Gaussian Process-based calculation of look-elsewhere trials factors

arXiv:2206.12328

Spåtind 2023 - Nordic Conference on Particle Physics

Victor Ananyev, Alex Read

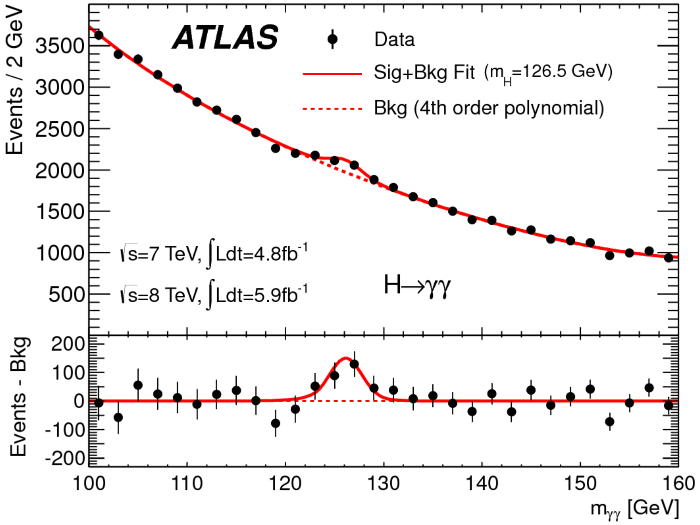
University of Oslo

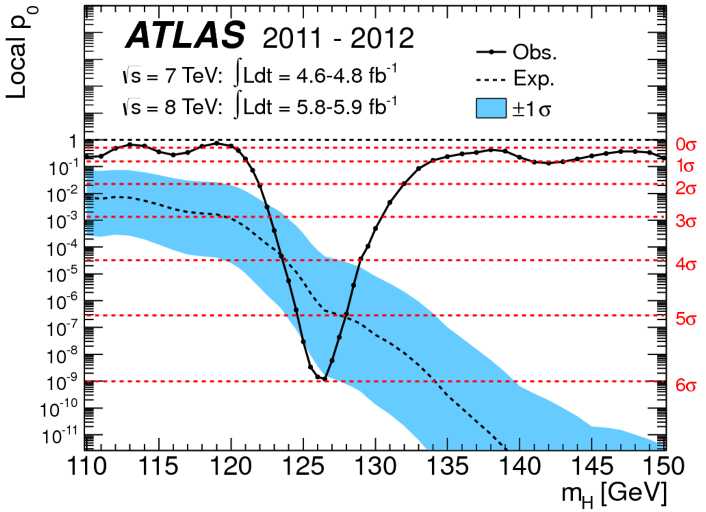
Introduction

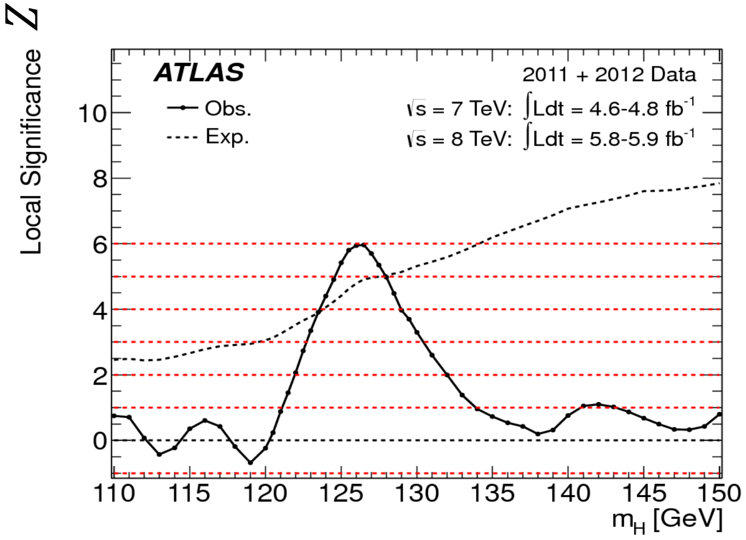
Recurring questions about how to efficiently and precisely (enough) calculate the global significance Z^{global} of a potential new resonance when a scan over a mass range yields a local significance Z^{local} at a particular mass value

1. Typical resonance search (bump hunt)
2. Look-elsewhere effect (trials factor)
3. Gaussian Processes
4. Leveraging Gaussian Processes for LEE/TF
    a) Given a covariance, estimate trials factor
    b) Estimating covariance in presence of background

Typical bump hunt



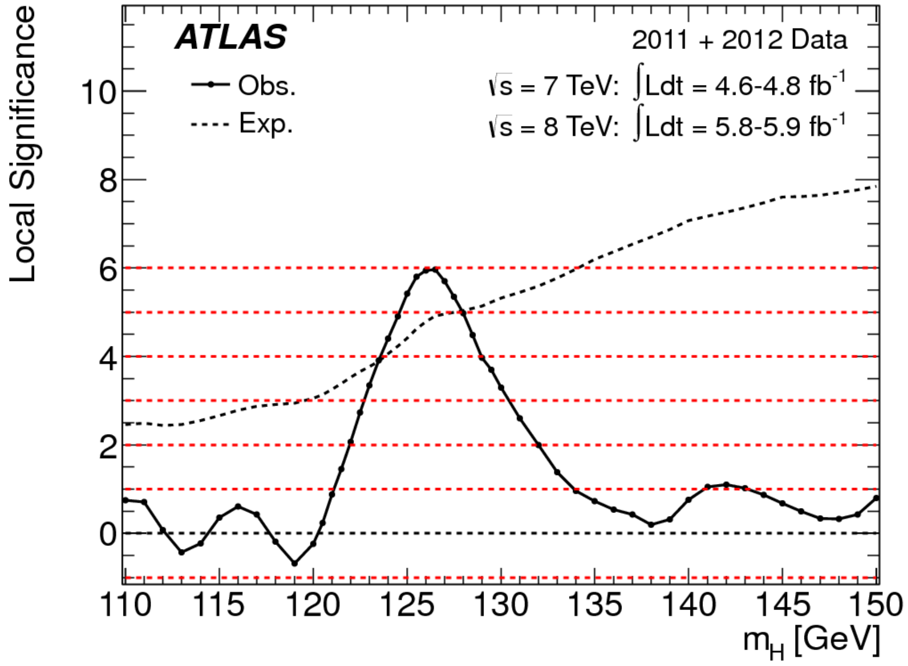




- Fit signal and background models to invariant mass spectrum
- Vary the mass parameter m of a hypothetical signal across the spectrum
- Calculate p_0, the probability to observe an excess *of the background* at least as big as the one we see in the data
  - Hope to see a narrow region with $p_0 \ll 1$
  - Report «local» significance ($Z \equiv N\sigma$)
  - Take into account trials factor (LEE)

Look-elsewhere effect (LEE) and Trials factor (TF)

- We focussed on the peak with maximum significance
- A peak from a background fluctation could arise anywhere in the spectrum: we have to «look elsewhere»
- The p_0 must be increased by a «trials factor»: This will give us a «global» p_0 and corresponding global Z with $Z^{global} < Z^{local}$



$$\boxed{\text{Trials factor } f_T \equiv p_o^{global} / p_0^{local}}$$

3 LEE options

1. **Outdated** rule-of-thumb
$$f_T \simeq \Delta m/\sigma_m$$
2. Estimate by **brute force** – fit a number of background-only pseudo-datasets or «toys»
3. Use a **cutting-edge** asymptotic approximation («G&V»)

THE EUROPEAN PHYSICAL JOURNAL C - PARTICLES AND FIELDS
Volume 70, Numbers 1-2, 525-530, DOI: 10.1140/epjc/s10052-010-1470-8 Open Access

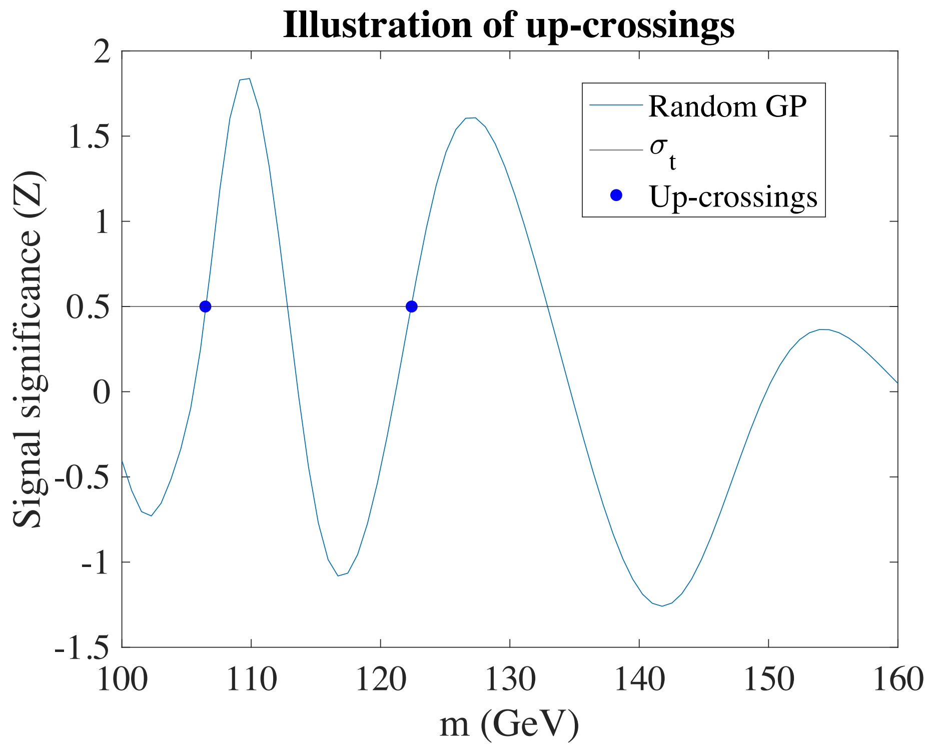
SPECIAL ARTICLE - TOOLS FOR EXPERIMENT AND THEORY
Trial factors for the look elsewhere effect in high energy physics
Eilam Gross and Ofer Vitells

