Including theoretical uncertainties in Higgs fits

Sylvain Fichet
International Institute of Physics, Natal

16/09/13

work with Grégory Moreau, Glen Cowan, pre-print to appear soon

Overview

Introduction

A Higgs-like particle at the LHC...

- But one may detect the existence of New Physics by looking accurately enough at the Higgs properties.

- Low-energy effects of any heavy new physics are captured by higher-dimensional operators,

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{\alpha_i}{\Lambda^2}\mathcal{O}_i + \text{higher order terms}$$

- Higgs precision physics consists in getting information on α_i, Λ

  [Eboli et al '12, B. Dumont/SF/G. von Gersdorff '13, SF '13]

Accurately enough ?

- Anomalous Higgs properties induced by New Physics may be tiny. A good understanding and treatment of all sources of uncertainty is necessary.

- Statistical uncertainties will decrease as $1/\sqrt{\mathcal{L}}$. Experimental systematics should naturally decrease with a refined understanding of detectors.

- So what about theoretical uncertainties ? They are already sizeable and could be dominant in the future.

Accurately enough ?

- Anomalous Higgs properties induced by New Physics may be tiny. A good understanding and treatment of all sources of uncertainty is necessary.
- Statistical uncertainties will decrease as $1/\sqrt{\mathcal{L}}$. Experimental systematics should naturally decrease with a refined understanding of detectors.
- So what about theoretical uncertainties ? They are already sizeable and could be dominant in the future.

⇨ Our main goal is to provide a detailed survey of the various treatments of theoretical uncertainties for Higgs fits.

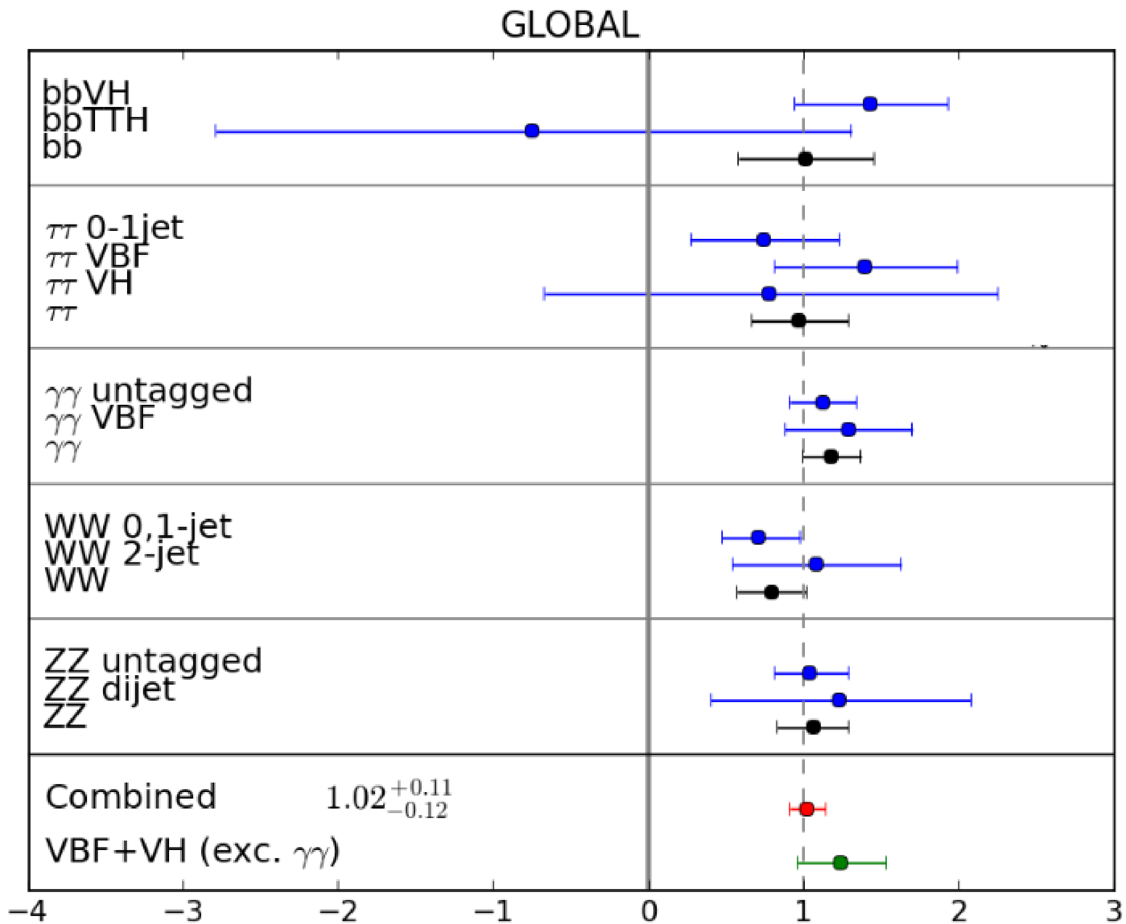
(However we use the latest data, so the fits themselves are also relevant)

- Accessible observables are signal strengths,

$$\mu_{XY} = \frac{N_{evts}}{\sigma(X \to h)\mathcal{B}(h \to Y)\epsilon_{XY} \times \mathcal{L}}$$

- Latest data:

~ 60 subchannels



Probing the Higgs couplings to the rest of the SM

- For a given mass, all Higgs decays rates are predicted in the SM. Any deviation would be a consequence of New Physics. The higher-dimensional operators contribute to Higgs anomalous couplings.

- The tree-level Higgs anomalous Lagrangian looks like

$$\mathcal{L}_h^{tree} = c_W \frac{h}{v} 2m_W^2 W_\mu^+ W_\mu^- + c_Z \frac{h}{v} m_Z^2 (Z_\mu)^2 - \sum_i c_{fi} \frac{h}{v} m_f \bar{\Psi}_L \Psi_R - h.c.$$
$$+\zeta_\gamma h (F_{\mu\nu})^2 + \zeta_g h (G_{\mu\nu})^2 + \zeta_Z h (Z_{\mu\nu})^2 + +\zeta_W h W_{\mu\nu}^+ W_{\mu\nu}^- + \zeta_{Z\gamma} h Z_{\mu\nu} F_{\mu\nu}$$

- The SM corresponds to $c_i = 1, \zeta_i = 0$.

