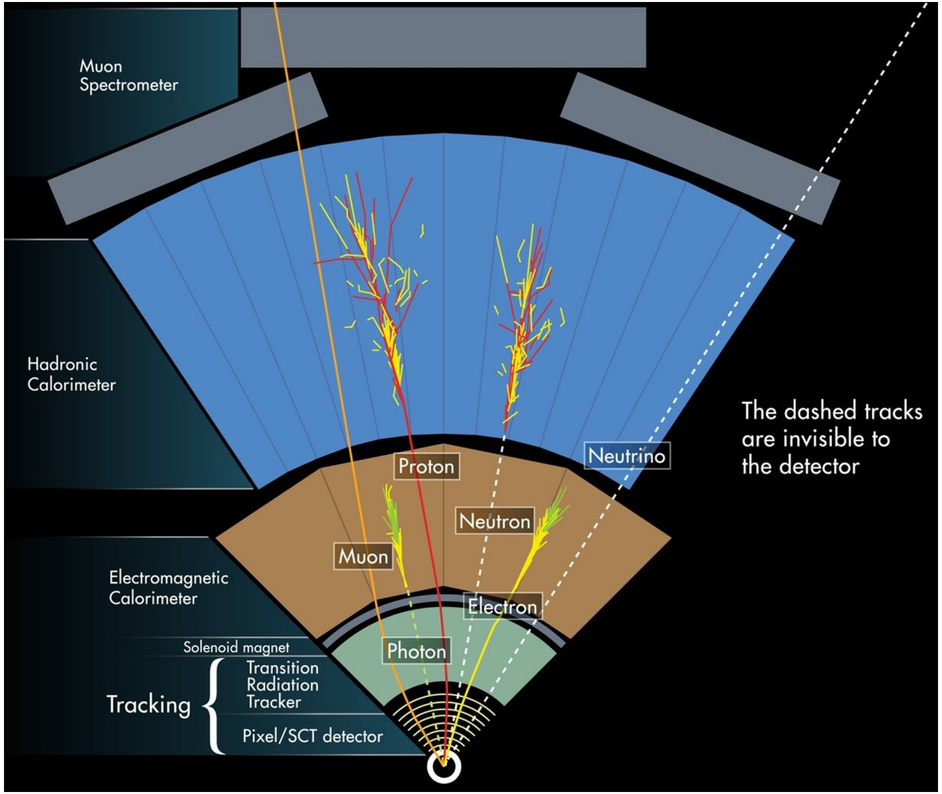


https://cds.cern.ch/record/1095924

A Toroidal LHC ApparatuS (ATLAS)



https://cds.cern.ch/record/1095924



https://cds.cern.ch/record/1505342

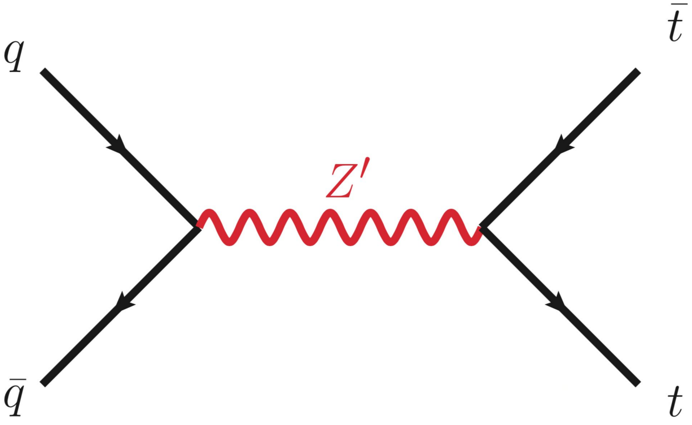
Outline

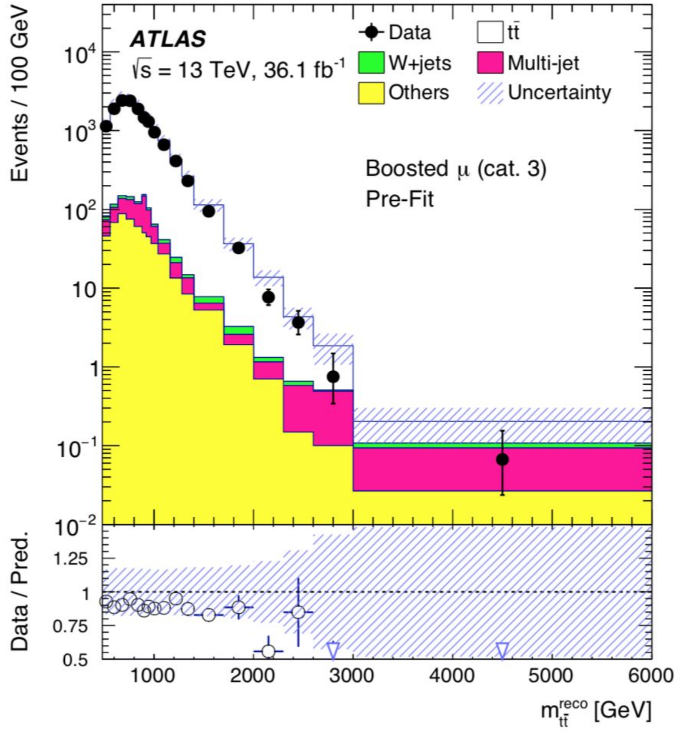
1. Searching for New Physics at the LHC
   - The Standard Model, the LHC and ATLAS
   - **Model Independent searches for New Physics**
   - Monitoring generic physics channels

2. Methods for Model Independent searches for New Physics
   - A Semi-supervised approach for anomaly detection
   - Gaussian Processes for resonance searches

Searching for New Physics - Model Dependent

- Select a NP model, simulate signal
- Probe specific signatures/hypotheses





Aaboud, M., Aad, G., Abbott, B. et al. Eur. Phys. J. C (2018) 78: 565.

Searching for New Physics - Model independent

All accessible signatures

Not feasible

General Searches

Model dependent

Model independent

Most LHC New Physics searches
(e.g. previous slide)

"Popular" signatures:

- Two objects: Dijet, ee, tt,...
- E_T^{miss} + jets
- ...

One-to-few signatures

What are General Searches?

Multi-signature + Model independent → General

Multi-signature

- Final states from combinations of objects
- Classify events using final states
- Number of classes:
  - 8-object final states with 5 kinds of objects:

All LHC model-dependent NP searches ~ 200 classes

$$\sum_{k=1}^{8} \left(\binom{5}{k} \right) > 1200$$

Automated analysis of high volumes of data!

Counting data and Monte Carlo

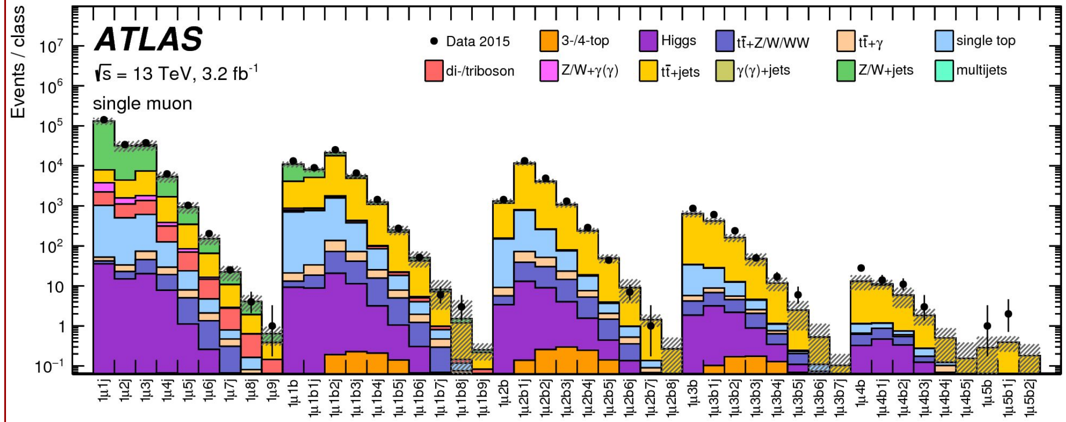
Multi-signature + Model independent → General

Model independent

- Search for data deviations to the SM
- Count data and SM in 622 classes
- Scan kinematic distributions:
  → **Quantify deviations**
- Not as sensitive as dedicated analyses

Assumption:

New Physics will appear in final states with high-p_T objects



- Good agreement in most channels
- Examine regions w/ largest deviations

[Aaboud, M., Aad, G., Abbott, B. et al. Eur. Phys. J. C (2019) 79: 120.]

Outline

1. Searching for New Physics at the LHC
    - The Standard Model, the LHC and ATLAS
    - Model Independent searches for New Physics
    - **Monitoring generic physics channels**

2. Methods for Model Independent searches for New Physics
    - A Semi-supervised approach for anomaly detection
    - Gaussian Processes for resonance searches

TADA: A Fast Monitoring System for ATLAS

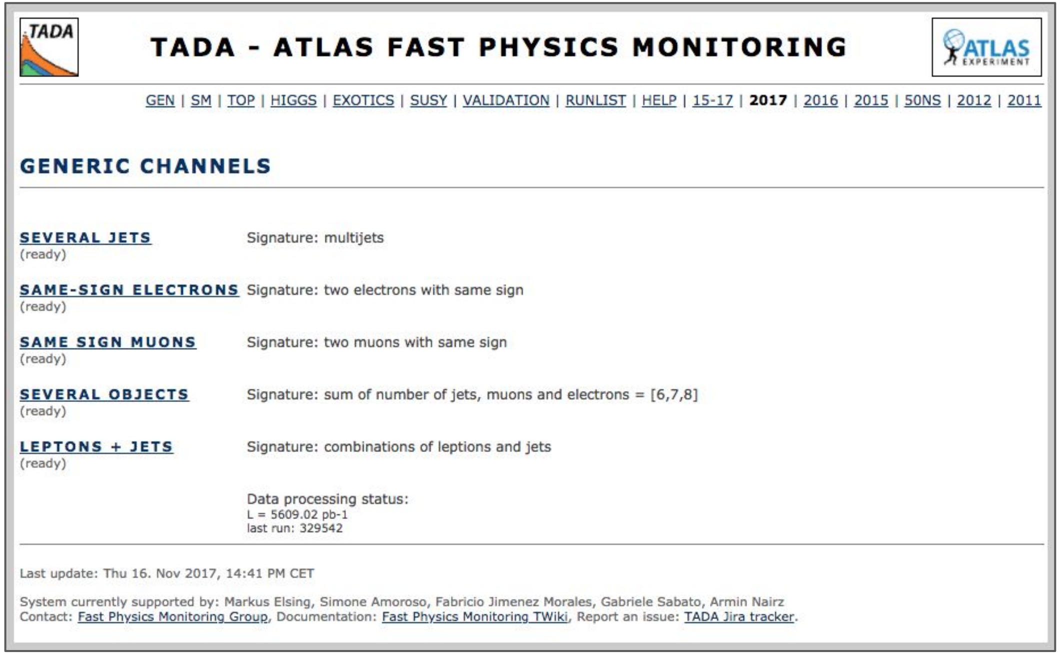
Monitoring (Python/C++) software for ATLAS → early warning system

- Hundreds of selections (more histograms)
- Data quality and sim. performance studies

TADA has a web interface
→ updated daily during data taking

Qualification task:

- Data and SM simulation validation
- Software maintenance
- **Generic Search system and webpage**


TADA - ATLAS FAST PHYSICS MONITORING

GEN | SM | TOP | HIGGS | EXOTICS | SUSY | VALIDATION | RUNLIST | HELP | 15-17 | **2017** | 2016 | 2015 | 50NS | 2012 | 2011

GENERIC CHANNELS

SEVERAL JETS (ready)	Signature: multijets
SAME-SIGN ELECTRONS (ready)	Signature: two electrons with same sign
SAME SIGN MUONS (ready)	Signature: two muons with same sign
SEVERAL OBJECTS (ready)	Signature: sum of number of jets, muons and electrons = [6,7,8]
LEPTONS + JETS (ready)	Signature: combinations of leptions and jets
	Data processing status: L = 5609.02 pb-1 last run: 329542

Last update: Thu 16. Nov 2017, 14:41 PM CET

System currently supported by: Markus Elsing, Simone Amoroso, Fabricio Jimenez Morales, Gabriele Sabato, Armin Nairz
Contact: Fast Physics Monitoring Group, Documentation: Fast Physics Monitoring TWiki, Report an issue: TADA Jira tracker.