Gross & Vitells Trials Factors (LEE)

- An elegant approximation based on average «up-crossings» $\langle N_{up} \rangle$
- Count «up-crossings» at low significance σ_t and extrapolate to high significance Z_{local} , i.e.

$$p_0^{global} \leq \; p_0^{local} + \langle N_{up}(\sigma_t) \rangle e^{-\frac{Z_{local}^2 - \sigma_t^2}{2}}$$

- Can use relatively small number of MC experiments to estimate $\langle N_{up}(\sigma_t) \rangle$, *but this can still be challenging*
- In worst case can use the data for *rough* estimate of $\langle N_{up}(\sigma_t) \rangle$

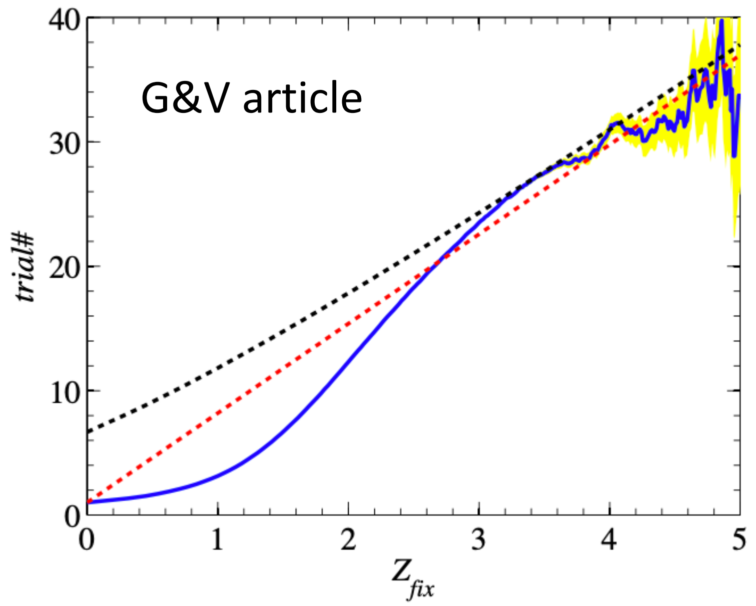


Gross & Vitells Trials Factors (LEE)

- What to do if not comfortable with large statistical uncertainty on TF (Z^{global}), besides throwing a lot of carbon at the problem?
- What to do about obvious over-conservatism of G&V UL below ~3σ?

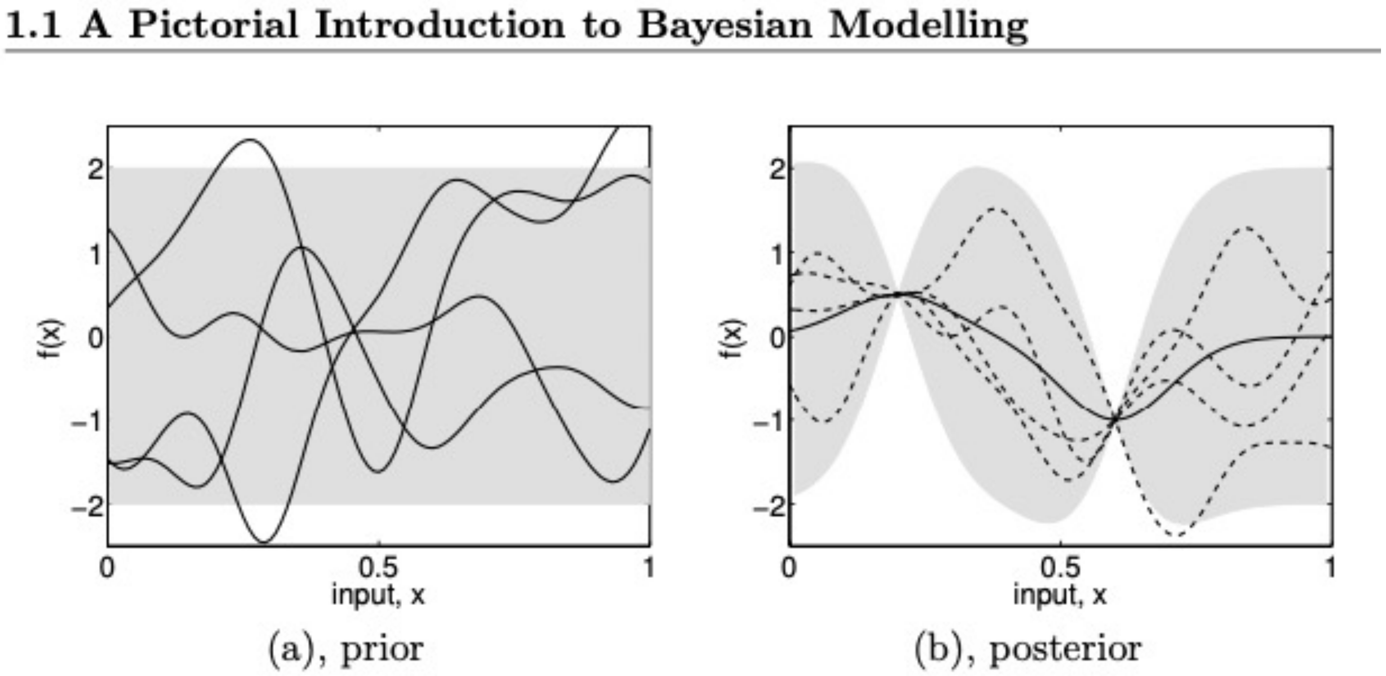
- **Can we do better?!**

- G&V article builds on extensive previous work on random χ^2 and Gaussian fields.
- I noticed the connection Rasmussen&Williams make in «Gaussian Processes for Machine Learning» (2006) between correlation length of Gaussian process and up-crossings.



MC simulation
MC uncertainty
G&V upper limit
High-Z approximation of χ_1^2

Gaussian process

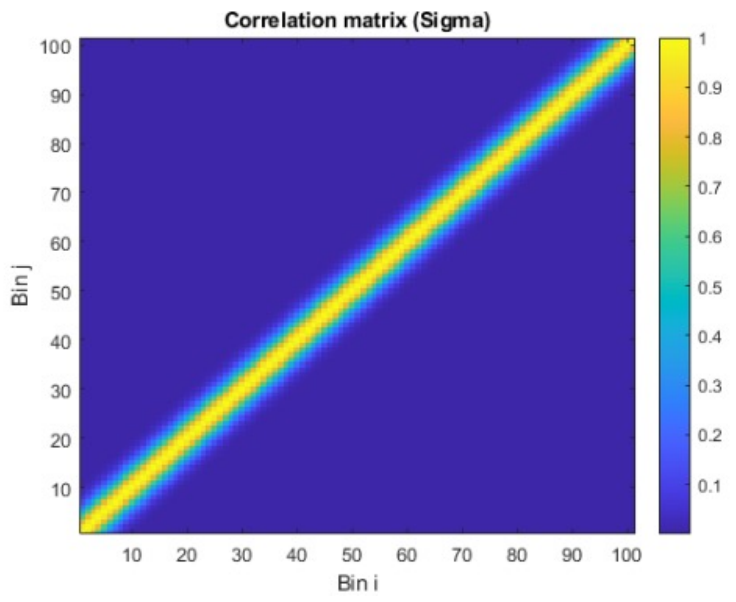


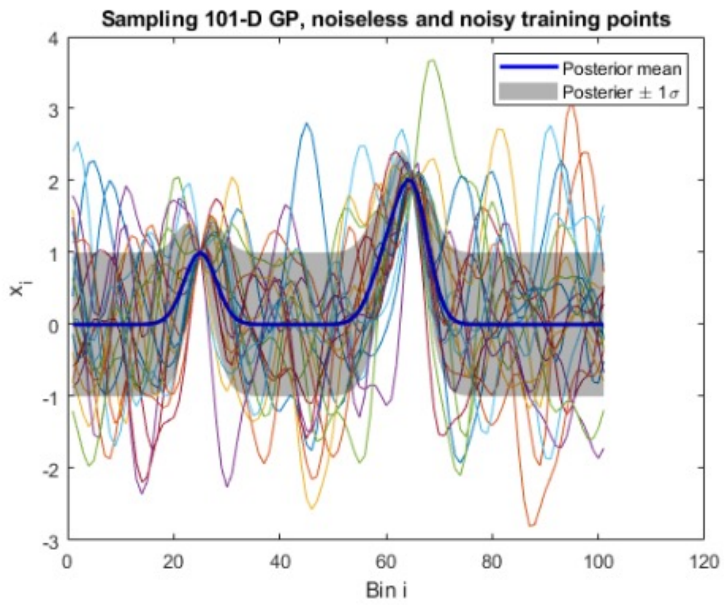
Rasmussen and Williams

- A GP is defined by a mean function $\mu(x)$ and a covariance function $\Sigma(x, x')$.
- In a typical (Bayesian) analysis one starts with priors and "trains" the process with data to obtain a posterior prediction $\mu^*(x), \Sigma^*(x, x')$.