- However in this work we will consider a two-parameter case : (c_V, c_f)

$$\mathcal{L}_h^{tree} = c_V \frac{h}{v} 2m_W^2 W_\mu^+ W_\mu^- + c_V \frac{h}{v} m_Z^2 (Z_\mu)^2 - c_f \sum_i \frac{h}{v} m_f \bar{\Psi}_L \Psi_R - h.c.$$

Comparing data and predictions

- We have the experimental signal strength

$$\mu_{XY}^{exp} = \frac{N_{evts}}{[\mathcal{L} \times \sigma(X \to h)\mathcal{B}(h \to Y)\epsilon_{XY}]_{SM}}$$

and the signal strength expected in case of New Physics,

$$\mu_{XY}^{th} = \frac{[\sigma(X \to h)\mathcal{B}(h \to Y)\epsilon_{XY}]_{SM+HDOs}}{[\sigma(X \to h)\mathcal{B}(h \to Y)\epsilon_{XY}]_{SM}}$$

with $\mu_{XY}^{th}(c_V, c_f)$

- One has to model the likelihood function, $p(\mu^{exp}|\mu^{th}) \equiv L(c_V, c_f)$

- It should capture all sources of uncertainty as best as possible

Standard procedure to take into account an uncertainty

- One parameterizes it by nuisance parameters entering the likelihood. Then one eliminates them following some principle.

- In Bayesian statistics, nuisance parameters are associated with a prior PDF and integrated over. ⇨ "marginalization"

$$\hat{L}(c_i) = \int_{\mathcal{D}_\delta} d\delta \, L(c_i, \delta) p(\delta)$$

- In Frequentist statistics, they are determined as to maximize the likelihood.
  ⇨ "profiling"
  (In the hybrid frequentist method adopted for the LHC, they are also associated with a PDF.)

$$\hat{L}(c_i) = \max_{\delta \in \mathcal{D}_\delta} [L(c_i, \delta) p(\delta)]$$

Modeling the leading theoretical uncertainties for Higgs fits:

- These are the QCD uncertainties on the Higgs production cross-sections, from parton PDFs and α_s^2 / renomalization scale.

- They reach roughly 10 % [LHCHWG]. They 100 % correlate all subchannels.

- They cancel in μ_{XY}^{th} but not in μ_{XY}^{exp} . So we introduce

$$\mu_i^{exp} \to \mu_i^{exp}(1 + \delta \times \Delta_i) \qquad \delta \in [-1, 1]$$

  where Δ_i sets the magnitude for a given subchannel.

- We identify two independent δ's. One is associated with the production modes ggF, tth, and the other one with VBF,Vh.

- As the uncertainty comes from QCD, we adopt a flat prior for δ , $p(\delta) = cste$

The likelihood often employed:

- The analysis released so far by the experiments allow us to model the likelihood enclosing the statistical and systematic uncertainties as a product of gaussians associated with each of the subchannels,

$$L = \prod_i L_i \qquad L_i = \mathcal{N}(\mu_i^{th} - \mu_i^{exp}, \sigma_i) \qquad \mathcal{N}(\mu_0, \sigma) = \exp[-\frac{(\mu - \mu_0)^2}{2\sigma^2}]$$

- All subchannels are decorrelated. All uncertainties are enclosed in the σ_i

- Apparently, in many cases, σ_i is taken to be

$$\sigma_i = \sqrt{(\sigma_i^{exp})^2 + (\sigma_i^{th})^2}$$

⇨ What does it mean ?

- In both frameworks, this treatment in "quadrature" corresponds exactly to assuming Gaussian uncorrelated theoretical uncertainties.

- In Bayesian, for a given subchannel one finds

$p(\delta)$

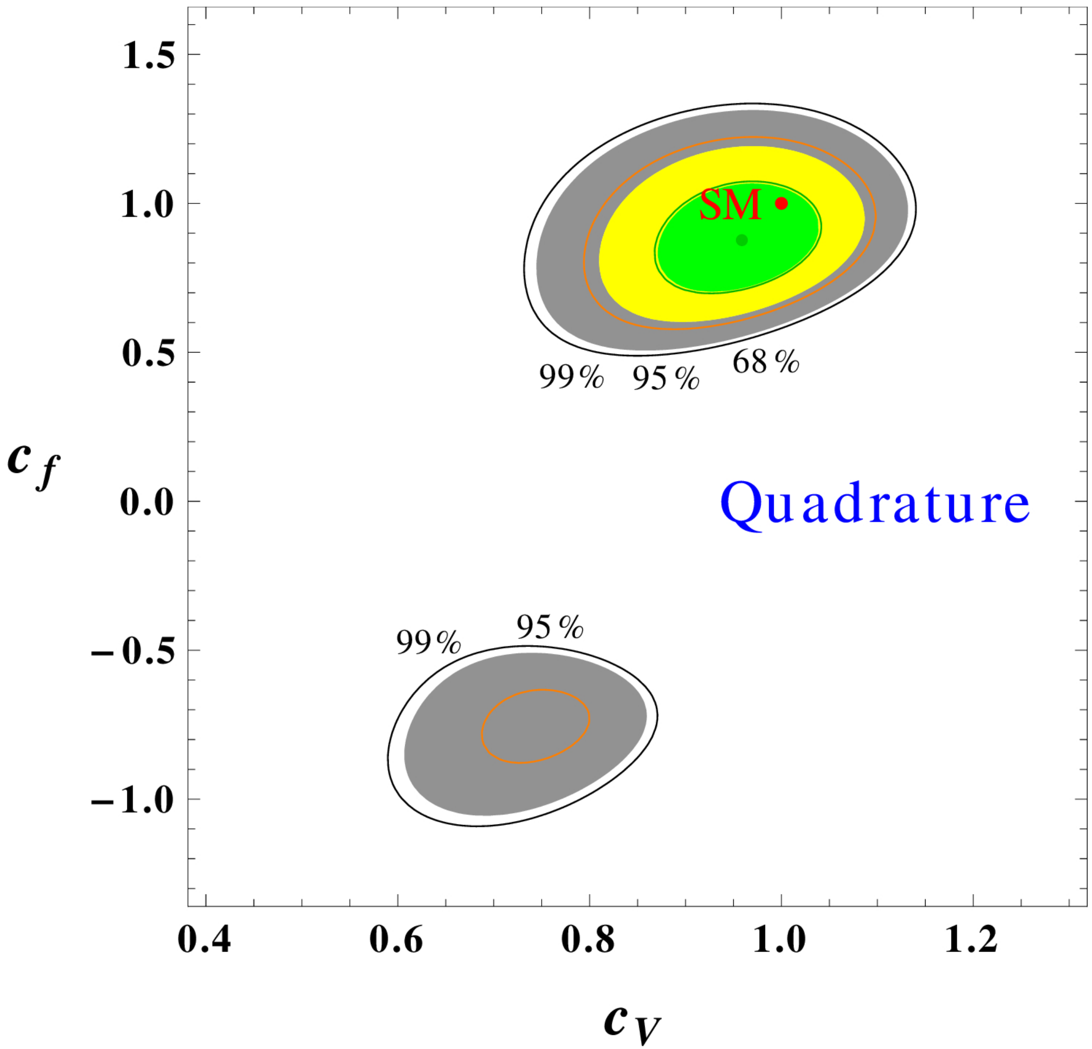
$$\int_{\mathcal{D}_\delta} d\delta\, \mathcal{N}(\mu^{th} - \mu^{exp}(1+\delta), \sigma^{exp})\, \mathcal{N}(\delta, \sigma^{th}/\mu^{exp})$$
$$= \mathcal{N}(\mu^{th} - \mu^{exp}, \sqrt{(\sigma^{exp})^2 + (\sigma^{th})^2})$$