Monitoring Generic Channels

Idea: monitor automated generic selections

- Inspired by General Searches
- Four physics variables monitored

$$M_{\text{inv}}, \qquad E_T^{\text{miss}}$$

$$H_T = \sum_{\text{objects}} |p_{T,\text{object}}|$$

$$M_{\text{eff}} = H_T + E_T^{\text{miss}}$$

- Selections transparent to **TADA**
- Automatic web page generation

Group	# Selections	Variables
Multijets	10	Number of jets = {6,7,8,9,10} $H_T > \{1,2\}$ TeV
Multiobjects	3	Number of objects = {6,7,8}
Several Photons	4	Number of photons = {2,3} $H_T > \{250,500\}$ GeV
Leptons plus jets	16	Number of leptons (e,μ) = {1,2} Number of jets = {2,3,4,5} $H_T > \{1,2\}$ TeV

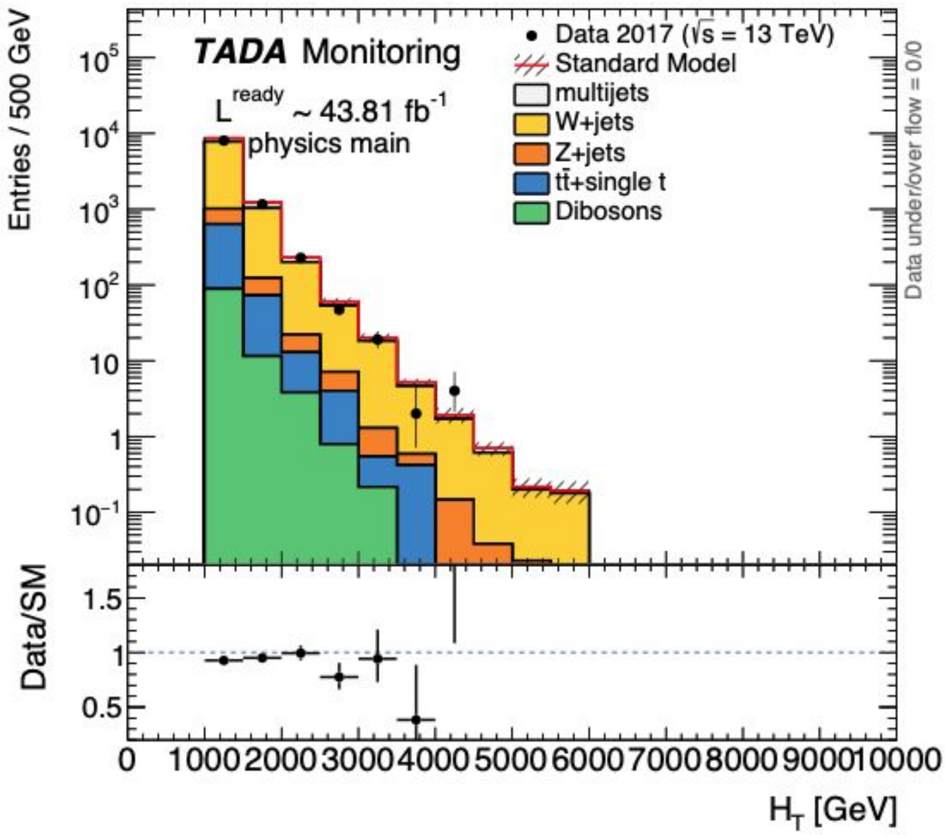
Monitoring Generic Channels

Example: leptons + jets (plot)

Number of leptons $(e, \mu) = \{1, 2\}$
Number of jets $= \{2, 3, 4, 5\}$
$H_T > \{1, 2\}$ TeV

Monitored during 2017 data-taking

- No significant excess or feature in data
- Multijet bkg difficult to model
- Luminosity doubled in the short term



Monitoring Generic Channels - conclusion

- Fast monitoring crucial during data-taking
- Generic signatures:
  - Proof of concept system
  - Easily extensible
  - Run-3?
- No systematic errors included
- No data-driven techniques for bkg. estimation

Outline

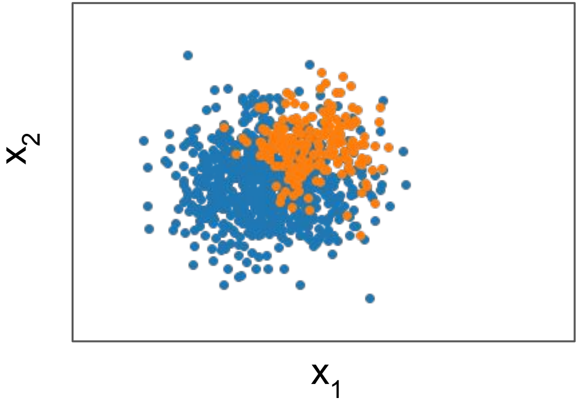
1. Searching for New Physics at the LHC
    - The Standard Model, the LHC and ATLAS
    - Model Independent searches for New Physics
    - Monitoring generic physics channels

2. **Methods for Model Independent searches for New Physics**
    - A Semi-supervised approach for anomaly detection
    - Gaussian Processes for resonance searches

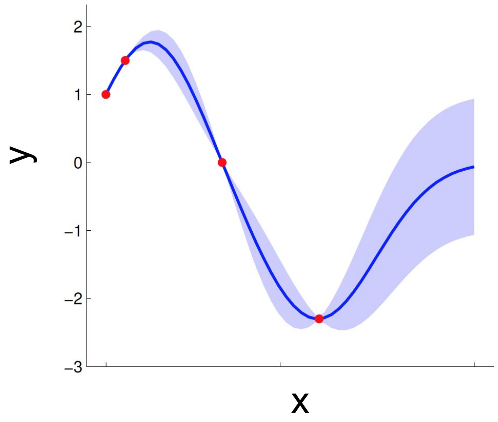
Gaussian Mixtures and Gaussian Processes

Two methods for model independent searches

Penalized Anomaly Detection
- Multiple dimensions (variable selection)
- Semi-supervised learning



Gaussian Processes for resonance searches
- Focus on one dimension
  - Smooth background + signal ID
- Resonance searches

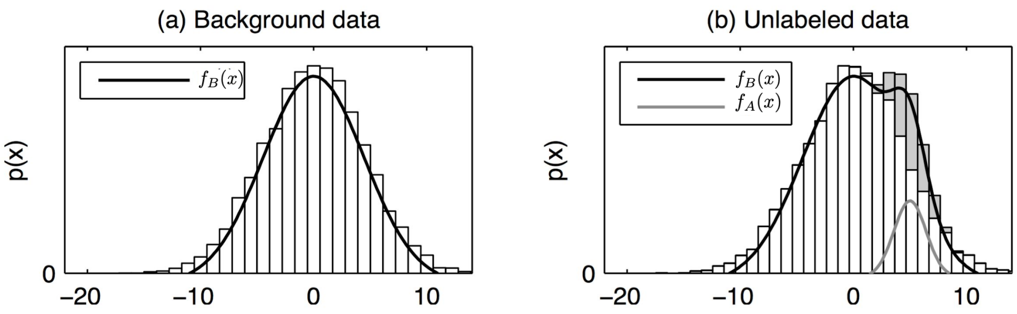


Detecting anomalies - Gaussian Mixture Models*

Fixed Background Model (FBM)

a. Learn a background model $f_B(x)$

b. Fit data keeping $f_B(x)$ fixed → $f_{FB}(x)$



mixing coefficient

$$f_{FB}(x) = (1-\lambda) f_B(x) + \lambda f_A(x)$$

Max. likelihood while keeping background fixed

$f_A(x)$ is the anomaly model

The anomaly could point to New Physics

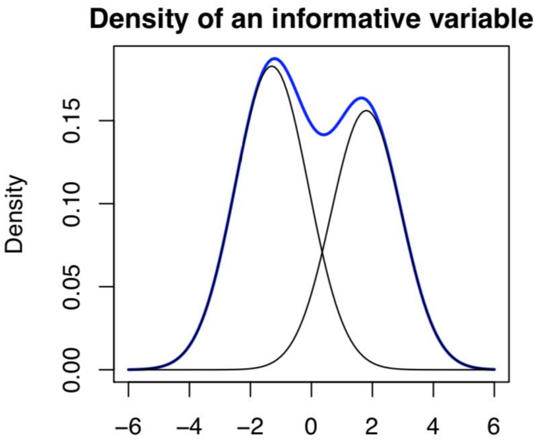
* Based on T. Vatanen, et. al. "Semi-supervised detection of collective anomalies with an application in high energy particle physics"

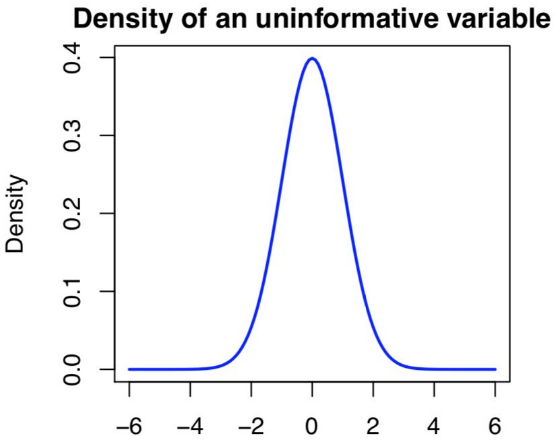
Penalized model-based clustering

GMMs are difficult to fit in high dimensions → Standard approach: Use Principal Components

Alternative: use regularization → dimensionality reduction

$$\log \mathcal{L}_p(\theta) = \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \Sigma_k\right) \right) - \gamma p(\Theta) \qquad [\Theta \subset \theta]$$





Anomaly detection
+
Variable selection
↓
Penalized Anomaly Detection (PAD)

Mean And Eigenvalue Shrinkage

$$\log \mathcal{L}_p(\boldsymbol{\theta}_B) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \Sigma_k)\right)$$

Regular GMM likelihood

$$+ \gamma_1 \sum_{j=1}^{D} \sqrt{\sum_{k=1}^{K} \pi_k \mu_{kj}^2}$$

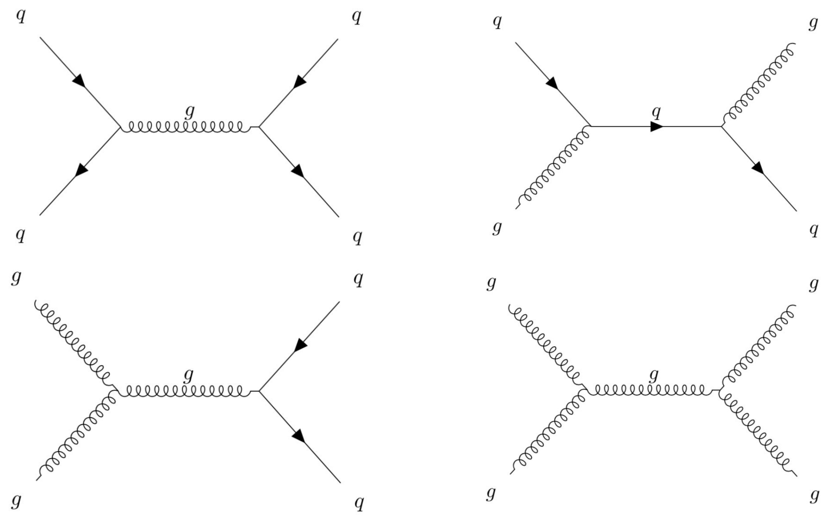
Shrinks the (squares of the) Gaussian means

$$+ \gamma_2 \sum_{k=1}^{K} \sum_{j=1}^{D} \max\left(\delta_{kj}, \epsilon_k\right)$$

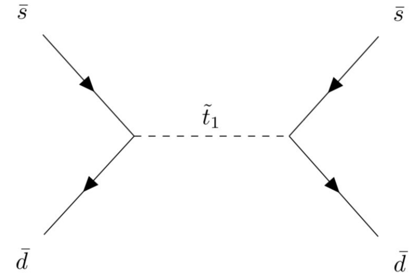
Shrinks the eigenvalue δ_k of covariance matrix Σ_k

Simple physics scenario: dijet simulation

QCD background: many ways to produce q and g



Signal: RPV-MSSM stop quark



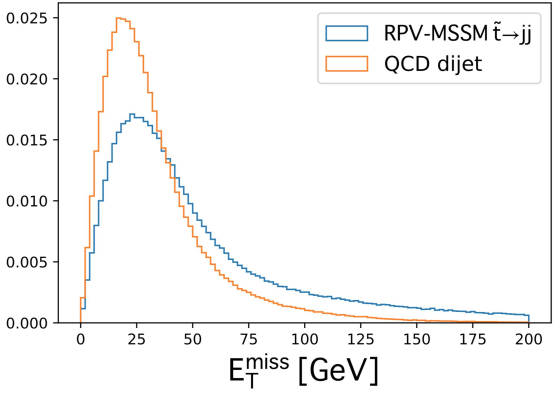
Variables extracted

Apply PAD

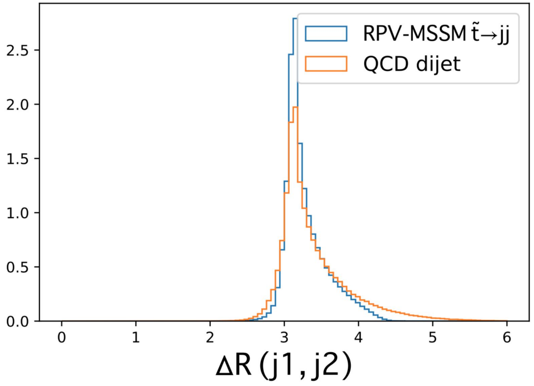
Variables

11 variables extracted from the simulation describe the physics in the event:

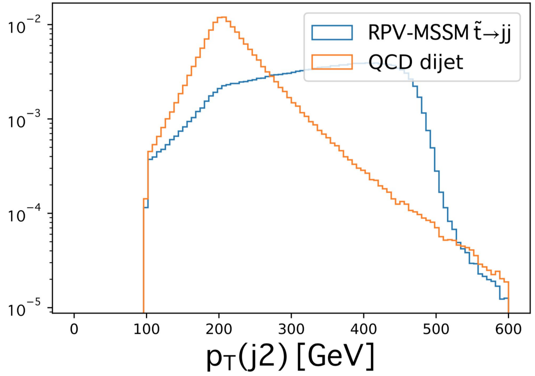
Event wide



Dijet system



Object (jet) information



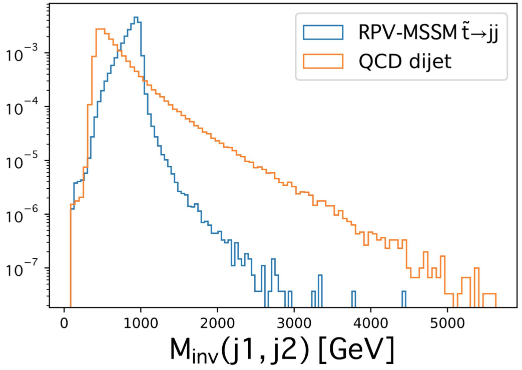
Sample preprocessing

GMMs: flexible, but skewed data require many Gaussian components

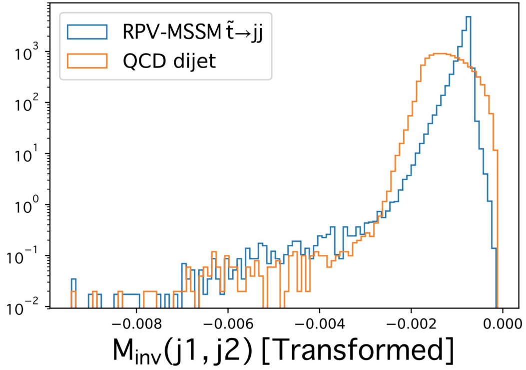
→ **Tukey ladder of powers transformation, makes distribution more Gaussian**

$$f(x) = \begin{cases} x^{\alpha}, & \text{for } \alpha > 0 \\ -x^{\alpha}, & \text{for } \alpha < 0 \\ \ln(x), & \text{for } \alpha = 0 \end{cases}$$