Gaussian process

- Similar example
$$\mu(x) = 0$$
$$\Sigma(x, x') = A^2 e^{-(x-x')^2/(2l^2)},$$
with $A = l = 1$ and train on 1 noiseless and 1 noisy observation.
- Random samples of the posterior GP look a lot like signed local significance (Z) scans! Same is true for prior.



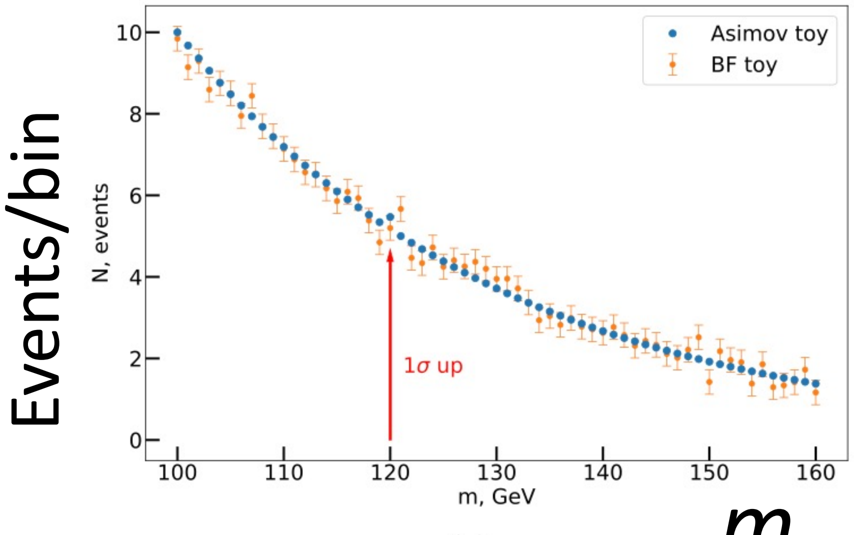


Gausssian process for LEE!

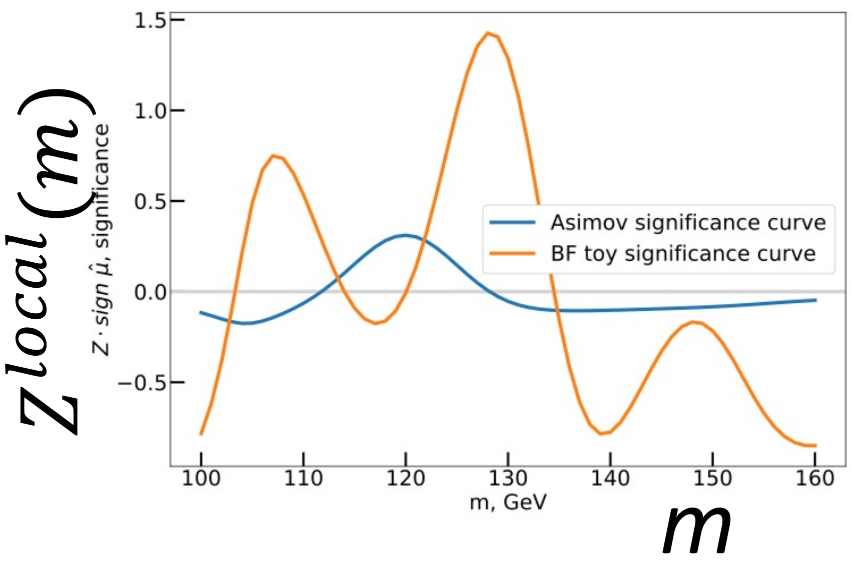
- Instead of defining priors and training the GP, we do *something completely different:* Construct the covariance directly and as efficiently as possible from information encoded in the fitting procedure used in the resonance search.
  - $\mu(x) \equiv \langle Z(m) \rangle = 0$ asymptotically under the background hypothesis
- Once we have $\Sigma(m, m')$, it is computationally cheap to sample the GP ("generate GP-toys") and measure with high precision:
  - up-crossings
  - directly the global p_0, especially at low and moderate signficance, but also at moderately high Z_{local}
- It was already worked out for us how to calculate the average up-crossings at <u>any significance level</u> directly from the covariance in *"Random Vibrations. Analysis of Structural and Mechanical Systems", Lutes, L.D. and Sarkani, S., Butterworth Heinemann, Boston (2004).*

A new (!) set of Asimov (background) datasets

- Sum the covariances $\Sigma_i(m, m')$ resulting from *individual background (bin i) fluctuations*.
- The standard background-only toy is a series of independent fluctuations!
  - Covariances for independent sources add linearly
  - There are cases with correlated data - not covered here
- Fitted signal amplitude $\hat{\mu}(m)$ must be proportional to the amount of «signal». A single fluctuation is a placeholder for all possible fluctuations of the bin in question, just like the fit to a standard Asimov dataset is a placeholder for an ensemble of fits to random toy datasets.
- Last step is a normalization (inspired by Gibbs for his GP kernel function) to render $\Sigma(m, m) = 1$.

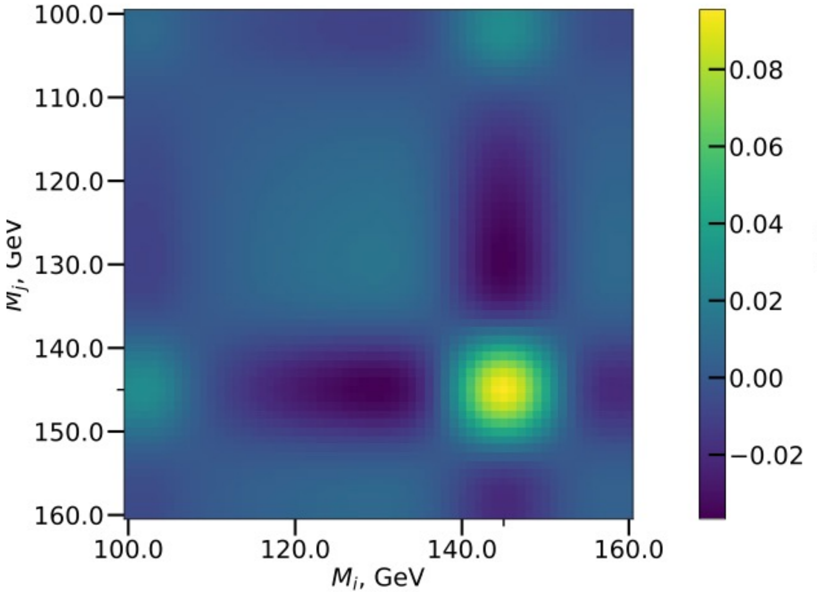


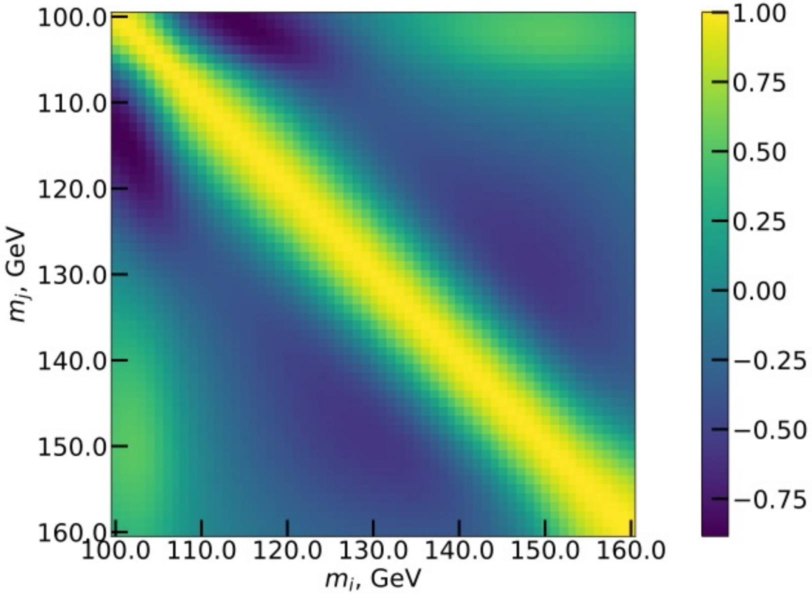
(a)