In Frequentist one finds:

$$\max_{\delta \in \mathcal{D}_\delta} [\mathcal{N}(\mu^{th} - \mu^{exp}(1+\delta), \sigma^{exp})\, \mathcal{N}(\delta, \sigma^{th}/\mu^{exp})]$$
$$= \mathcal{N}(\mu^{th} - \mu^{exp}, \sqrt{(\sigma^{exp})^2 + (\sigma^{th})^2})$$

One is missing the crucial feature of 100 % correlation between subchannels.

And the Gaussian shape for $p(\delta)$ is also quite discutable.



Colored Regions :
Freq. confidence regions
($\Delta\chi^2$ at two d.o.f's)

Contours :
Bayesian credible regions
(X % of the volume)

Theoretical uncertainties from marginalization

- We denote both Frequentist profiling and Bayesian integration by Marginalization.

- The common philosophy is that the nuisance parameters participate into the fit. Through these procedures, values of the nuisance parameters are automatically chosen as to favor the best fit.

- In Bayesian: $\hat{L}(c_V, c_f) = \int_{\mathcal{D}_\delta} d\delta_1 d\delta_2 \, L(c_V, c_f, \delta_i) p(\delta_1, \delta_2)$
- In Frequentist : $\hat{L}(c_V, c_f) = \max_{\delta_1, \delta_2 \in \mathcal{D}_\delta} \left[L(c_V, c_f, \delta_i) p(\delta_1, \delta_2) \right]$

With $\mu_i^{exp} \to \mu_i^{exp}(1 + \delta \times \Delta_i)$, $\delta \in [-1, 1]$, $p(\delta) = cste$

- In Frequentist : $\hat{L}(c_V, c_f) = L(c_V, c_f, \xi_X)$

$$\xi_X = \begin{cases} \zeta_X/\eta_X & \text{if } \xi_X \in [\delta_{X1}, \delta_{X2}] \\ \delta_{X1} & \text{if } \xi_X < \delta_{X1} \\ \delta_{X2} & \text{if } \xi_X > \delta_{X2} \end{cases}$$

$$\eta_X = \sum_i \Delta_i^2 \mu^t_{exp\,i} \Sigma_i^{-1} \mu_{exp\,i} \qquad \zeta_X = \sum_i \Delta_i (\mu_i - \mu_{exp\,i})^t \Sigma_i^{-1} \mu_{exp\,i}$$

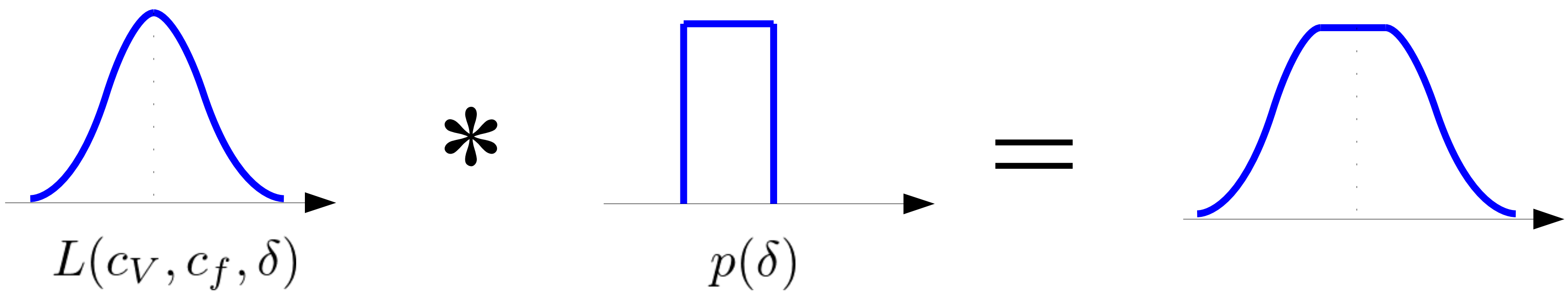
- In Bayesian :

$$\hat{L} = L_0 \prod_X e^{\zeta_X^2/2\eta_X} \left\{ \mathrm{Erf}\left(\frac{\sqrt{\eta_X}\delta_{X2}}{\sqrt{2}} - \frac{\zeta_X}{\sqrt{2\eta_X}} \right) - \mathrm{Erf}\left(\frac{\sqrt{\eta_X}\delta_{X1}}{\sqrt{2}} - \frac{\zeta_X}{\sqrt{2\eta_X}} \right) \right\} ,$$