Example: dijet invariant mass

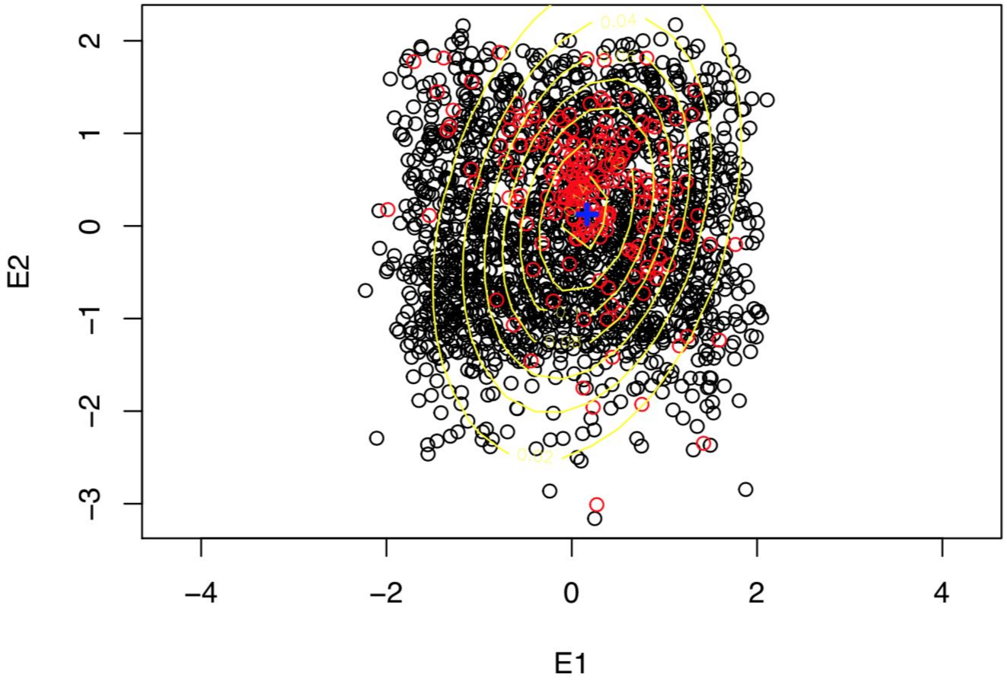


$\alpha = -1.05$ →



PAD - transformed jet energies

Data: Signal, Background **Signal model**: cross + level curves



PAD - signal extraction

Apply the method → estimate signal strength, classify observations

Method	λ	Average estimate $\hat{\lambda}$	Average AUC
PAD	0 (spurious)	0.10(0.03)	-
PAD	0.05	0.04(0.01)	0.7(0.1)
PAD	0.10	0.06(0.01)	0.81(0.01)
PAD	0.15	0.09(0.01)	0.87(0.02)
PAD	0.20	0.112(0.006)	0.88(0.01)
FBM	0 (spurious)	0.12(0.03)	-
FBM	0.05	0.025(0.009)	0.7(0.1)
FBM	0.10	0.046(0.008)	0.76(0.08)
FBM	0.15	0.070(0.006)	0.77(0.07)
FBM	0.20	0.10(0.01)	0.78(0.05)

- PAD is able to identify uninformative variables, in this case 2 (p_{T2} and E_T^{miss})
- Better performance than FBM in classification

PAD - Conclusion and Outlook

- Novel method for Collective Anomaly Detection
- Results promising but signal underestimated
- Possible improvement directions
  - Simplify pre-processing step
  - Use penalty terms added for the shrinkage of (mean and covariance) parameters
  - Consider non gaussian finite mixture models

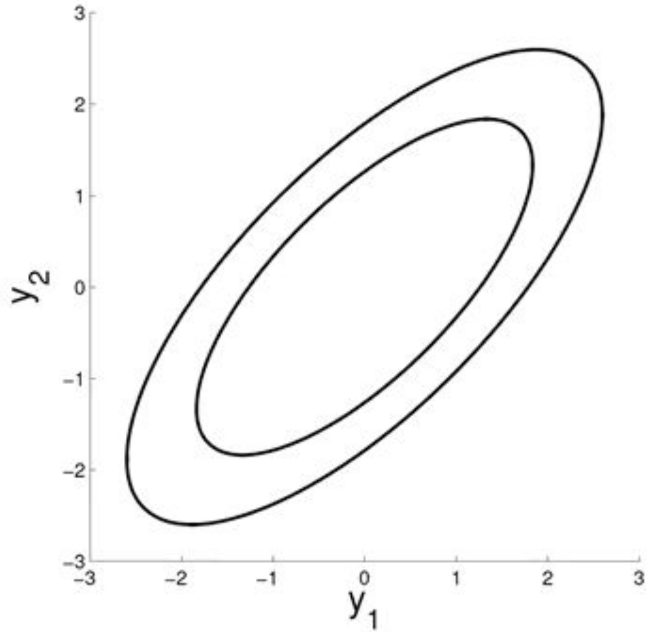
Outline

1. Searching for New Physics at the LHC
    - The Standard Model, the LHC and ATLAS
    - Model Independent searches for New Physics
    - Monitoring generic physics channels

2. Methods for Model Independent searches for New Physics
    - A Semi-supervised approach for anomaly detection
    - **Gaussian Processes for resonance searches**

Bivariate Gaussian

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^\top\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$

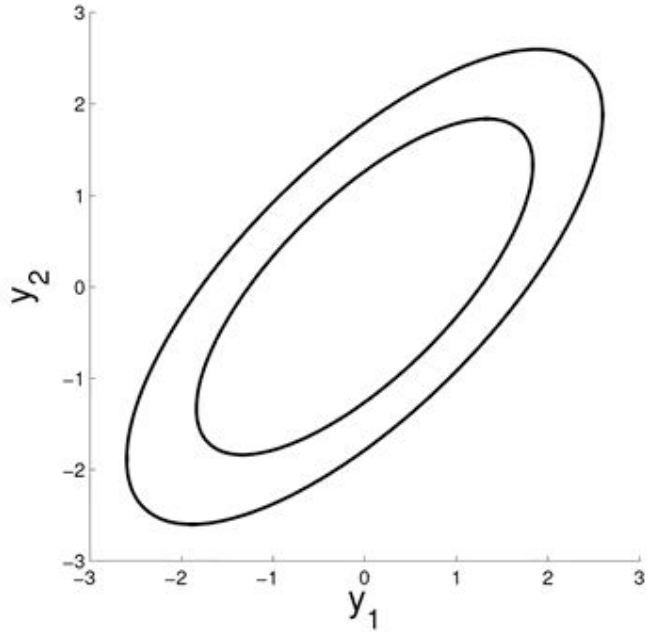


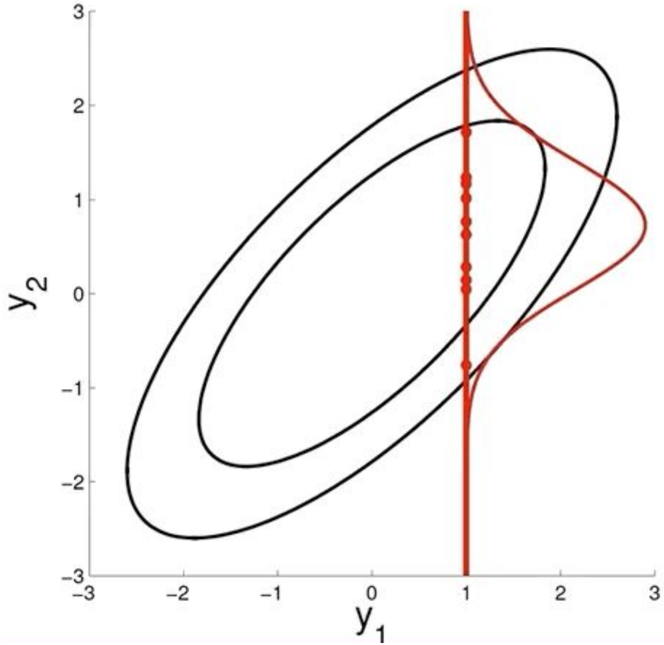
https://youtu.be/92-98SYOdlY

Bivariate Gaussian

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\tfrac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$

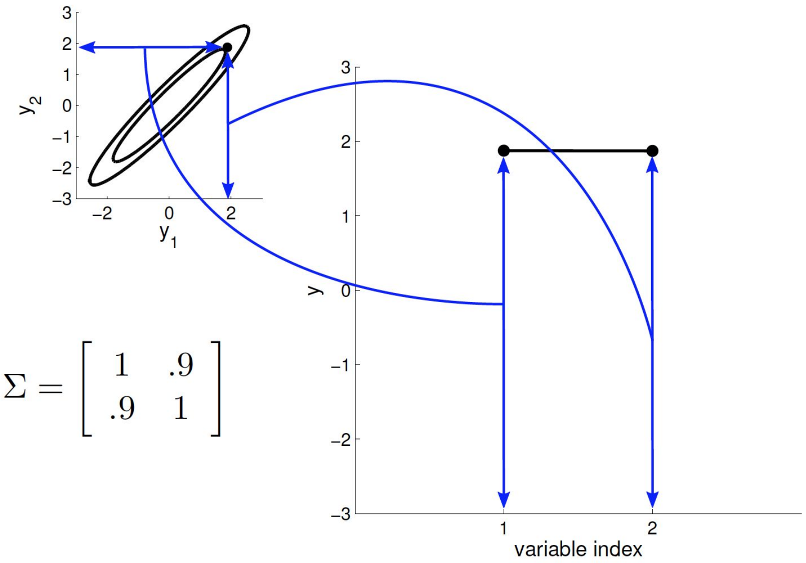
$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\tfrac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$





https://youtu.be/92-98SYOdlY

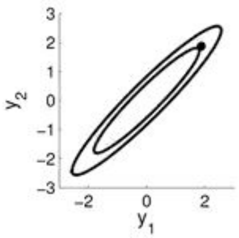
MacKay’s visualization of N-dim Gaussians



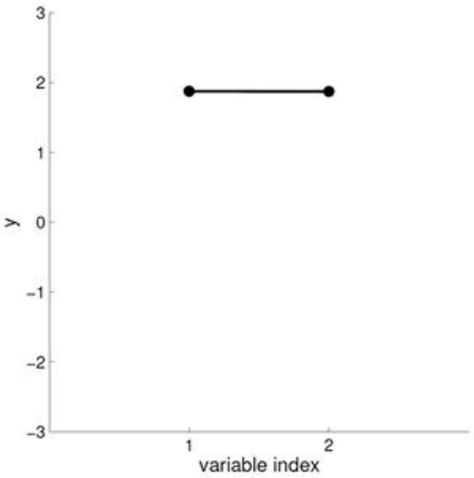
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