Covariances $\Sigma(m, m')$

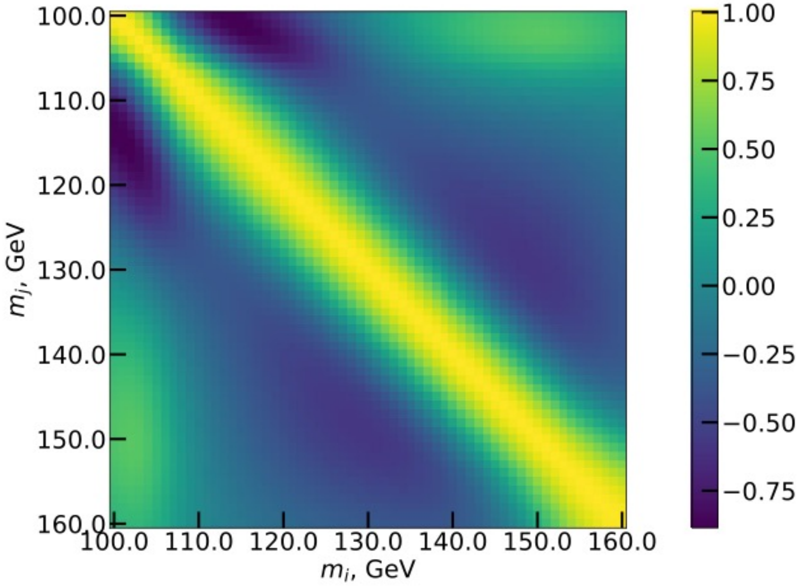
Contr. to covariance for fluctuation at 145 GeV

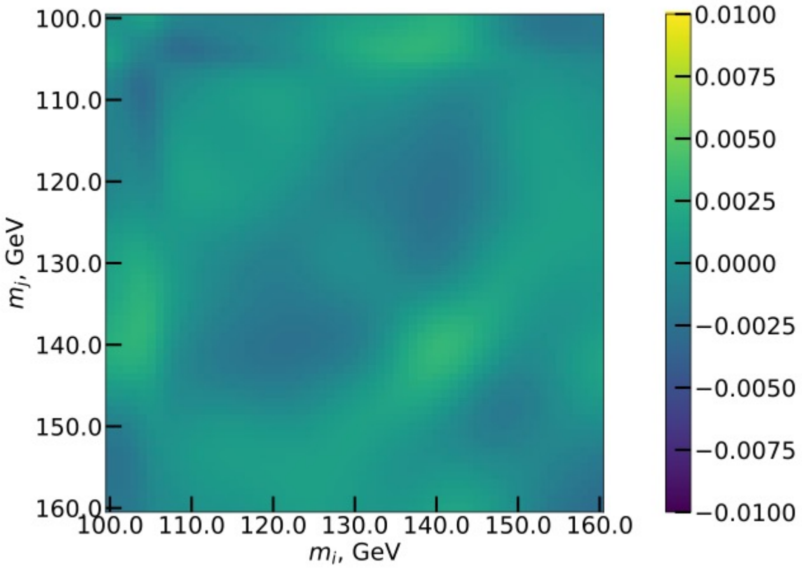




Covariance from new Asimovs

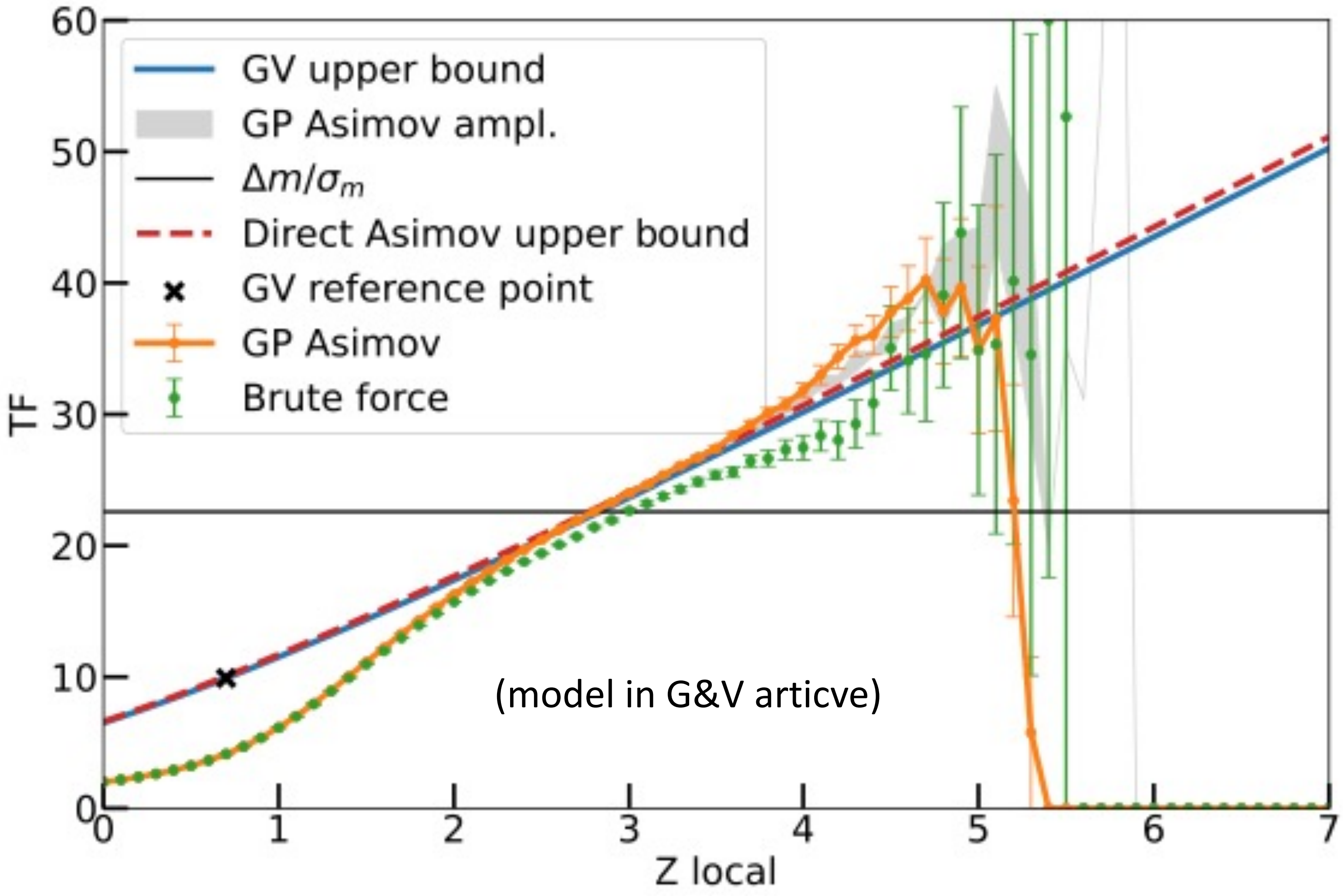
Covariance from Brute-force toys





Absolute difference New Asimov – Brute force toys

Various paths to the trials factor (and Z_{global})



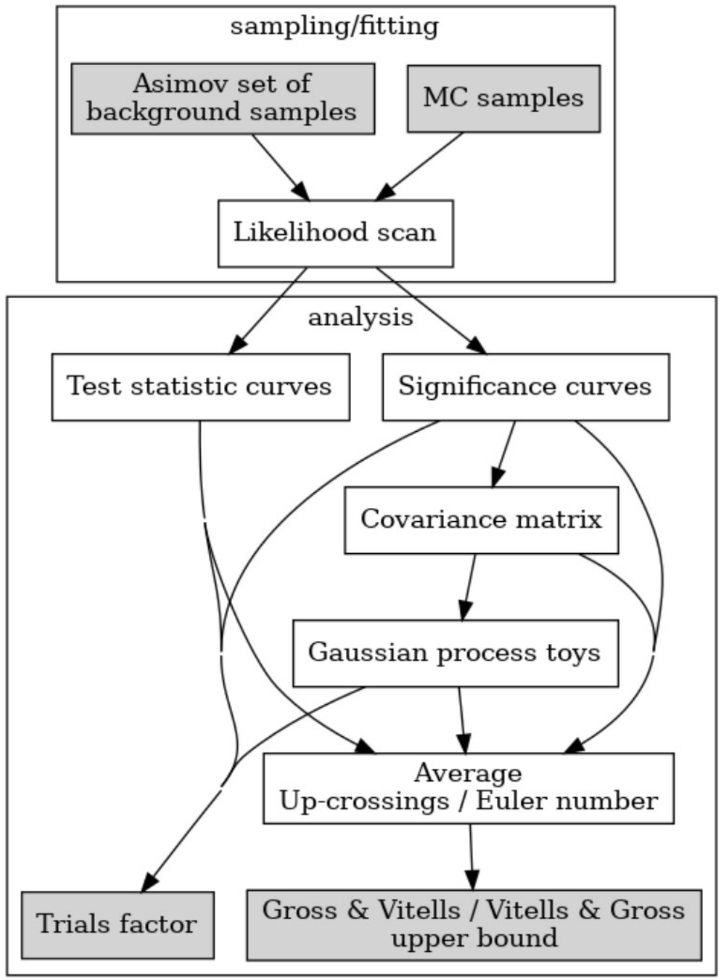
(model in G&V articve)