- In our case, $X = \{ggF/tth; VBF/Vh\}$

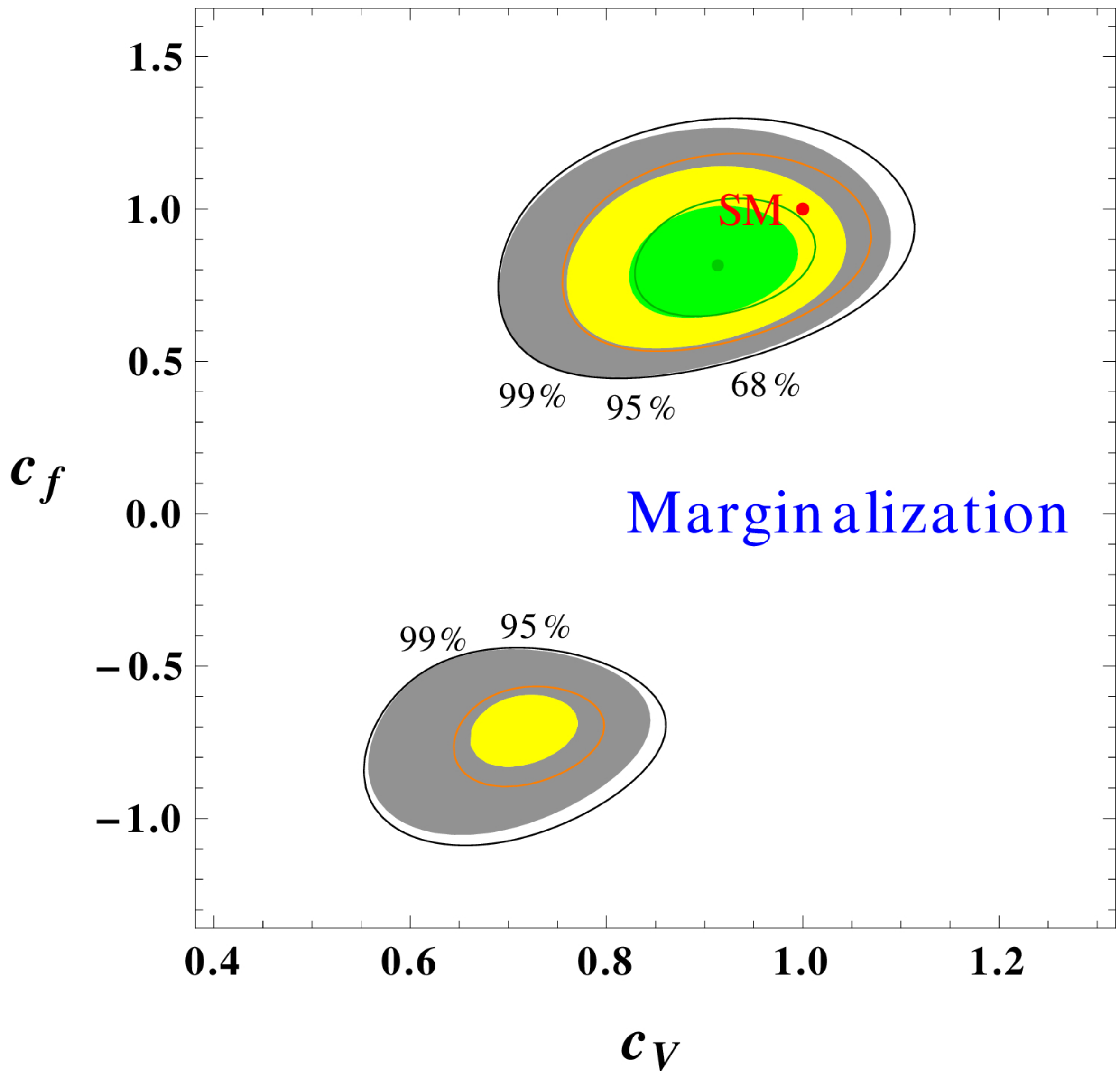
- Both Frequentist and Bayesian are analytic : computationally light