https://youtu.be/92-98SYOdlY

MacKay’s visualization of N-dim Gaussians

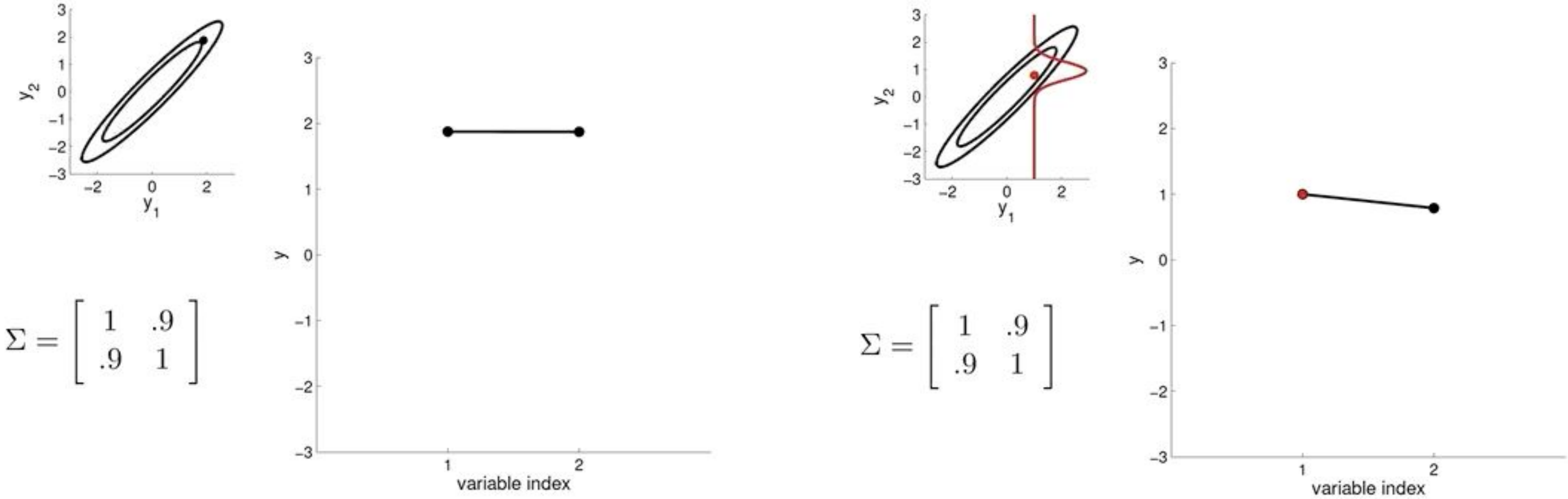


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

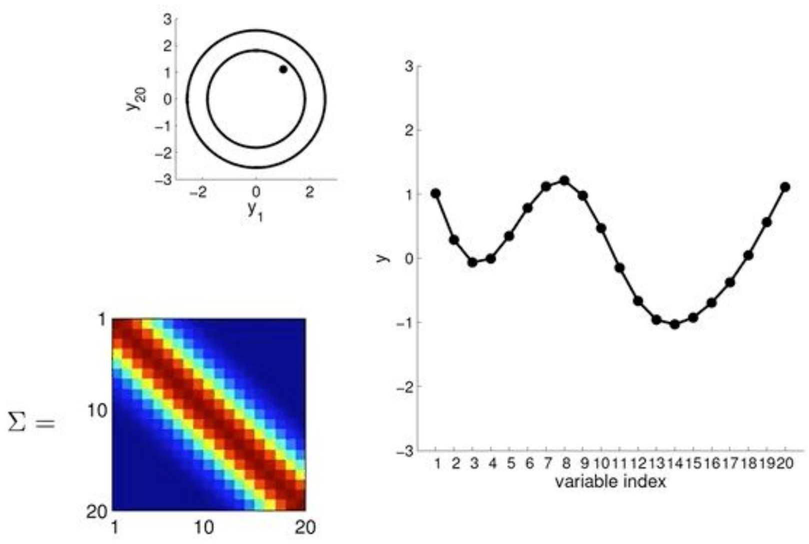


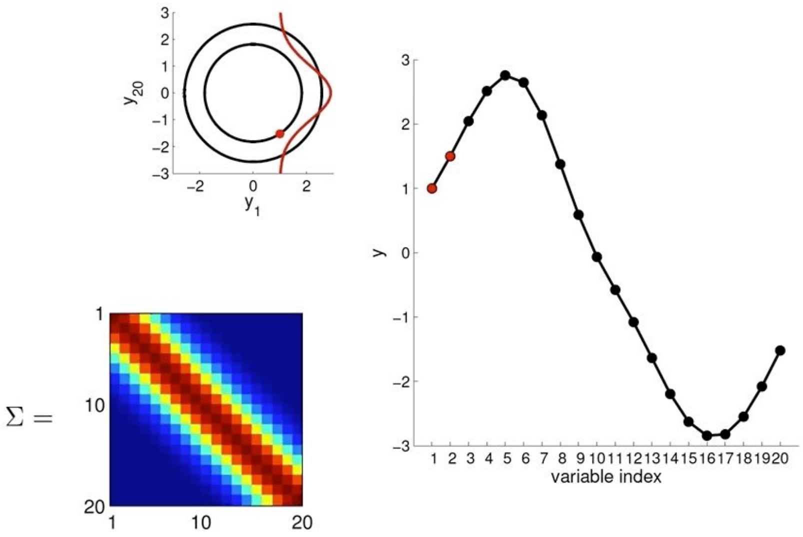
https://youtu.be/92-98SYOdlY

MacKay’s visualization of N-dim Gaussians



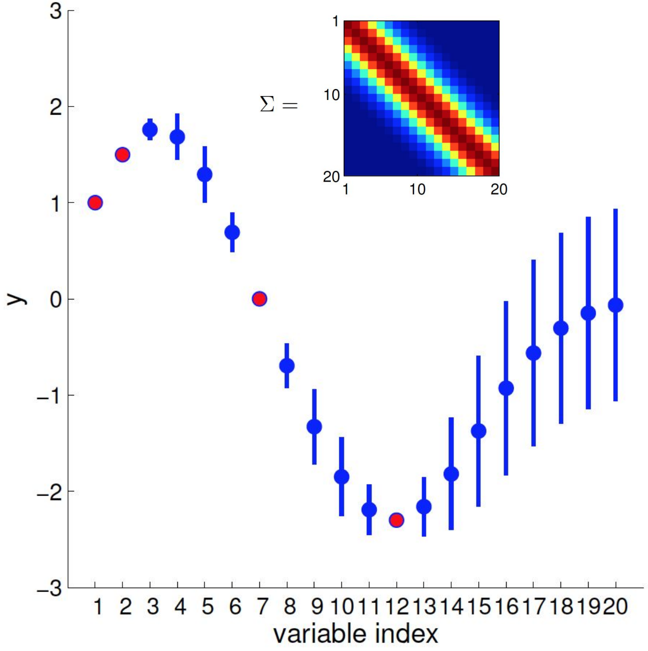
20-dimensional Gaussian

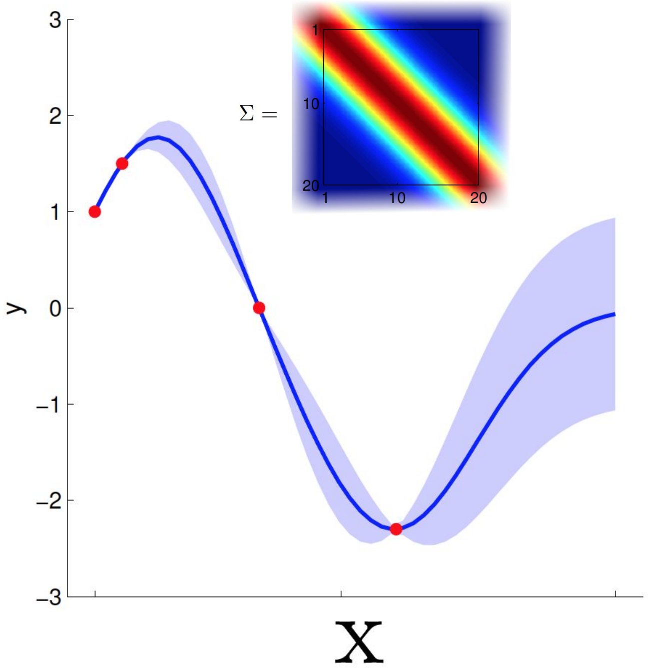




https://youtu.be/92-98SYOdlY

Infinite dimensions - Gaussian Process (GP)





https://youtu.be/92-98SYOdlY

Gaussian Processes (GPs)

GP: associate a multivariate gaussian distribution to a set of random variables

→ The gaussian will have as many dimensions as random variables we have

A set of N values (bin counts) **y** can be associated with

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \sim \mathrm{Gaus}(\boldsymbol{\mu}, \Sigma)$$

Infer new values y_* by extending (the dim. of) the Gaussian distribution

Use a *kernel* or *measure of similarity* between points (bin centers) and a *mean function*

→ A kernel example is the exponential squared

$$k(x_i, x_j) = A \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$$

where A and l are (hyper)parameters to be fixed

Gaussian Processes (GPs)

Infer a new value y_* located in x_*, using the following

$$p(\boldsymbol{y}_*|\boldsymbol{x}_*, \boldsymbol{x}, \boldsymbol{y}) = \mathrm{Gaus}(\boldsymbol{y}_*|\boldsymbol{\mu}_*, \Sigma_*)$$

$$\boldsymbol{\mu}_* = m(\boldsymbol{x}_*) + \boldsymbol{K}_*^T \Sigma^{-1}(\boldsymbol{y} - m(\boldsymbol{x}))$$

$$\Sigma_* = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T \Sigma^{-1} \boldsymbol{K}_*$$

With $\boldsymbol{K}_* = k(\boldsymbol{x}, \boldsymbol{x}_*)$, $\boldsymbol{K}_{**} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$

Note:

- Kernel hyperparameters are optimized using e.g. Maximum Likelihood
- GPs are flexible enough to model the mean of the distribution having m(x) = 0

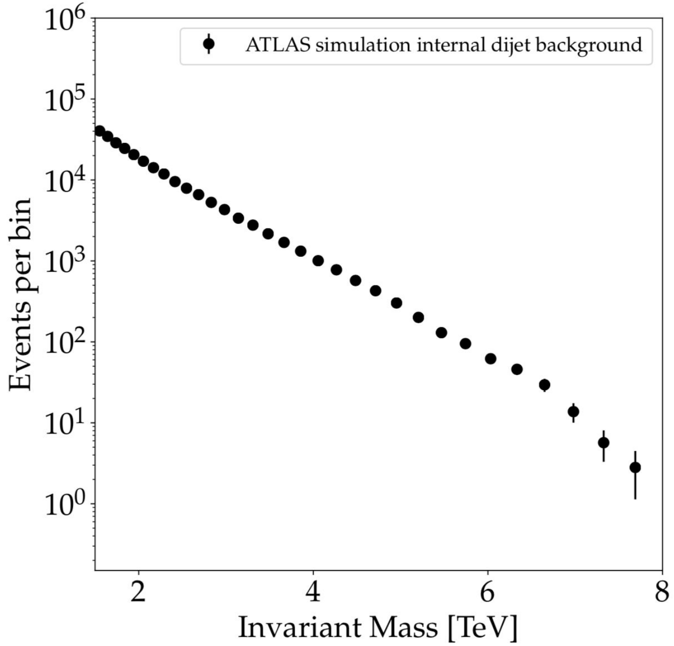
Two use cases in invariant mass spectra:

two jets & jets and leptons (top pair)

Here we use the george package for GP inference and optimization

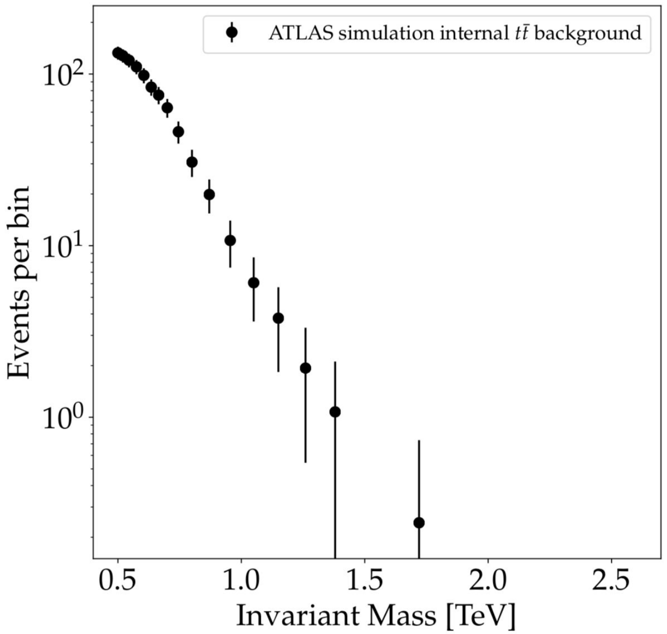
Mass spectra used

Dijet signature from General Search Analysis



[Aaboud, M., Aad, G., Abbott, B. et al. Eur. Phys. J. C (2019) 79: 120.]

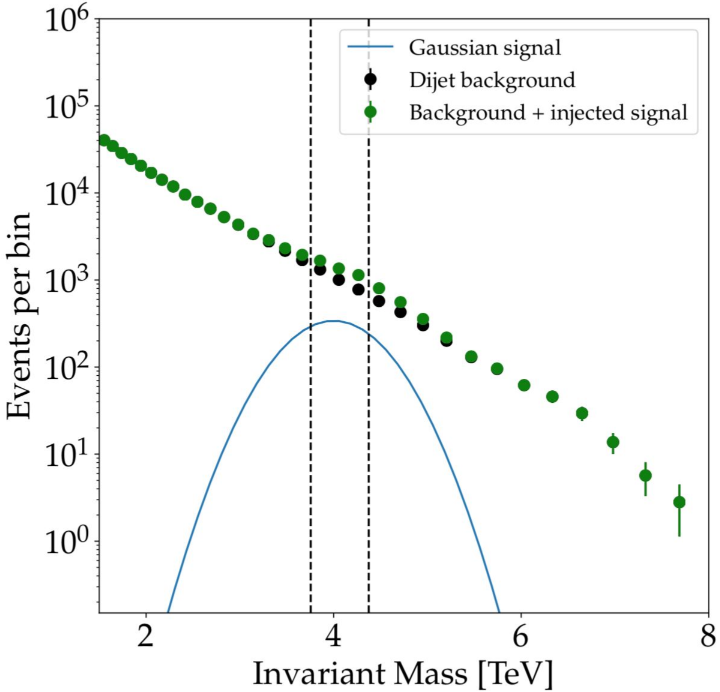
Lepton+jets signature from X→tt Search



[Aaboud, M., Aad, G., Abbott, B. et al. Eur. Phys. J. C (2018) 78: 565.]