Various studies and cross-checks

- 3 different statistical models:
  1) G&V model based on background template,
  2) parametric background inspired by $H \to \gamma\gamma$, and
  3) inspired by $H \to \gamma\gamma$ with additional scan over width of mass peak
- We (more precisely ☺) reproduce the published G&V example results (1)
- 2 different software implementations: python (published results) and MATLAB (cross-check)
- Effect of poor choice of bin width (affects sensitivity, but no bias)
- 10x larger data sample for G&V model (smaller diff. BF-GP toys)

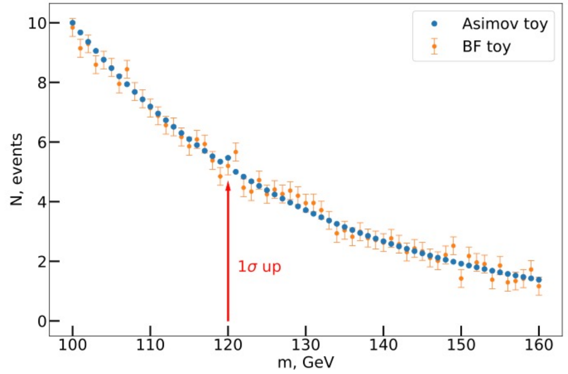
SigCorr: Python software on gitlab.cern.ch

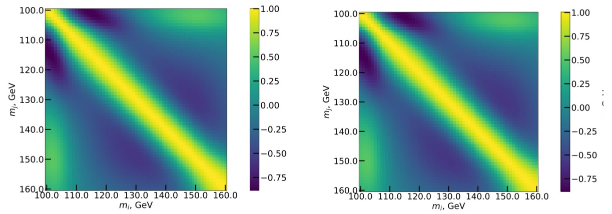
- A framework to study the trials factor
- Several ways to estimate the trials factor (see figure)
- Utilities that consistently operate on defined data structures that allow the user to build their own pipeline
- See the docs (sigcorr.docs.cern.ch/dev/) for installation details, usage examples and tutorials.

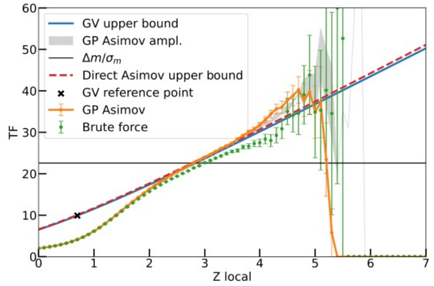


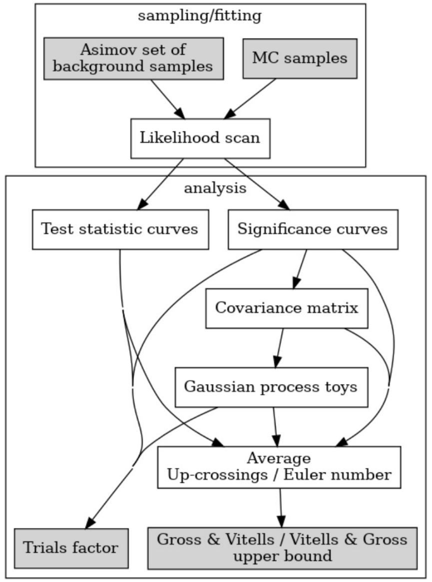
Conclusions

- We propose to model the covariance of the significance for bump hunts with fit-scans of special datasets consisting of Asimov background plus 1-bin-at-a-time fluctuations.
- With 3 quite different toy models of searches we have excellent agreement between large samples of brute-force toys and GP-toys (and reproduce the G&V published result on their toy model).
- The G&V approximation *for high local significances* based on $\langle N_{up} \rangle$ at low significance is still important, but now we have a GP-based method that gives precise and accurate results for moderate to low significances as well as precise and accurate estimates of the up-crossings at low significance.
- We will soon submit a publication showing how to obtain an estimate of the covariance with only a single Asmimov fit (and some derivative calculations) with the same accuracy as our previous, already carbon-friendly, procedure.