- For a single constraint: just enlargement

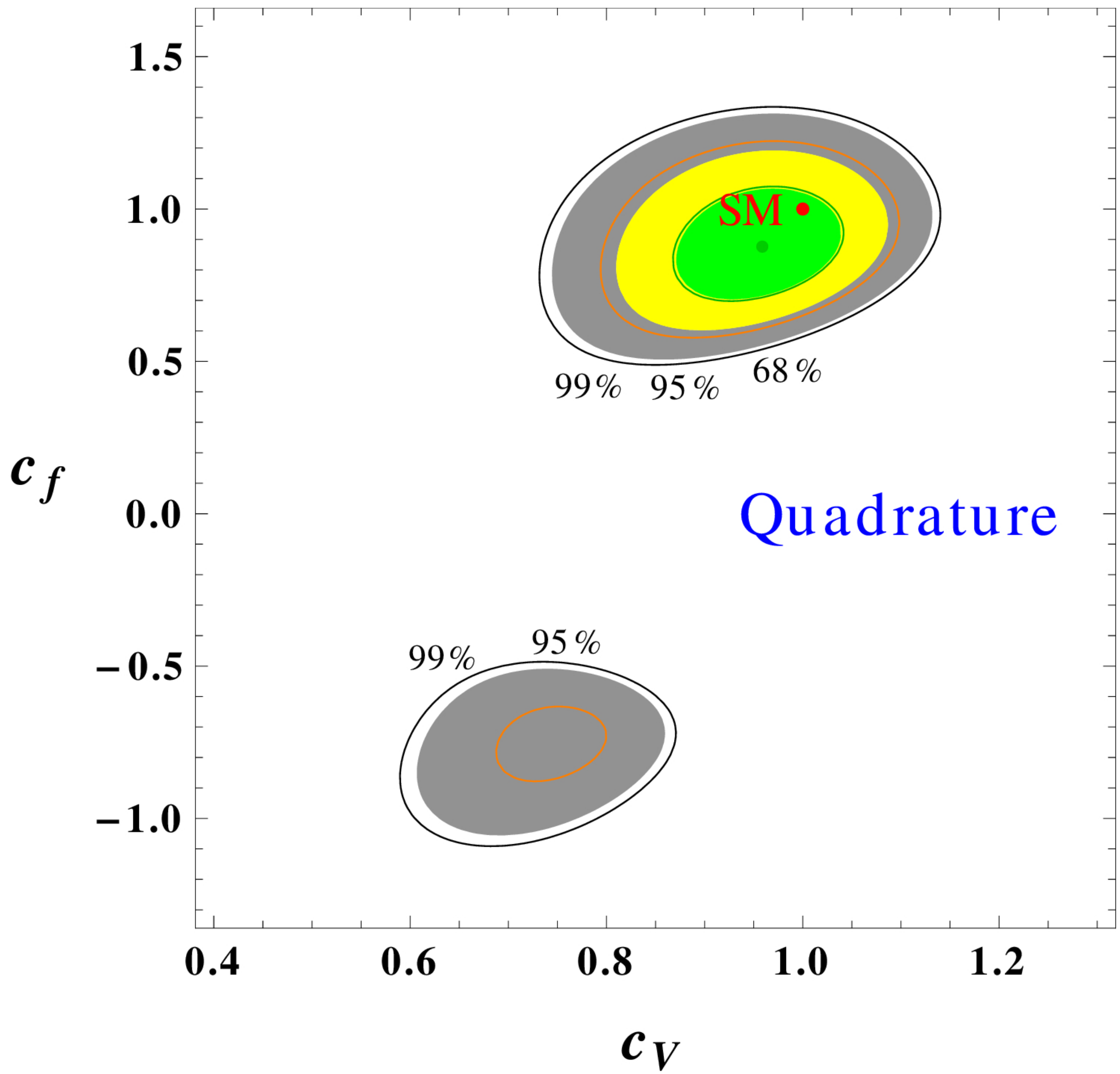


- For several constraints with correlations : δ also changes the tensions between the different subchannels

- Theoretical correlations induce a shift
- SM point goes out from the 68% regions

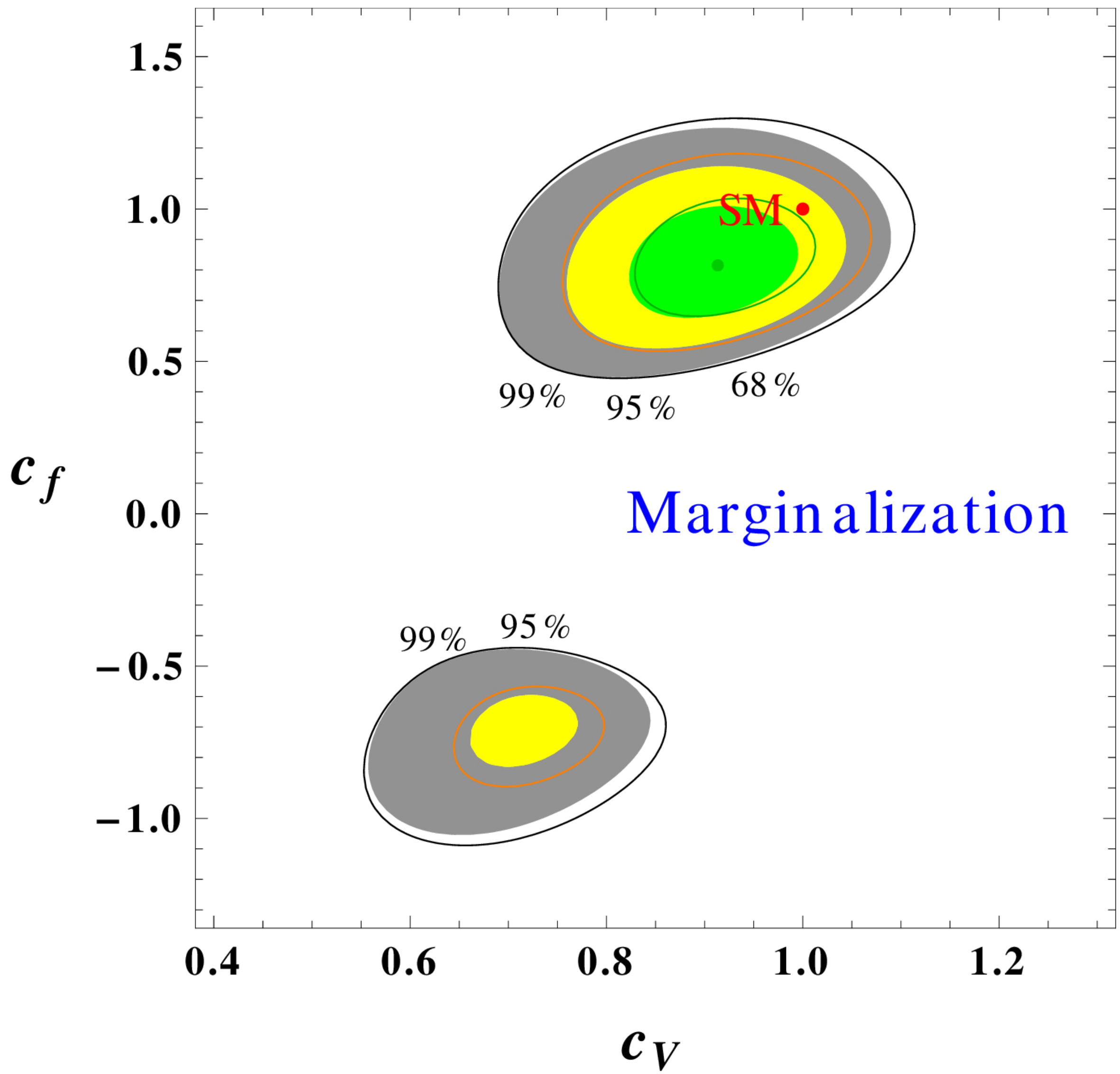


- Theoretical correlations induce a shift
- SM point goes out from the 68% regions



- Theoretical correlations induce a shift
- SM point goes out from the 68% regions
- The value $\delta = -1$ is preferred

(here $\delta \equiv \delta_1 = \delta_2$ for simplicity)