Injecting signals

Construct a window (dashed) around signal mean

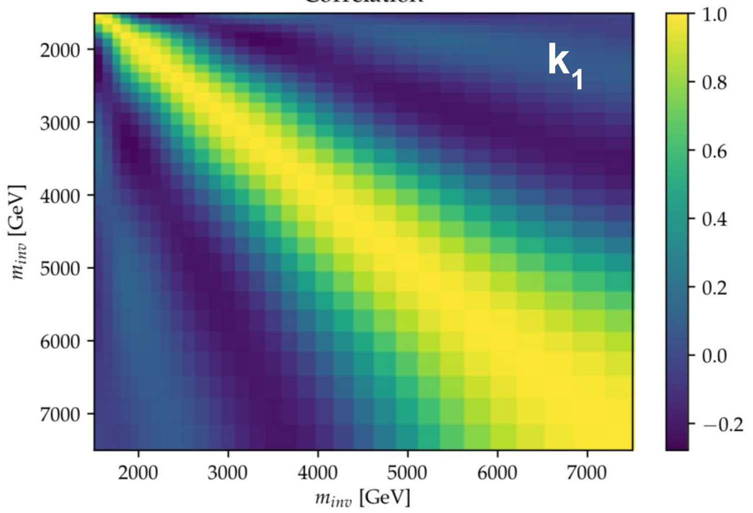


$$R = \frac{\text{Injected signal events in the window}}{\text{Background events in the window}}$$

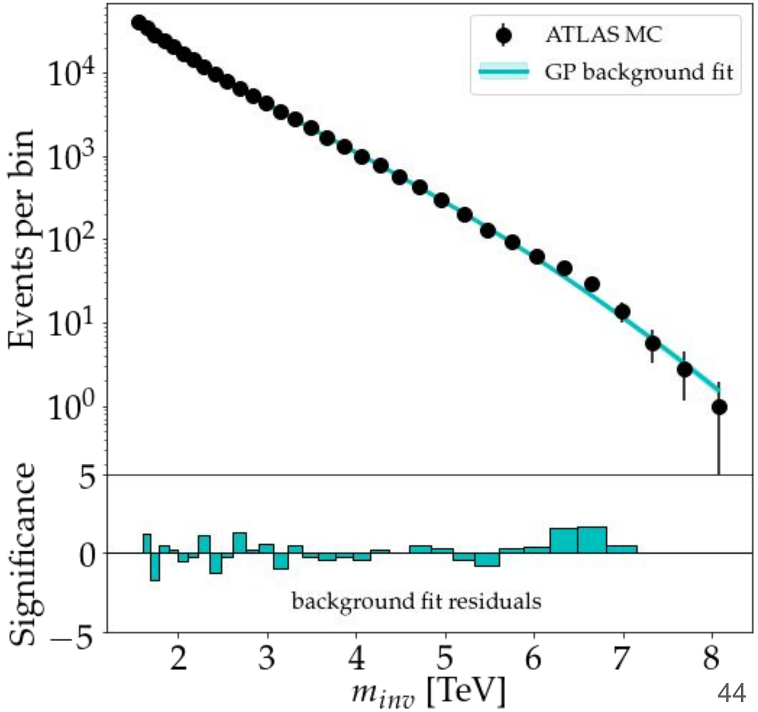
2-step procedure in the dijet spectrum*

1. GP fit for SM background
   → Use background kernel $\mathbf{k_1}$ and mean $\mathbf{m(x) = 0}$

$$A\ \exp\left(\frac{d-(x+x')}{2a}\right)\sqrt{\frac{2l(x)l(x')}{l(x)^2+l(x')^2}}\exp\left(\frac{-(x-x')^2}{l(x)^2+l(x')^2}\right).$$



Dijet data from the 2015 General Search analysis (3.2/fb)



* Based on 1709.05681

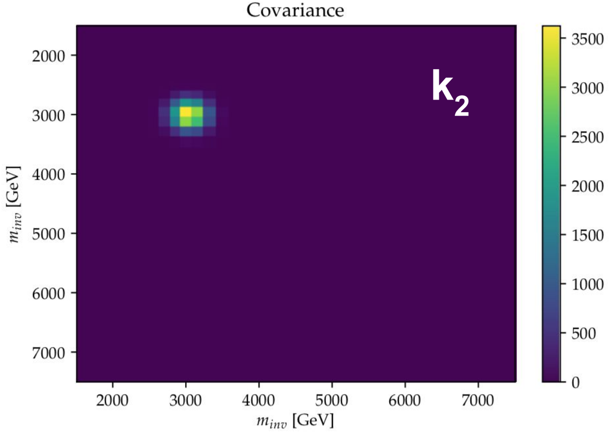
2-step procedure*

2. GP fit using background pseudodata and injected signal

$$A_S \ \exp\left(-\frac{1}{2}\left(x-x'\right)^2/l^2\right)\exp\left(-\frac{1}{2}\left((x-m)^2+\left(x'-m\right)^2\right)/t^2\right).$$

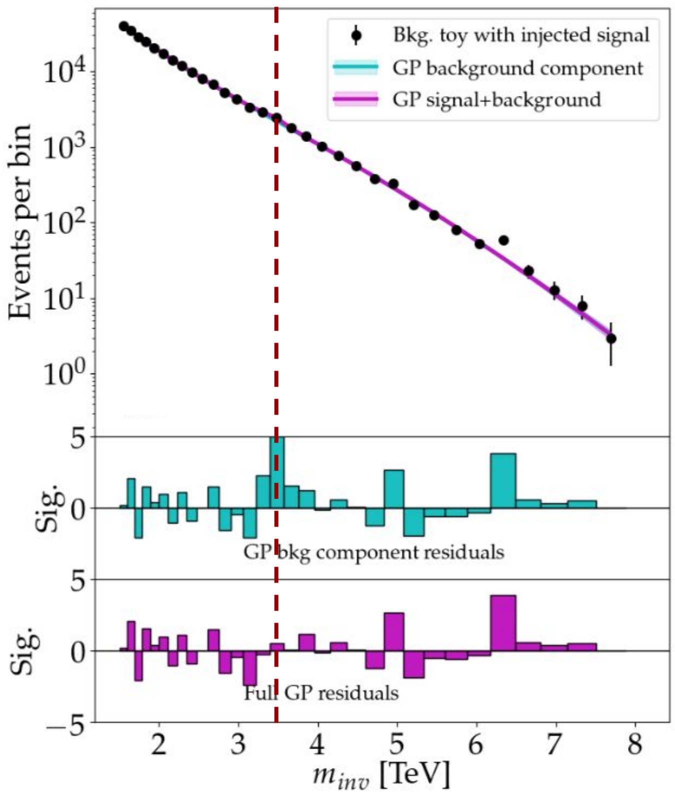
→ Use background + signal kernel → $k_1 + k_2$

→ Keep bkg. kernel parameters frozen (step 1)



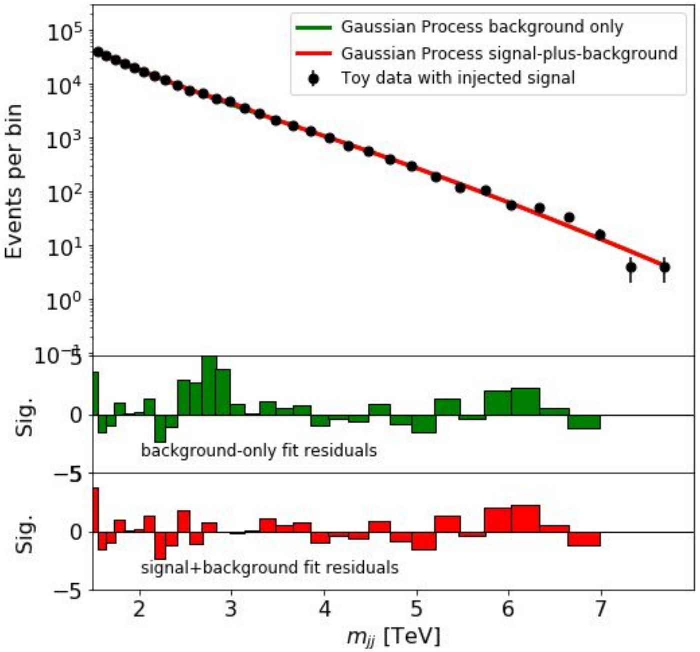
Extract signal parameters from signal kernel

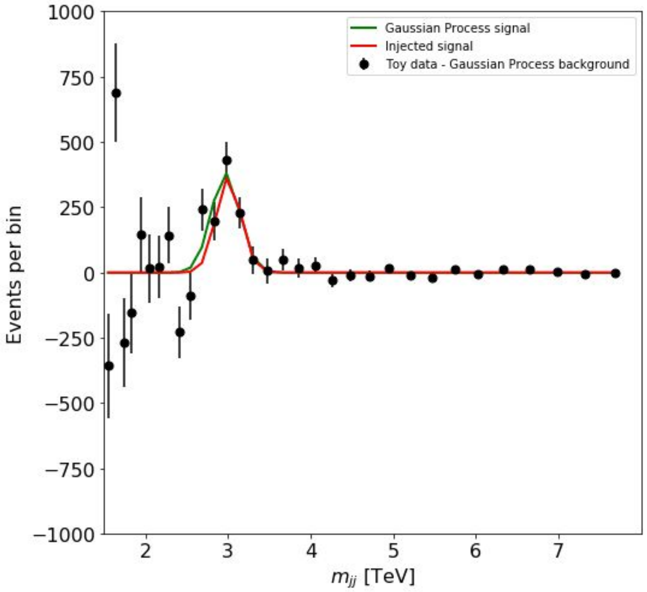
Dijet data from the 2015 General Search analysis (3.2/fb)



* Based on 1709.05681

2-jet mass spectrum with signal injected



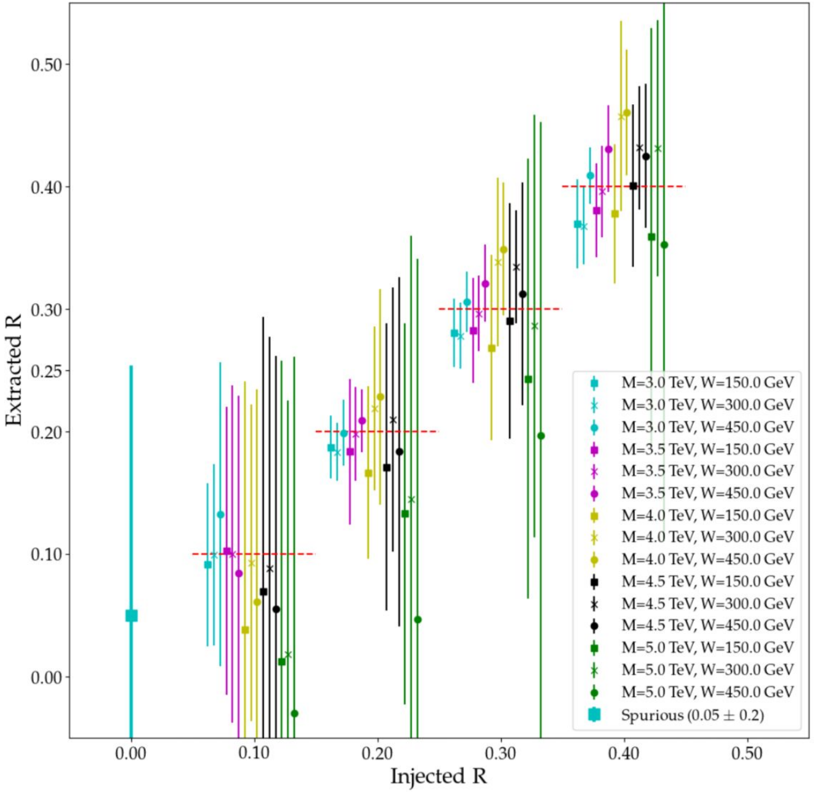


Left - residuals:
Data + signal vs. bkg GP
Data vs. bkg GP

Right:
Extracted signal
→ Final sig kernel hyperparams

Tested 60 signal hypotheses with 100 toys each

Testing linearity



Points and errors extracted:
→ the mean and deviation from values (all toys)

Extract information from spurious signal fitting

Further tests of this GP method:

- Variations of the 2-step procedure
- 3-step procedure for m_{tt} spectrum

https://cds.cern.ch/record/2652616?

Two-step procedure in the m_{tt} spectrum

Case: mass spectrum from ([1804.10823](1804.10823))

Search for X→ top pair (signature w/ jets and leptons)

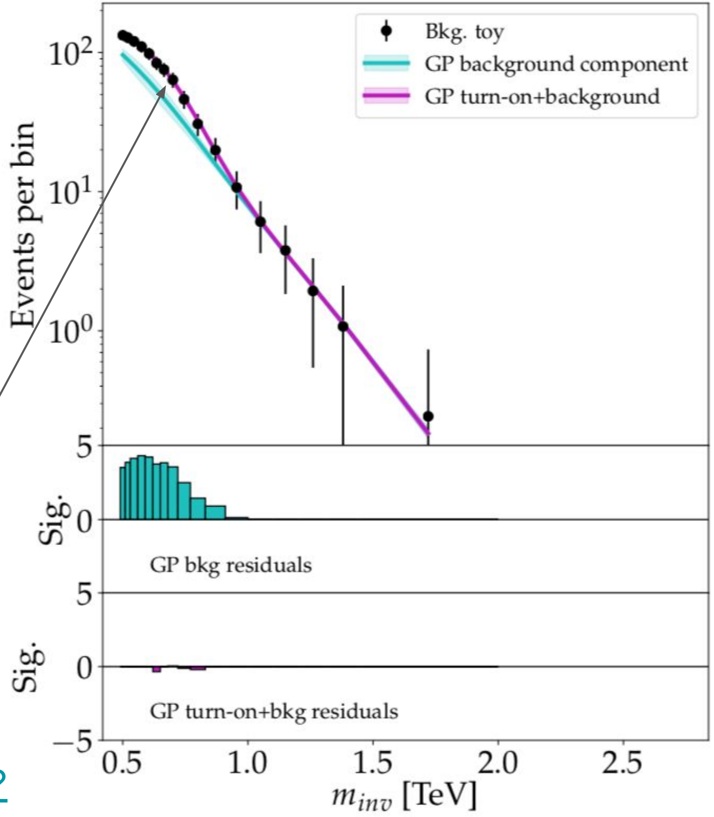
One top decays hadronically

Another into b-jet and muon+neutrino

ATLAS simulations (36.1 /fb analysis)

The background spectrum presents a turn-on in the low-mass region

Kernel in the second step (k_2) accommodates turn on



https://cds.cern.ch/record/2652616?