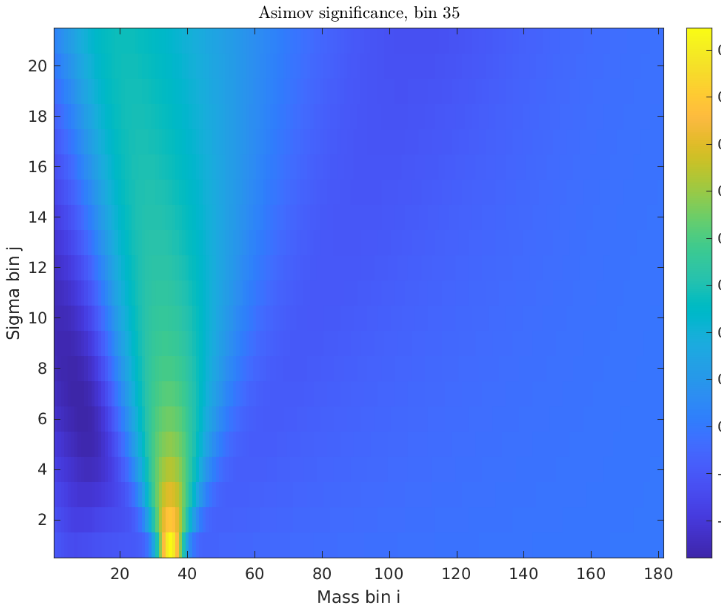




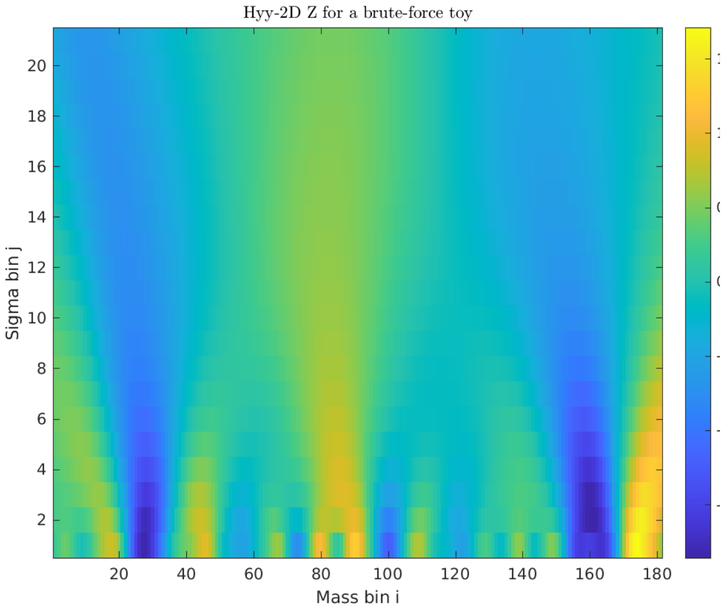


Backup

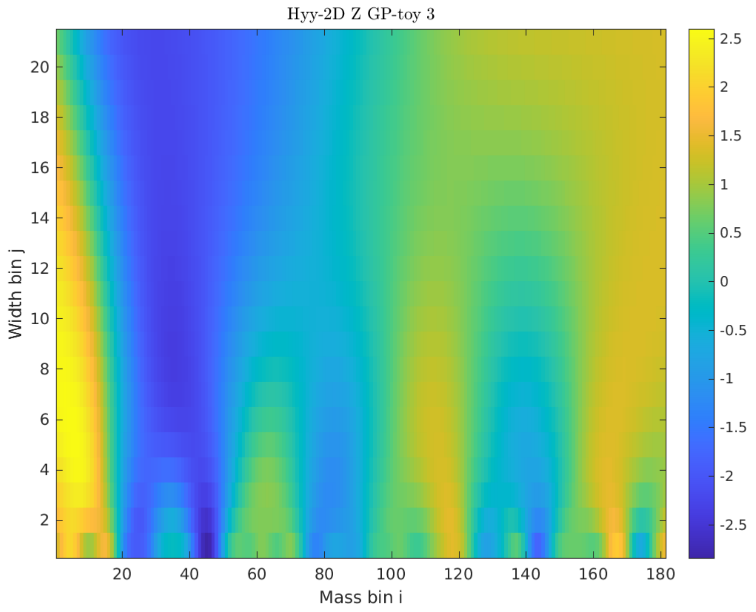
Significance scans of 2-D toys



Special Asimov toy

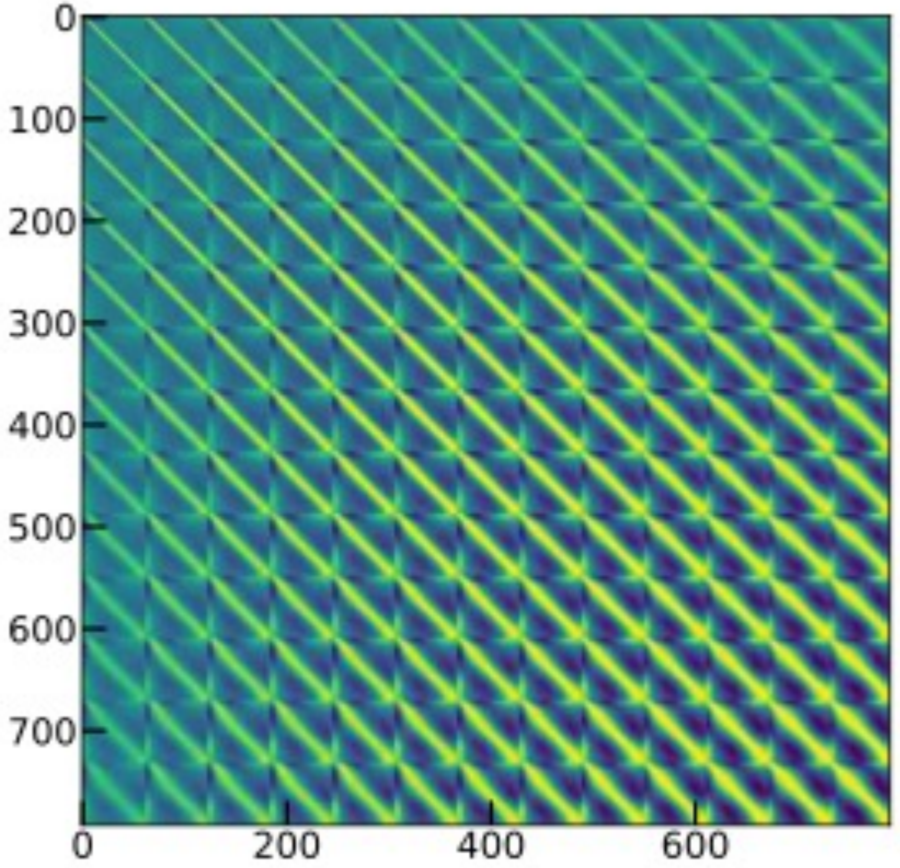


Brute-force toy



GP toy

Covariance for 2-D scan of mass and width



2 x 2-D (mass and width) → 4-D covariance matrix
Unwrapped into 2-D

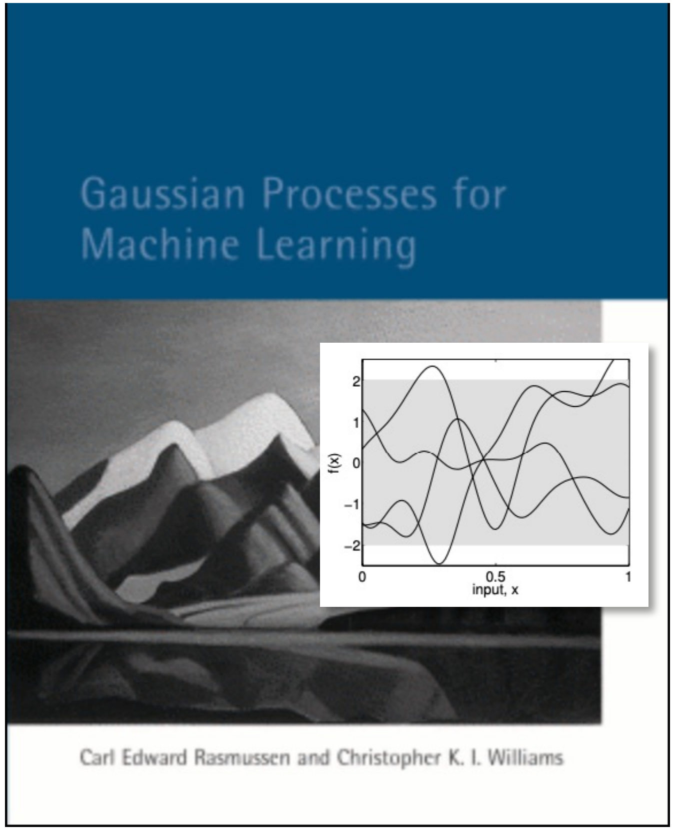
$$\{M_1, M_2, \ldots\}$$
$$\Downarrow$$
$$\{M_{11}, \ldots, M_{1k}, M_{21}, \ldots, M_{2k}, M_{31}, \ldots\}.$$

Gross & Vitells Trials Factors (LEE)

- Using high-*Z* approximation of χ_1^2 , can show that: $f_T \simeq 1 + \sqrt{\frac{\pi}{2}} \mathcal{N} Z_{local}$, where $\mathcal{N}$ is «effective # of independent search regions»
    - Outdated rule-of-thumb is ~OK for $Z_{local} \simeq 2 - 3$, but *wrong* for large *Z*!
    - Bob Cousins called this «an important discovery»!

Prediction of up-crossings?

- G&V: «The function C(θ) [that $\langle N_{up}(\sigma_t) \rangle$ depends on] can in general be difficult to calculate.»
- The G&V results are based on the properties of random Gaussian fields, i.e. Gaussian Processes



Analytic prediction of up-crossings?

- Adler (1981) Theorem 4.1.1 states that expectation value of number of up-crossings per unit interval at level u is

$$\mathbb{E}(N_u) = \frac{1}{2\pi}\sqrt{\frac{-k''(0)}{k(0)}}\, e^{-u^2/(2k(0))},$$

where $k(x, x') = k(x - x')$, i.e., a «stationary kernel» (i.e. covariance)

- For exp-squared kernel with correlation length l, $k(x) = \sigma^2 e^{-x^2/2l^2}$, we have

$$\mathbb{E}(N_u) = \frac{1}{2\pi l} e^{-u^2/(2\sigma^2)}$$

  - For significance field $u \equiv Z$ and $\sigma \equiv 1$
  - $\mathbb{E}(N_u) = \frac{1}{2\pi l} e^{-Z^2/2}$ (basis of G&V extrapolation from low to high *Z*)

Analytic prediction of up-crossings?

- For mass range Δm and constant Gaussian mass resolution $\sigma_m = l$ (*), the predicted average number of up-crossings at any threshold is

$$\mathbb{E}(n_u) = \frac{\Delta m}{2\pi\sigma_m} e^{-Z^2/2}$$

- (*) A comment in Frate, Cramner et al., Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes, arXiv:1709.05681v1 about an additional $\sqrt{2}$ is misleading for the significance field.

Analytic prediction of up-crossings for $\sigma_m(m)$?

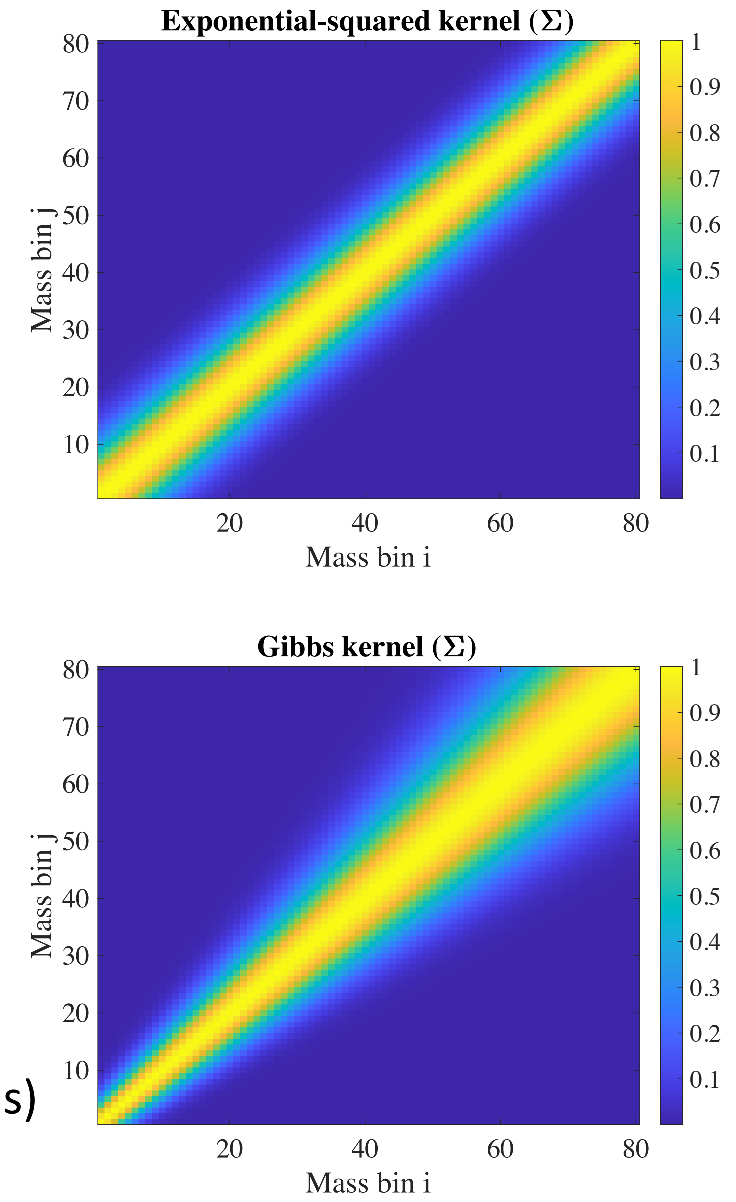
- Propose $\mathbb{E}(N_u) = \frac{\int_{m_0}^{m_1} \mathbb{E}(N_u(m))dm}{\int_{m_0}^{m_1} dm} = \frac{1}{2\pi} e^{-\frac{Z^2}{2}} \mathbb{E}(\frac{1}{l(m)})$
- For linear $\sigma(m)$, e.g. «Gibbs kernel» (next page), we have