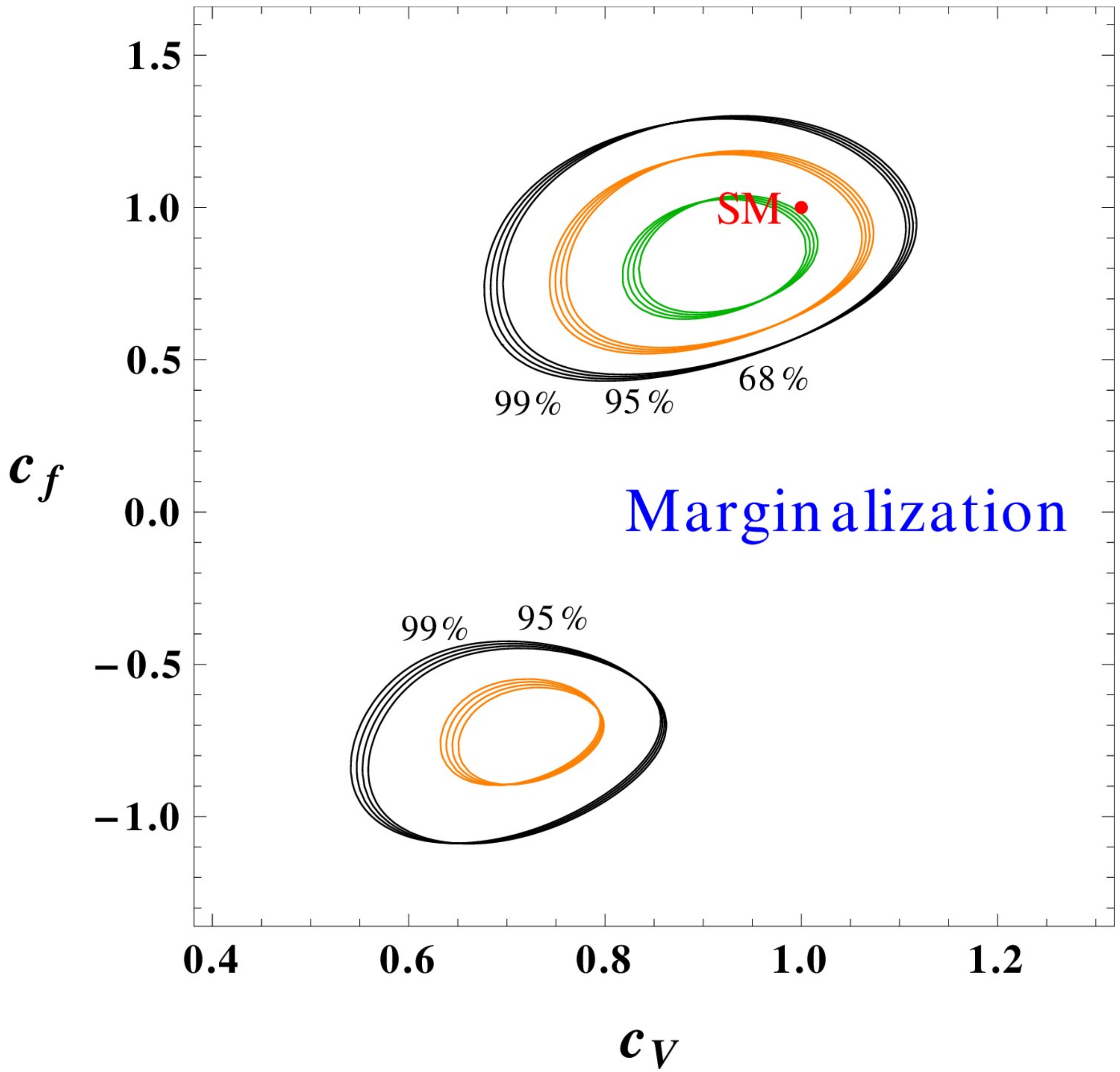
- One should check the prior dependence : how our results vary for an estimation more optimistic or pessimistic of the QCD uncertainty.

- Let us rescale everything (amounts to change the range of δ)

$$\mu_i^{exp} \to \mu_i^{exp}(1 + \delta \times \Delta_i) \qquad \delta \in [-1, 1]$$

- For

$$\delta = [-0.9, 0.9], [-1, 1], [-1.1, 1.1], [-1.2, 1.2]$$



Marginalization

Although marginalization is well-defined, one may find not conservative enough the fact that the nuisance parameters contribute to goodness-of-fit

⇨ Let's find an alternative/complementary approach to marginalization

Theoretical uncertainties from bias

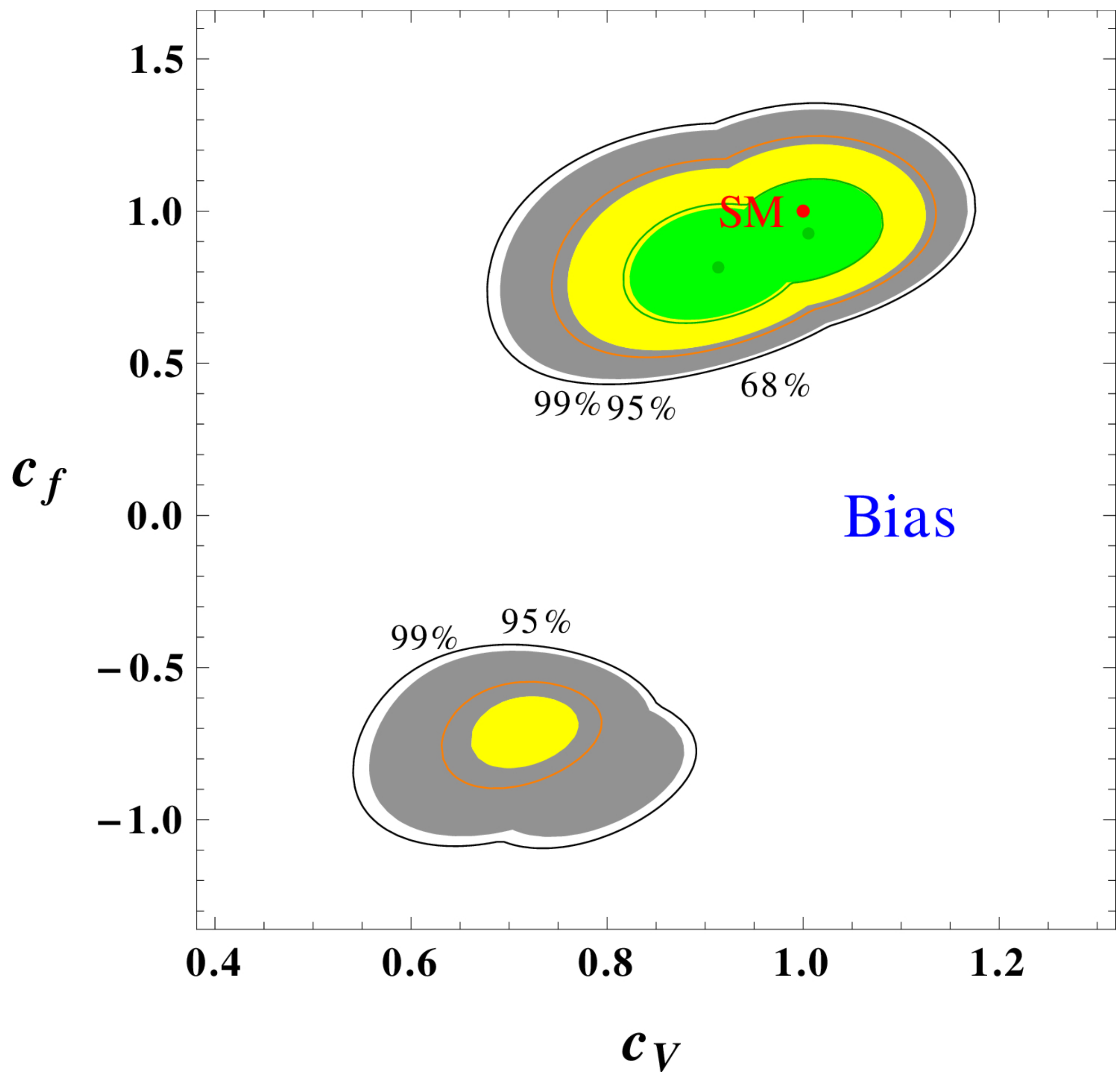
- Let us demand that the nuisance parameters do NOT participate into the fit. We denote this approach BIAS in both Bayesian and Frequentist.