Three-step procedure

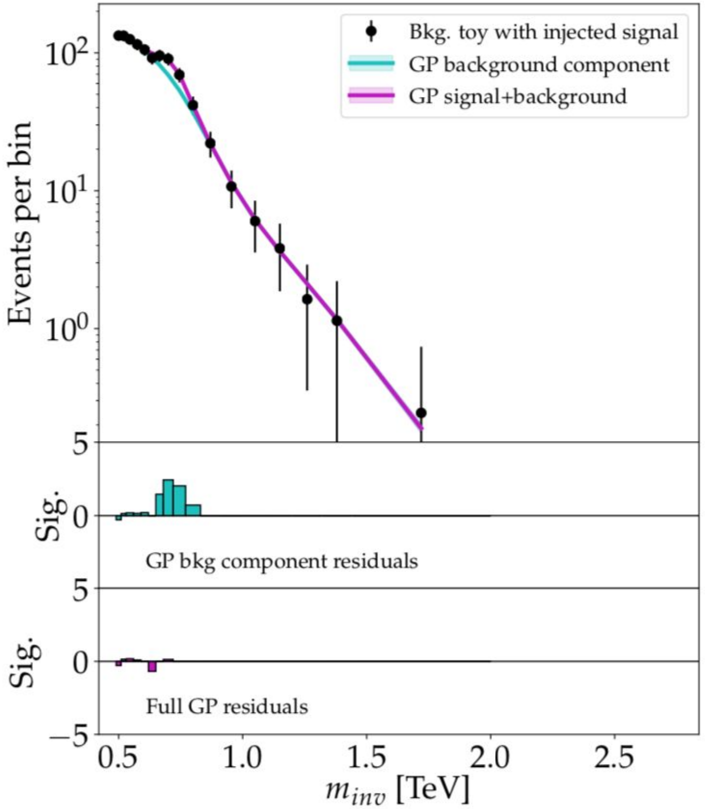
1. On background data: use $\mathbf{k} = \mathbf{k}_1$
2. On same background data, use $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$ (keep $\mathbf{k}_1$ parameters fixed)
3. On background+signal data: use $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ (params $\mathbf{k}_1$, $\mathbf{k}_2$ fixed)
   Where $\mathbf{k}_3$ is the same as $\mathbf{k}_2$ with e.g. a constraint on small widths

Three-step procedure in the m_{tt} spectrum

Third step is used for signal extraction

Signal benchmarks
Z' @ 750 GeV with an amplitude of 1.85 pb (right plot)
Z' @ 1250 GeV with an amplitude of 1 pb

Note: Need to amplify the signal to achieve detection



Three-step procedure in the m_{tt} spectrum

Hypothesis (factor)	injected R	m [GeV]	R
No signal	0	610 ± 50	0.04 ± 0.05
750 GeV (5)	0.1	630 ± 20	0.16 ± 0.06
750 GeV (10)	0.2	680 ± 20	0.36 ± 0.08
750 GeV (15)	0.3	700 ± 10	0.39 ± 0.07
1250 GeV (5)	0.18	900 ± 100	0.01 ± 0.01
1250 GeV (10)	0.35	1000 ± 100	0.19 ± 0.07
1250 GeV (15)	0.53	1110 ± 30	0.35 ± 0.07

GPs in resonance searches - conclusion+outlook

- Method working in two different mass spectra
  - Background modelling
  - Signal extraction possible (biased)
- Future improvements
  - Towards data-driven background estimation
  - Choice of kernels
  - Automated procedure for other signatures

Conclusions

- Abundance of data at the LHC and upgrade for probing the Standard Model and beyond

- Time is ripe for devoting effort in developing Model-Independent methods
    - Methods presented in this work plus several other
    - Profit from advances in the Machine Learning community

This Report is part of a project that has received funding from the **European Union's Horizon 2020 research and innovation programme under grant agreement Nº675440**

Thank you!

Backup

General Search approach

HERA proposed a 1-D (one variable in a class) search algorithm, used in ATLAS

Calculate the value of the estimator **p** given by

$$p = \begin{cases} A\int_0^\infty db\,\boxed{G(b, N_{SM}, \delta N_{SM})}\,\boxed{\sum_{i=N_{obs}}^{\infty} \frac{e^{-b} b^i}{i!}} & N_{obs} > N_{SM} \\ A\int_0^\infty db\,\boxed{G(b, N_{SM}, \delta N_{SM})}\,\boxed{\sum_{i=0}^{N_{obs}} \frac{e^{-b} b^i}{i!}} & N_{obs} < N_{SM} \end{cases}$$

Gaussian pdf with mean N_{SM} and width δN_{SM} * Poisson pdf

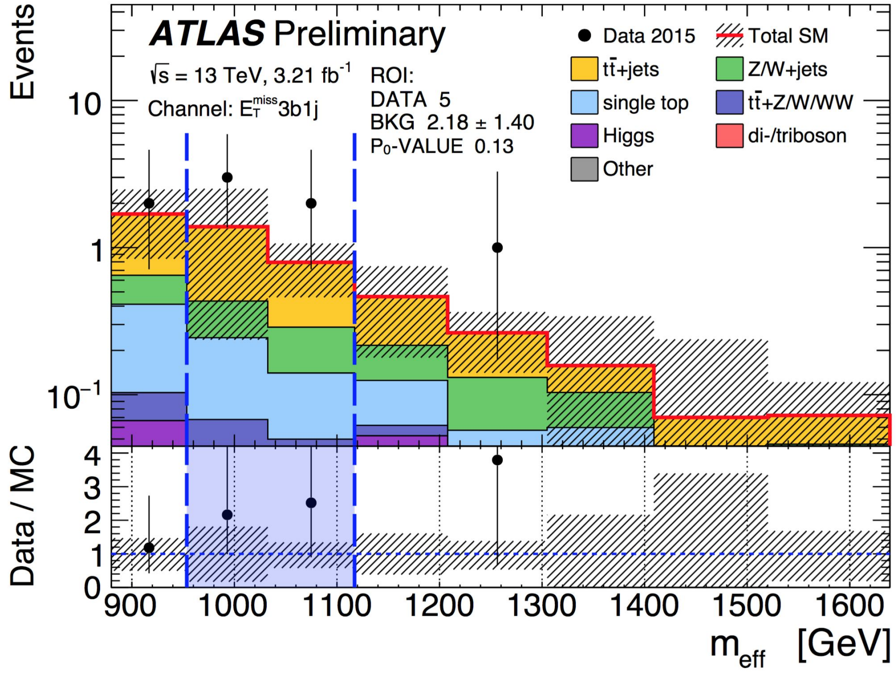
in all possible regions (connected bins), with:

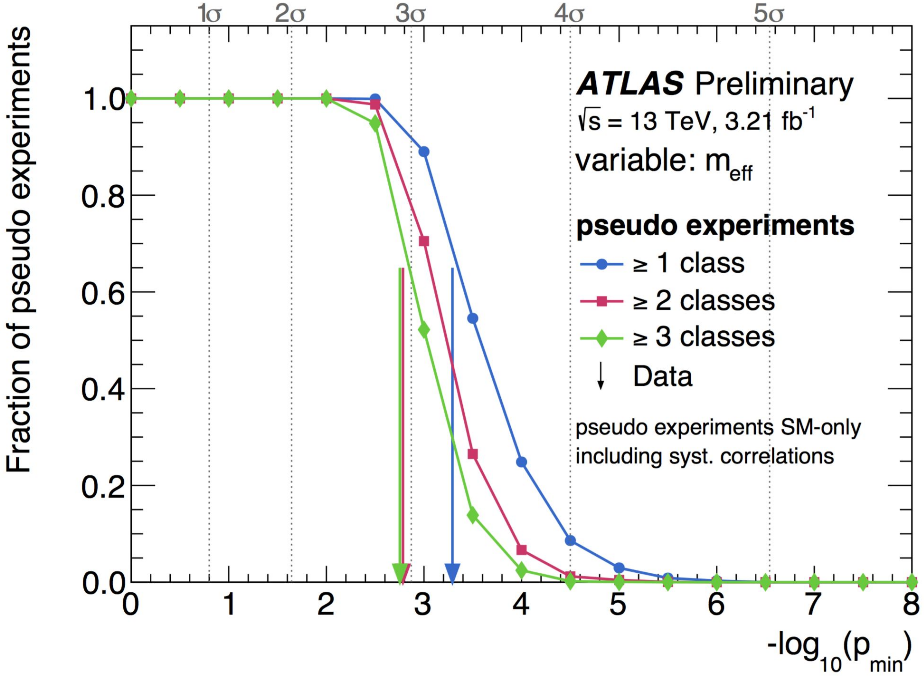
- N_{SM} = expected number of events
- δN_{SM} = systematic unc. on N_{SM}
- N_{obs} = observed number of events
- A = norm. constant

Probability of a N_{SM} fluctuation as extreme as N_{obs} in the region

p → local p-value

General Search @ ATLAS





TADA: format, data + simulations

TAG: Condensed data format produced at ATLAS Tier-0

Leading: 6 e, 6 μ, 4 τ, 4 γ, 10 jets → For each object: kinematics, info on PID, isolation, quality

E_T^{miss} info, trigger counters, global event info

(Object selection + overlap removal → backup)

MC simulations (detail in backup)

- Top pairs + single top
- Dibosons
- W/Z+jets
- Multijets
- Diphoton

Data

- pp collisions @ √s = 13 TeV
- Period monitored: 2017 (43.8 / fb)
- Athena release 21
- Updated twice a day when runs available

Object selection requirements in TADA

Object	p_T	$\lvert\eta\rvert$	Other requirements
e	> 10 GeV	$\in (0, 1.37) \cup (1.52, 2.47)$	ElectronIDLikelihoodLoose
μ	> 10 GeV	< 2.7	isCombined LooseID Cosmics veto Reject second muons with $dR < 0.01$
γ	> 20 GeV	$\in (0, 1.37) \cup (1.52, 2.37)$	PhotonIDLoose
(b-)jet	> 40 GeV	< 2.8	AntiKt4TopoJets LooseBadTool (`mv2c10` b-jet tagger)
τ	> 20 GeV	$\in (0, 1.37) \cup (1.52, 2.5)$	JetBDTSigMedium

Overlap removal in TADA

Rank	Overlap removal	separation
1	remove jets overlapping with electrons	$dR < 0.2$
2	remove taus overlapping with muons	$dR < 0.2$
3	remove jets overlapping with taus	$dR < 0.4$
4	remove electrons overlapping with jets	$dR < 0.4$
5	remove muons overlapping with jets	$dR < 0.4$
6	remove photons overlapping with electrons	$dR < 0.2$
7	remove jets overlapping with photons	$dR < 0.4$

MC simulations for TADA

- $t\bar{t}$+single-t. Samples for top quark pairs ($t\bar{t}$) were produced using Powheg [48, 49] (limiting the `hdamp` parameter to 1.5 times the top mass), Pythia 8 [50, 51] (using the A14 tune [52] and `nnpdf23` [53] at leading-order (LO)) and EvtGen [54]. Single top-quark (single-t) samples were generated using Powheg [48, 49], Pythia 6 [50] (with the Perugia 2012 tune [55]) and EvtGen [54].

- Dibosons. These samples, corresponding to processes generating WW, ZW or ZZ, were generated using Sherpa 2.2.1 [56] using `nnpdf30` at next-to-next-to-leading-order (NNLO).

- W/Z+jets. The processes corresponding to final states with a weak boson plus jets, were also generated with Sherpa 2.2.1 [56] using `nnpdf30` at NNLO for leptonic decays and Sherpa 2.1.1 and the `CT10` pdf [57].

- Multijets. These were dijet samples generated with Pythia 8 [50, 51] using the A14 tune [52] and `nnpdf23` [53] LO and EvtGen [54]. Multijet samples are the combination of samples generated at different ranges of the leading jet p_T value (known as *slices*).

Monitoring generic channels

Group	# Selections	Variables
Leptons plus jets	16	Number of leptons $(e, \mu) = \{1,2\}$ Number of jets = $\{2,3,4,5\}$ $H_T > \{1,2\}$ TeV

```
    'Selection': '''
        ( TrigEmuOneMu || TrigEmuOneEl) &&
        ( IsGoodJetMET ) &&
        ( (NLooseElectron + NLooseMuon) == {nlep} ) &&
        ( LooseElectronPt1 > 26000 || LooseMuonPt1 > 26000 ) &&
        ( NJet == {njet} ) &&
        ( HT >= {ht} )
    ''',
    'SelectionFunc': [require_good_lepton, require_good_jet],
    'Variables': {
        'nlep' : [1,2],
        'njet' : [2,3,4,5],
        'ht'   : [1000*_GeV, 2000*_GeV],
        }
```

ML in model independent NP searches

A number of applications:

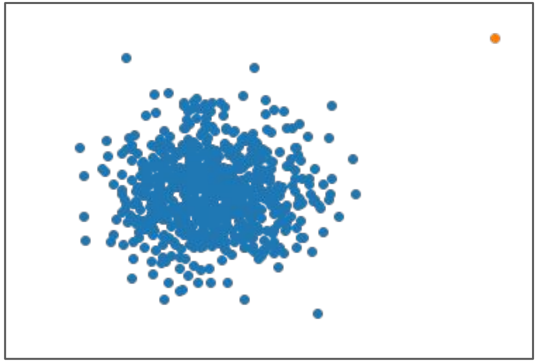
- Variational Autoencoders for outlier detection
- Compare B and S+B samples using Nearest Neighbors and Kullback-Leibler divergence
- Neural Networks as universal approximators for comparing B and S+B samples
- Using auxiliary measurements to improve Bump Hunting

Anomalies

Departure from some "normal" behavior in data

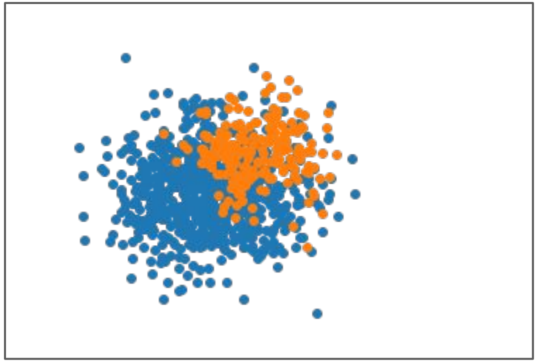
Point anomaly

Observation* that differs from other



Collective anomaly

Individual observations that are not (necessarily) anomalous, but a set of which is unusual