$$l(m) = l_0 + (m - m_0)(l_1 - l_0)/(m_1 - m_2)$$

and find

$$\mathbb{E}(n_u) = \frac{1}{l_1 - l_0} \ln(\frac{l_1}{l_0}) \frac{1}{2\pi} e^{-\frac{Z^2}{2}}$$

Two basic Gaussian Process kernels



- Gaussian signal with constant mass resolution
    - The unit «exponential-squared» kernel:

$$\Sigma(m, m', \sigma_m) = e^{-(m-m')^2/2\sigma_m^2}$$

    - Example with $\sigma_\mathrm{m} = 5$ GeV (8.3 bins)
- Gaussian signal with mass-dependent resolution $\sigma_m(m)$:
    - The unit «Gibbs» kernel:

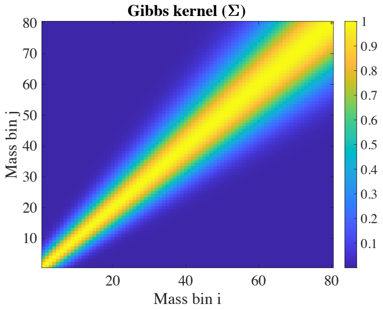
$$\Sigma(\mathrm{m}, \mathrm{m}') = \sqrt{\frac{2\sigma_m(m)\sigma_m(m')}{\sigma_m^2(m)+\sigma_m^2(m')}}\, e^{-(m-m')^2/(\sigma_m^2(m)+\sigma_m^2(m'))}.$$

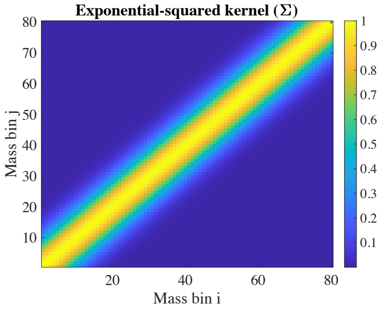
    - Example with linear $\sigma_m(m)$ between 2-10 GeV (2.7-13.3 bins)

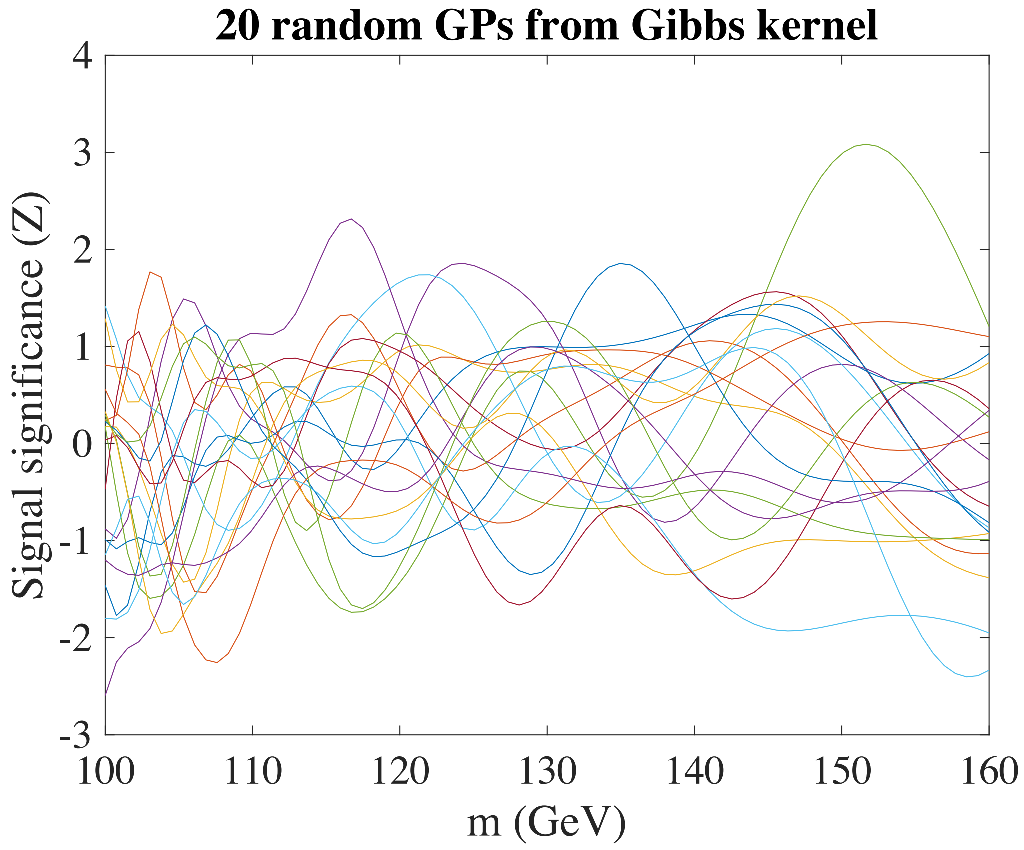
Gaussian Process toys prediction of up-crossings?

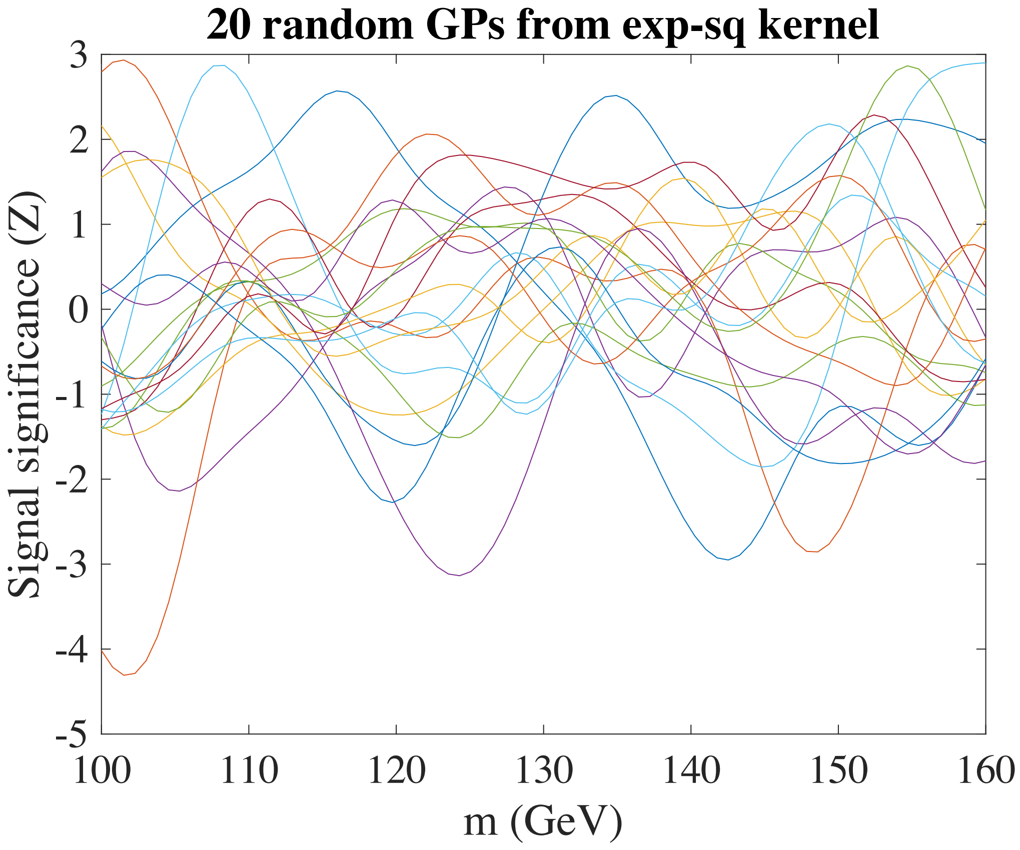
- A Gaussian Process for a set of points $\boldsymbol{x}$ is defined by a mean function $\mu(\boldsymbol{x})$ and a kernel $k(\boldsymbol{x}, \boldsymbol{x}')$, i.e. a covariance matrix $\Sigma(\boldsymbol{x}, \boldsymbol{x}')$.
- Fitting a signal model to an ensemble of background spectra at a particular mass hypothesis x_i should result in mean significance of 0 and standard deviation of 1.
- The shape of the signal model must strongly influence the covariance between 2 points $\Sigma(\boldsymbol{x}, \boldsymbol{x}')$, *however, is it a good approximation to neglect the influence of background*?
- With $\mu(\boldsymbol{x}) = 0$ and $\Sigma(\boldsymbol{x}, \boldsymbol{x}')$ in hand we can easily generate huge numbers of MC experiments for $Z(m)$ from the multivariate Gaussian distribution or «GP-toys»!
  - No need to perform zillions of time-consuming fits!
  - Reliable results for small Z!