- Simplest version of the idea: showing the ellipses for the extreme values of δ

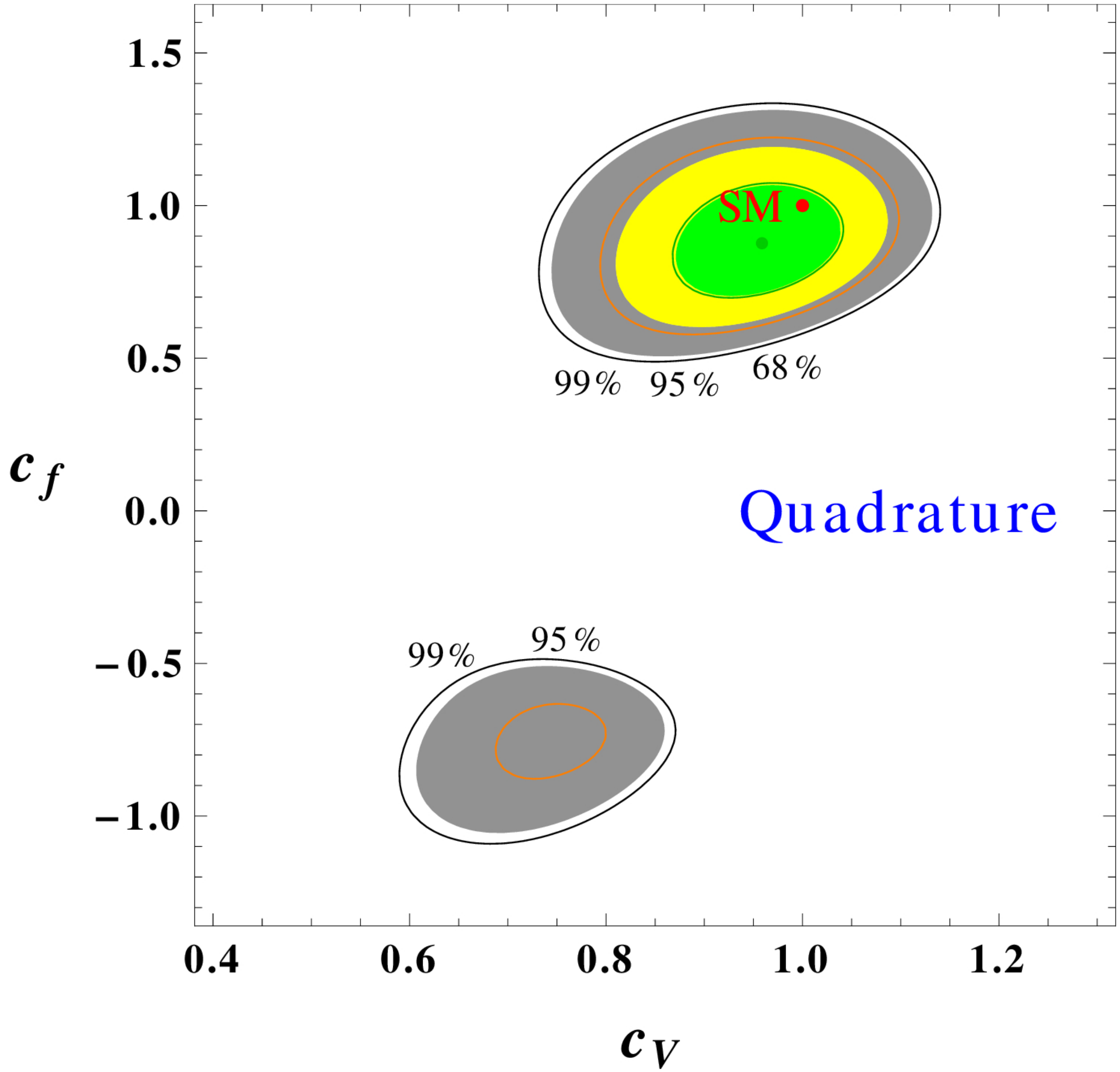
  Approach more conservative. Requires hard bounds on δ .