* Observation ↔ LHC collision event

Expectation-Maximization (EM) for GMM

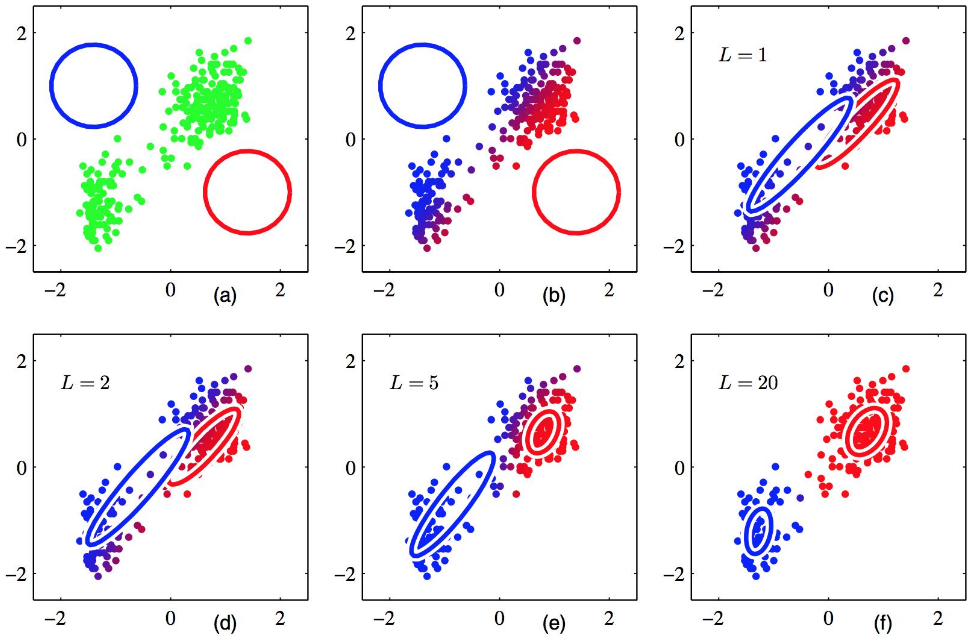
Start with a mixture of J multivariate Gaussian pdf

$$f(x|\theta) = \sum_{j=1}^{J} \pi_j \mathcal{N}\left(x_i|\mu_j, \Sigma_j\right)$$

At each iteration, there are two steps (backup):

- E - expectation: assign prob. to data
- M - maximization: update model

⇒ Increase likelihood

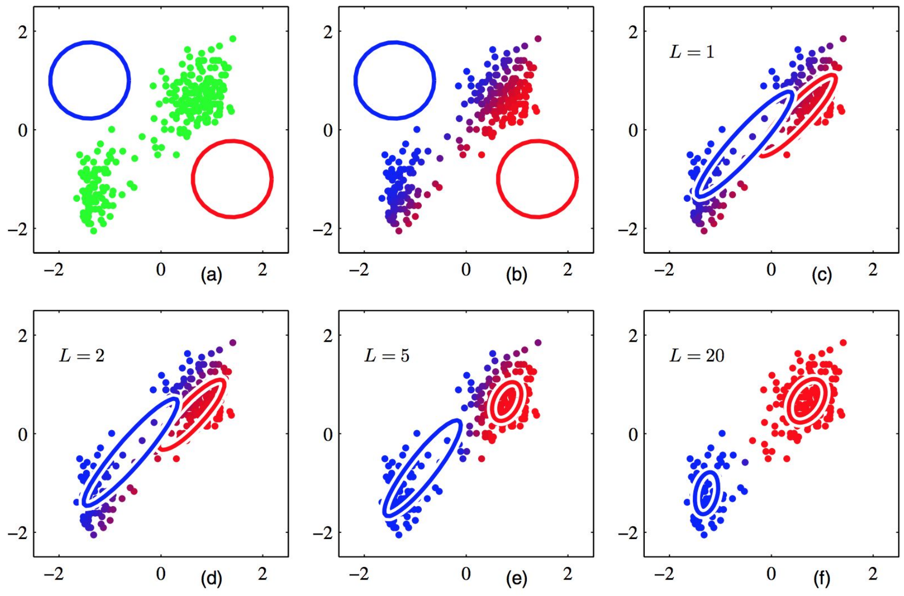


Bishop “Pattern recognition and machine learning,” Figure 9.8

Extend this to fit the anomaly model (two step procedure)

Expectation-Maximization (EM) algorithm for GMM

From Bishop's "Pattern recognition and machine learning," Figure 9.8



Start with a mixture of J multivariate Gaussian pdfs:

$$f(x|\theta) = \sum_{j=1}^{J} \pi_j \phi(x_i|\mu_j, \Sigma_j)$$

θ is the set of parameters π_j, μ_j, Σ_j ; ϕ is a Gaussian

At each iteration, there are two steps (see backup):

- **E - expectation step:**

  Prob. for each point x_i to have been generated by the jth Gaussian

- **M - maximization step:**

  Update the θ values using the probability from E step

Explore θ values until a local maximum of the (log) likelihood $l(\theta) = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{J} \pi_j \phi(x_i|\mu_j, \Sigma_j)\right)$ is found

Extend this to fit the anomaly model f_A

Expectation-Maximization (EM) algorithm for GMM

Start with a mixture of J multivariate Gaussian distributions:

$$p(x|\theta) = \sum_{j=1}^{J} \pi_j \phi(x_i|\mu_j, \Sigma_j)$$

where θ is the set of parameters π_j, μ_j, Σ_j

Expectation step (kth iteration):
Compute prob for each point x_i to have been generated by the jth Gaussian

$$p(z_{ij}|x_i, \theta^k) = \frac{\pi_j^k \phi(x_i|\mu_j^k, \Sigma_j^k)}{\sum_{j'=1}^{J} \pi_{j'}^k \phi(x_i|\mu_{j'}^k, \Sigma_{j'}^k)} \equiv \gamma_{ij}^k$$

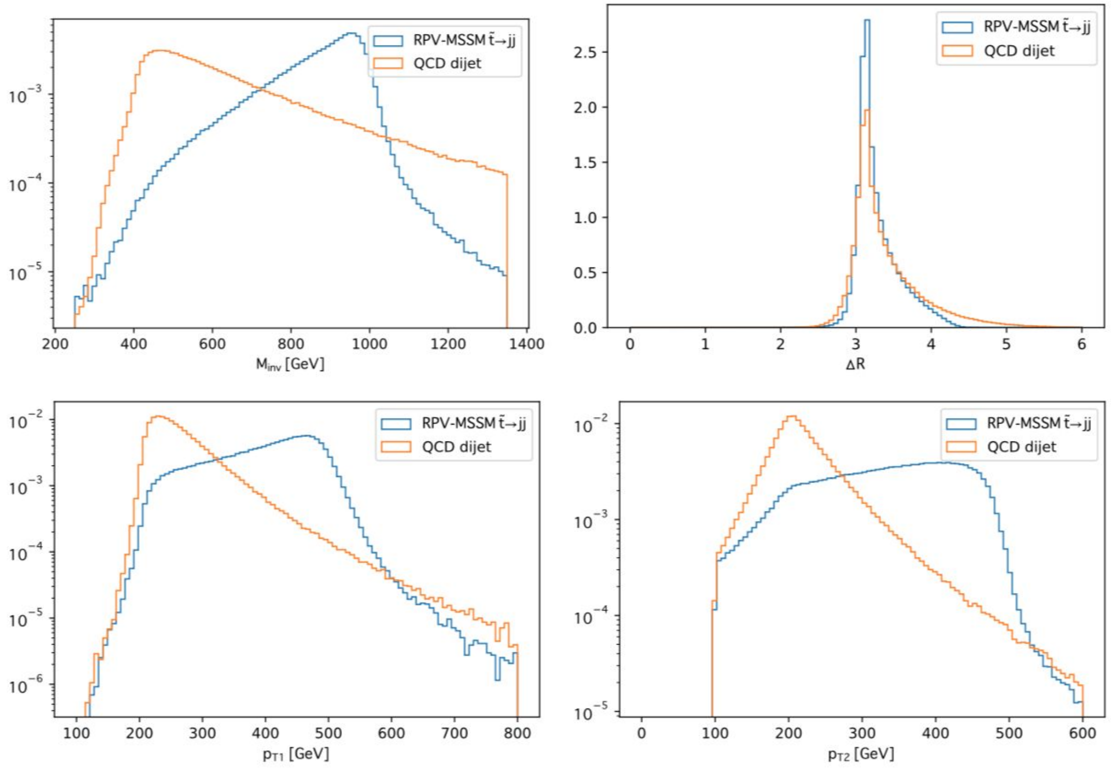
Maximization step: update the values

$$\pi_j^{k+1} = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ij}^k, \qquad \mu_j^{k+1} = \frac{\sum_{i=1}^{N} \gamma_{ij}^k x_i}{\sum_{i=1}^{N} \gamma_{ij}^k}$$

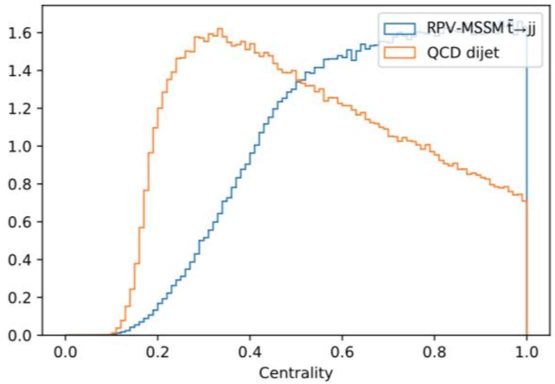
$$\Sigma_j^{k+1} = \frac{\sum_{i=1}^{N} \gamma_{ij}^k (x_i - \mu_j^{k+1})(x_i - \mu_j^{k+1})^T}{\sum_{i=1}^{N} \gamma_{ij}^k}$$

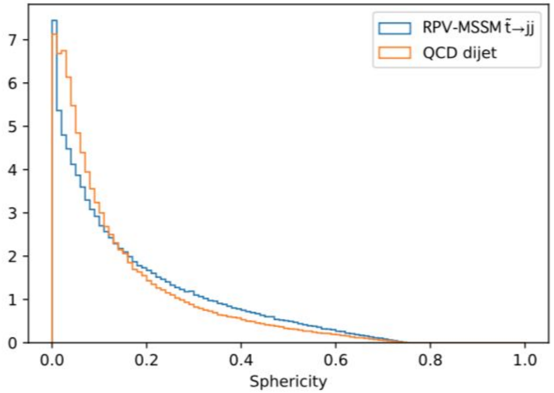
EM algo. increases the (log) likelihood $l(\theta) = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{J} \pi_j \phi(x_i|\mu_j, \Sigma_j)\right)$ until a local minimum is found