Random samples of *background-only* Z-GP's



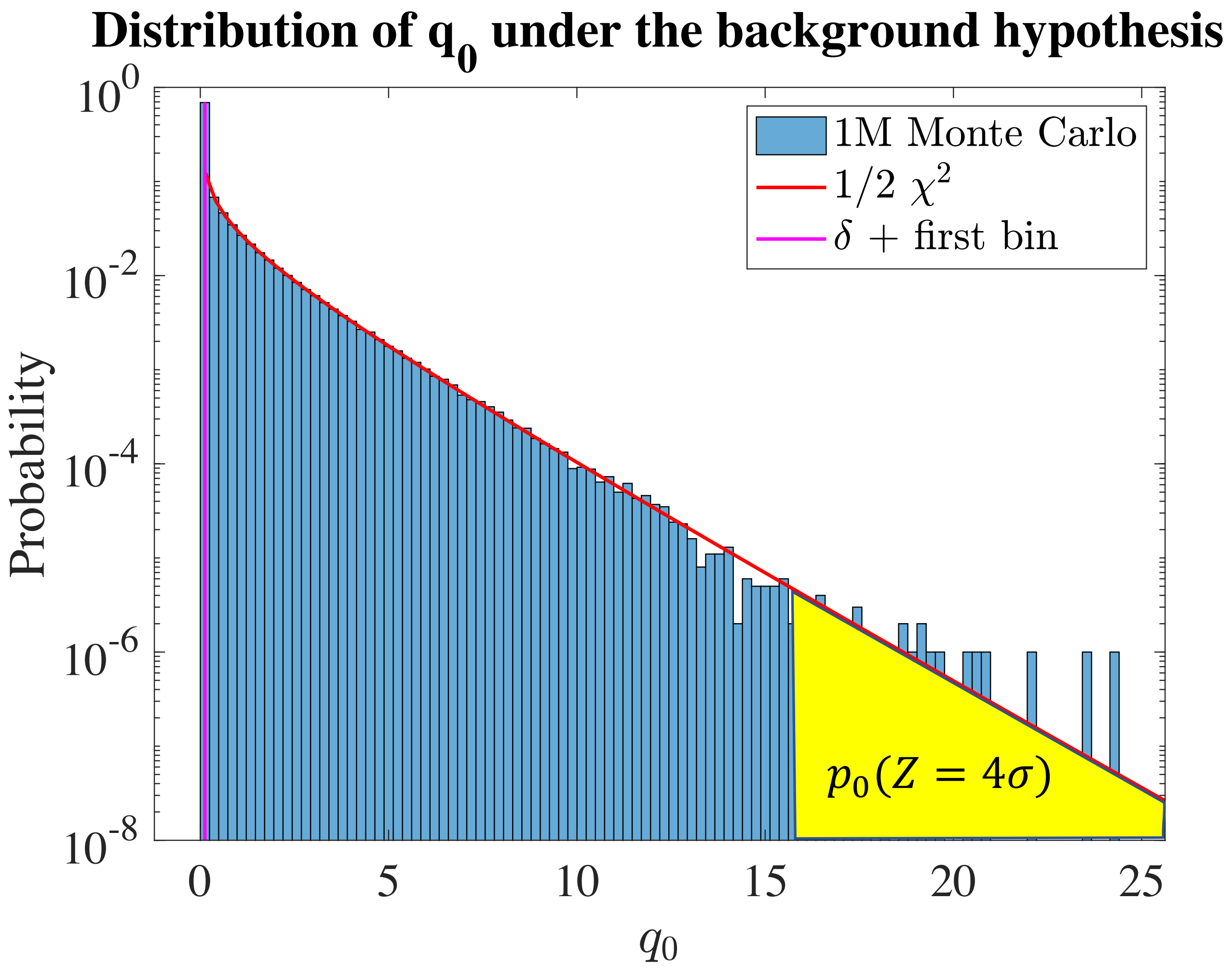






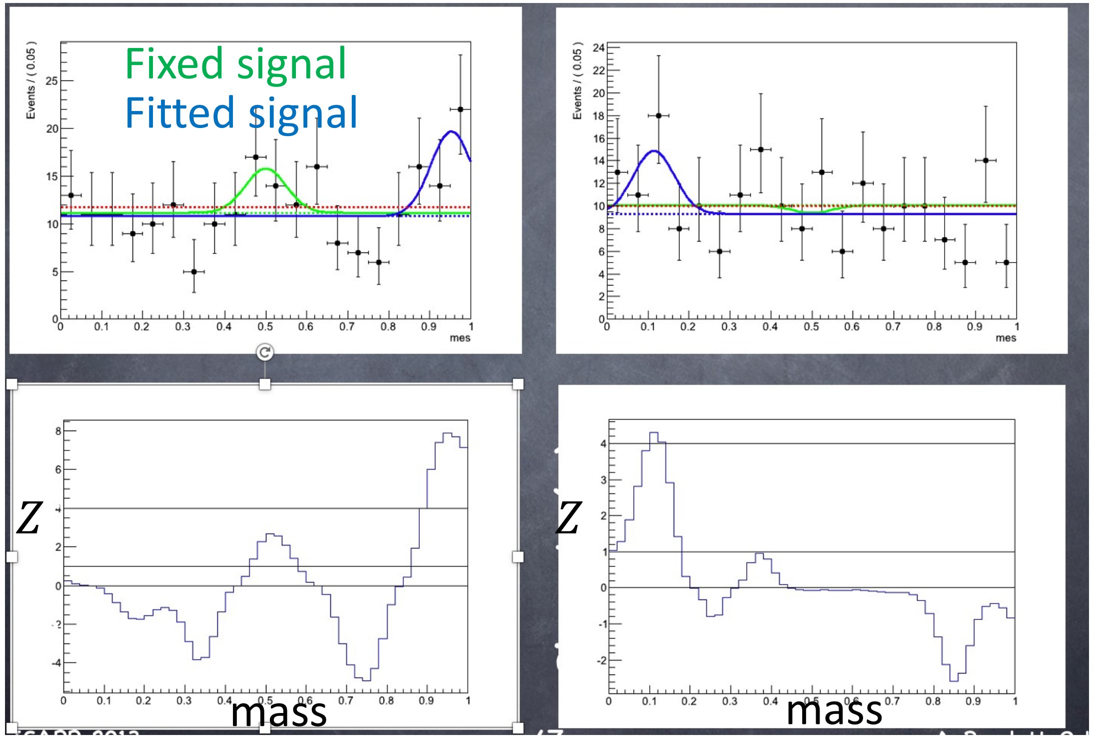
Typical bump hunt

- At the LHC we often use $q_0 \equiv -2\ln\frac{\mathcal{L}(0,m,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},m,\hat{\theta})}$
  - θ is a vector of nuisance parameters (background, systematics)
  - m is the mass parameter we are scanning over
  - μ is the amplitude or «strength» of the signal
- q_0 is distributed as χ_1^2 for background and large enough data samples (asymptotic regime)
- p_0 is the upper-tail probability $p_0 \equiv P(q_0 \geq q_0^{obs})$
- $\sqrt{q_0}$ corresponds to significance of excess, e.g.
  $$q_0 = 16 \Longrightarrow Z = 4, i.e.\ 4\sigma$$
- Use MC «toys» beyond asymptotic regime for $P(q_0 \geq q_0^{obs})$



3 LEE options

1. Outdated rule-of-thumb
$$f_T \simeq \Delta m / \sigma_m$$
2. Estimate by brute force – fit a number of background-only pseudo-datasets or «toys»
3. Use an asymptotic approximation («G&V»)




THE EUROPEAN PHYSICAL JOURNAL C - PARTICLES AND FIELDS
Volume 70, Numbers 1-2, 525-530, DOI: 10.1140/epjc/s10052-010-1470-8 Open Access

SPECIAL ARTICLE - TOOLS FOR EXPERIMENT AND THEORY

Trial factors for the look elsewhere effect in high energy physics

Eilam Gross and Ofer Vitells