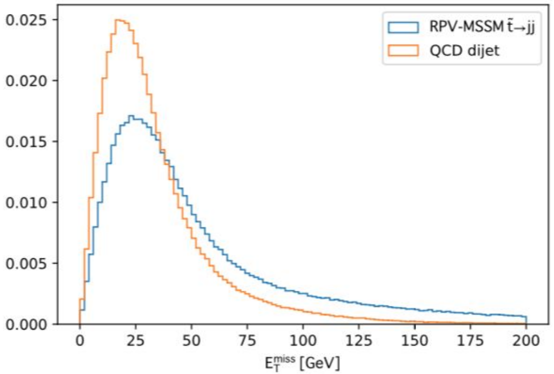
Dijet sample



Dijet sample

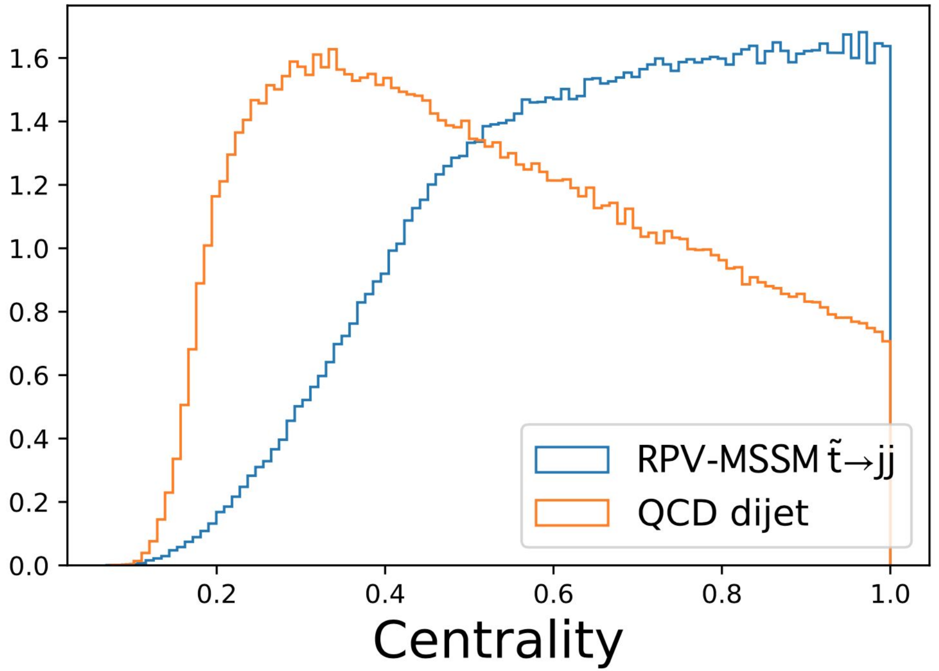


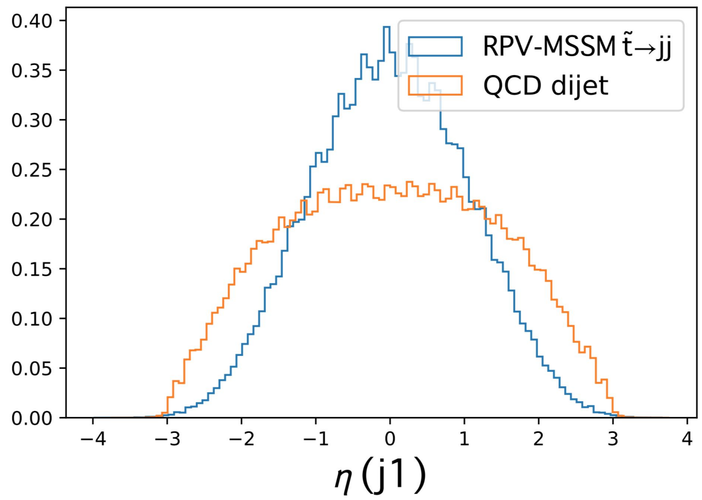


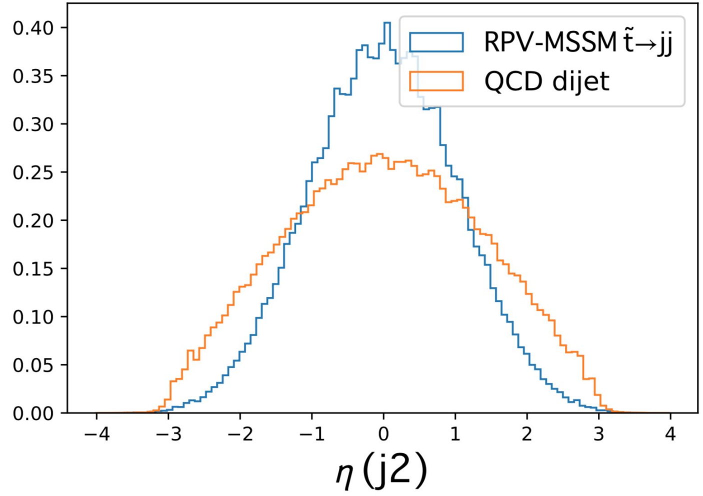


Centrality in simulated dijet

$$C = \frac{\sum_i E_{Ti}}{\sum_i E_i}$$





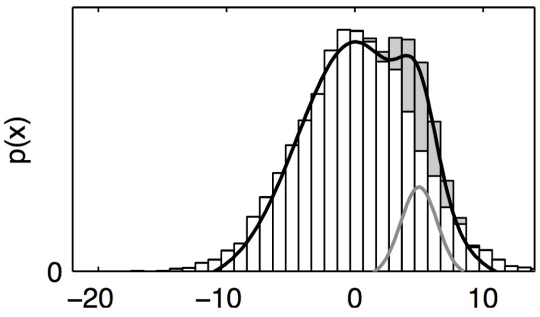


F

Gaussian Mixture Models & Gaussian Processes

In my work, I have used mainly two methods for model-independent searches. More below

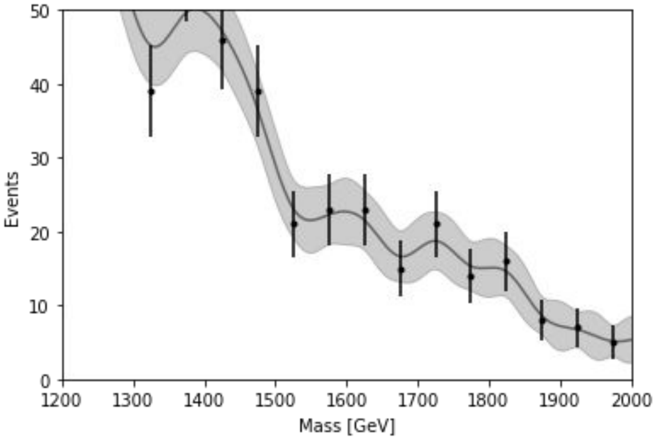
1. Semi-supervised anomaly detection using **Gaussian Mixture Models**:
   - Model background and background+signal prob. distribution of events
   - Use a linear combination of Gaussian pdfs + penalized maximum likelihood



Gaussian Mixture Models & Gaussian Processes

In my work, I have used mainly two methods for model-independent searches. More below

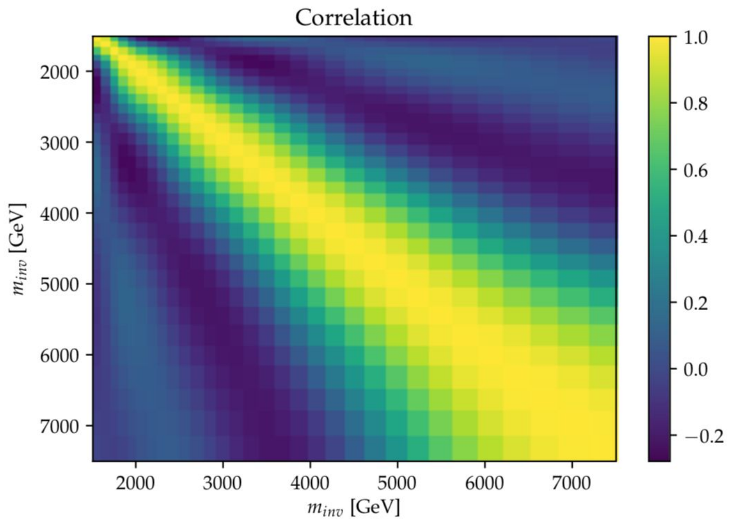
2. Model smooth backgrounds and generic signals in 1-D distributions with **Gaussian Processes**
   - Associate a multivariate Gaussian with as many dimensions as data points bins in a histogram
   - Use Maximum Likelihood to estimate the correlations between points



Kernels used (after fit example)

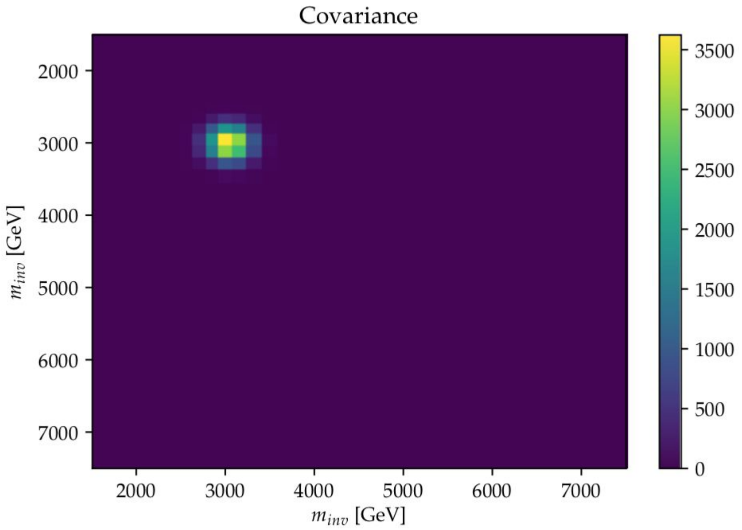
k_1

$$A\ \exp\left(\frac{d-(x+x')}{2a}\right)\sqrt{\frac{2l(x)l(x')}{l(x)^2+l(x')^2}}\exp\left(\frac{-(x-x')^2}{l(x)^2+l(x')^2}\right).$$

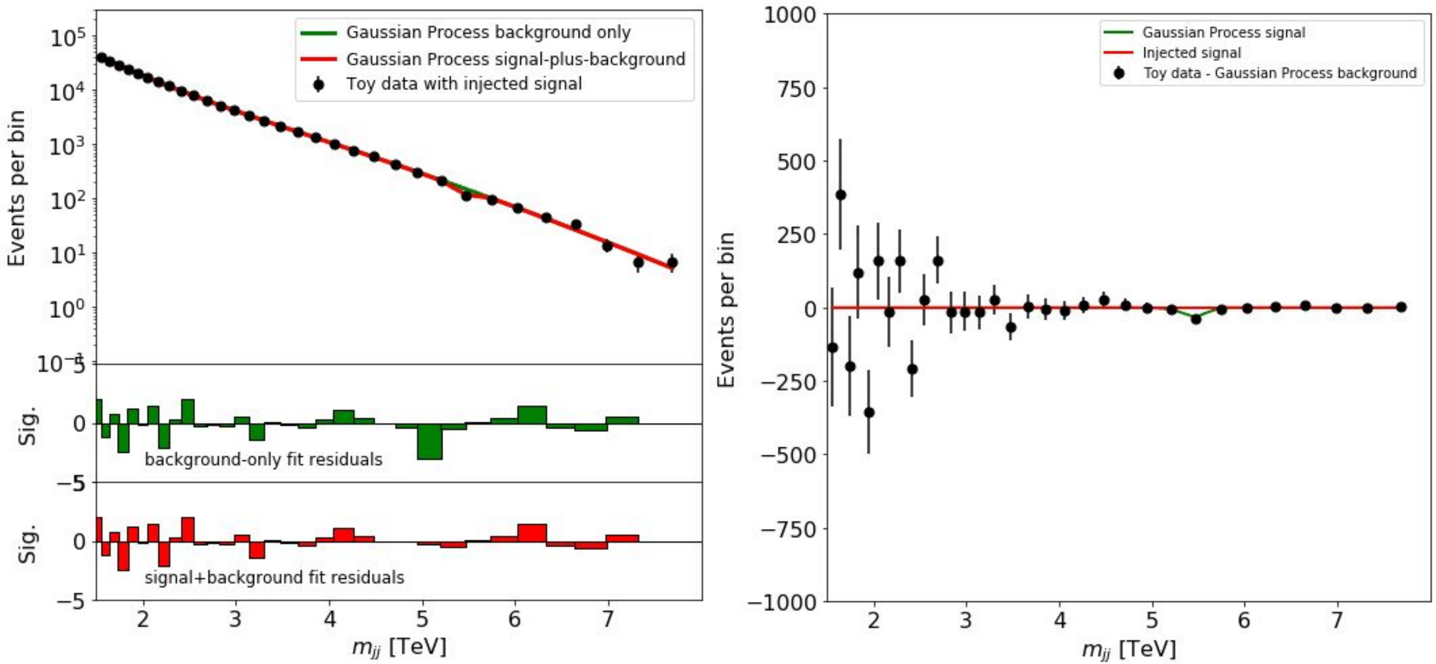


k_2

$$A_S\ \exp\left(-\frac{1}{2}(x-x')^2/l^2\right)\exp\left(-\frac{1}{2}\left((x-m)^2+(x'-m)^2\right)/t^2\right).$$



2-jet mass spectrum with no signal injected



Dijet simulated sample

Produced a fast simulation of signal and background events:

MadGraph5	→	Pythia8	→	Delphes3
Hard process generation		Showering and hadronization		Fast detector simulation

→ Use an object/event selection inspired in the one from ATLAS dijet analysis

11 variables extracted from the simulation describe the physics in the event:

Event wide

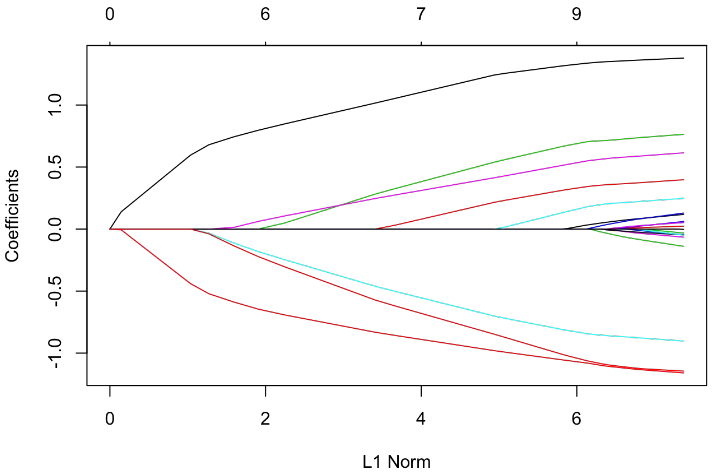
Dijet system

Object (jet) information

Penalized model-based clustering

Penalized likelihood approach → Add a term to the likelihood:

$$l(\theta) = \sum_{i=1}^{N} \left(\sum_{j=1}^{J} \pi_j \phi(x_i | \mu_j, \Sigma_j) \right) - \gamma p(\Theta)$$



p(.) is a penalty function, γ a regularization coefficient

→ The penalty term contains (combinations of) parameters Θ that are constrained in the model

Used for variable selection, i.e. **remove uninformative variables**; some choices are penalties on:

- The means of the gaussians, e.g. L2 norm: $p(\boldsymbol{\mu}) = \sum_{k=1}^{p} \sqrt{\sum_{j=1}^{J} \mu_{jk}^2}$ for p variables
- The values on the covariances

Sample preprocessing

Mixtures of Gaussians: Flexible model, but skewed data requires many Gaussian components

→ **Use the Tukey ladder of powers transformation, makes distribution more Gaussian**

$$f(x) = \begin{cases} x^{\alpha}, & \text{for } \alpha > 0 \\ -x^{\alpha}, & \text{for } \alpha < 0 \\ \ln(x), & \text{for } \alpha = 0 \end{cases}$$

- x is the variable to be transformed
- α is a parameter to be found (maximizes Shapiro-Wilk test statistic)
- Data is then standardized with respect to the background distribution

Variable	E_1	η_1	ϕ_1	p_{T1}
α	-0.65	0.6	0.775	-2.1

Variable	E_2	η_2	ϕ_2	p_{T2}
α	-0.6	0.55	0.825	-0.475

Variable	$\Delta R(j_1, j_2)$	$M_{\text{inv}}(j_1, j_2)$	E_T^{miss}	Sphericity	Centrality
α	-0.05	-1.05	0.125	0.25	0.5