- Bias with extreme ellipses for $\delta \equiv \delta_1 = \delta_2$

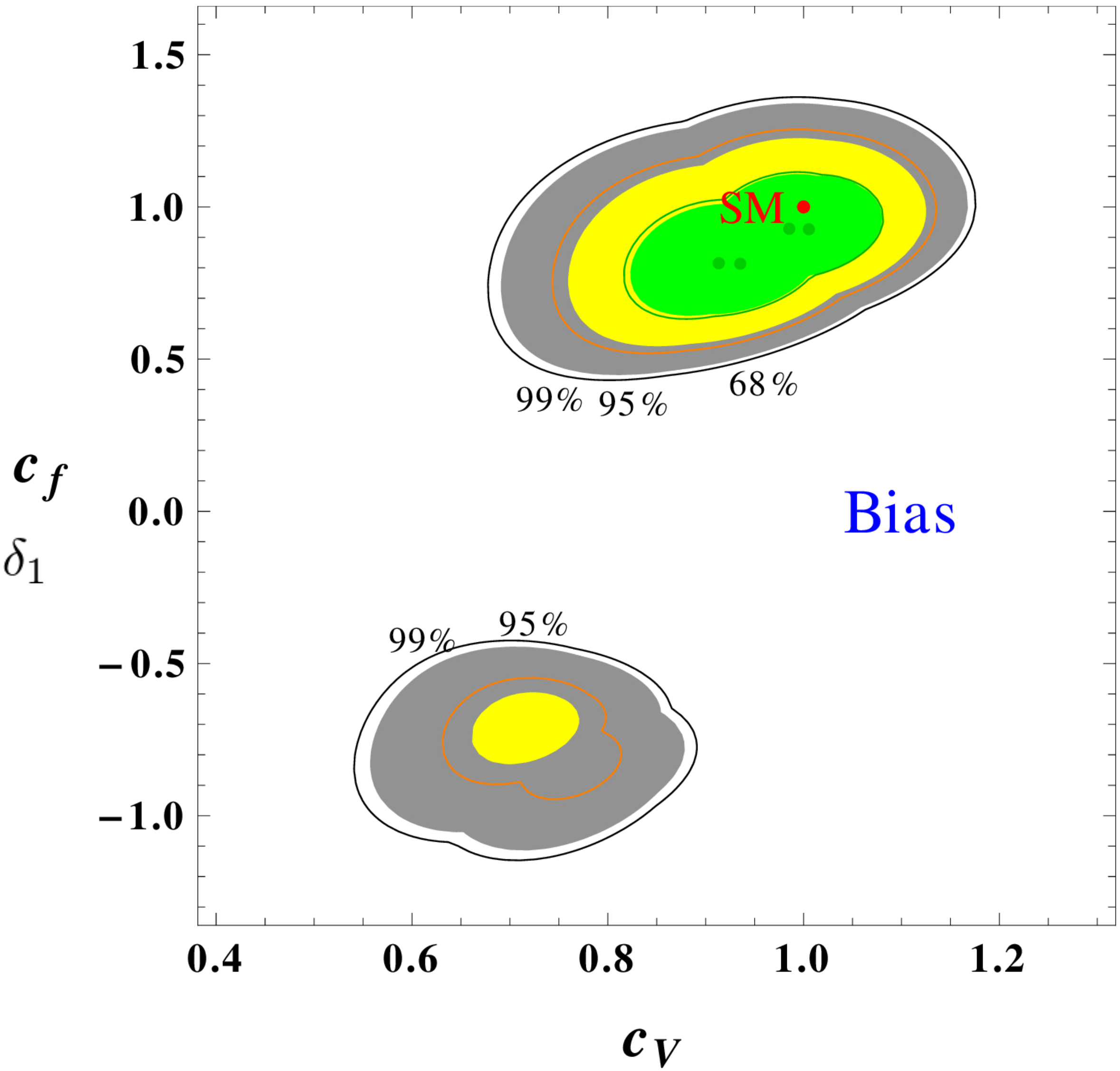
- The lower ellipse corresponds roughly to the one of marginalization



- Bias with extreme ellipses for $\delta \equiv \delta_1 = \delta_2$
- The lower ellipse corresponds roughly to the one of marginalization



- Bias with four ellipses for independent δ_1, δ_2
- The larger shift comes from δ_1 associated with ggF, tth



- Let us demand that the nuisance parameters do NOT participate into the fit. We denote this approach BIAS in both Bayesian and Frequentist.

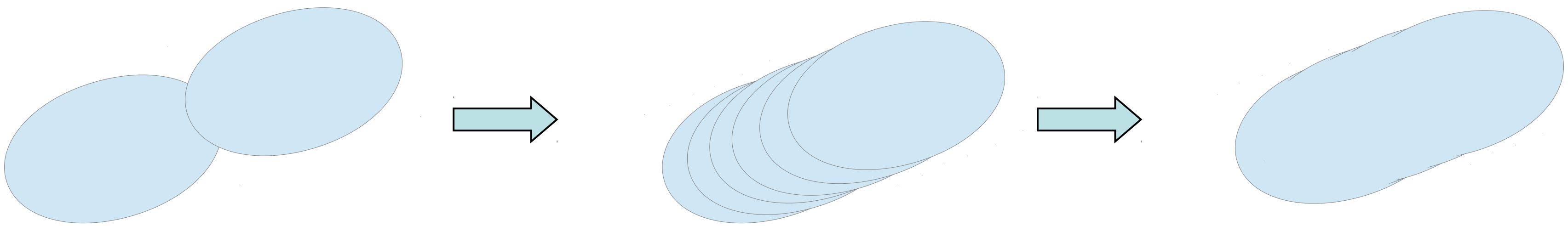
- Simplest version of the idea: showing the ellipses for the extreme values of δ

  Approach more conservative. Requires hard bounds on δ .

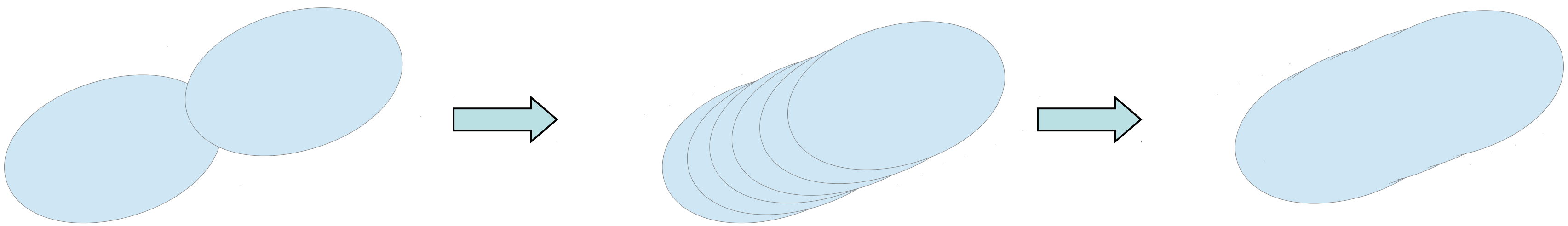
- Fine for these Higgs fits, but not so appropriate for a more general situation.

  ⇨ Let's try to define a continuous version

- Instead of taking the extrema, let us consider the union of regions for N different values of δ, and let $N \to \infty$. The boundary defines a contour $\mathcal{C}$.



- Instead of taking the extrema, let us consider the union of regions for N different values of δ, and let $N \to \infty$. The boundary defines a contour $\mathcal{C}$.



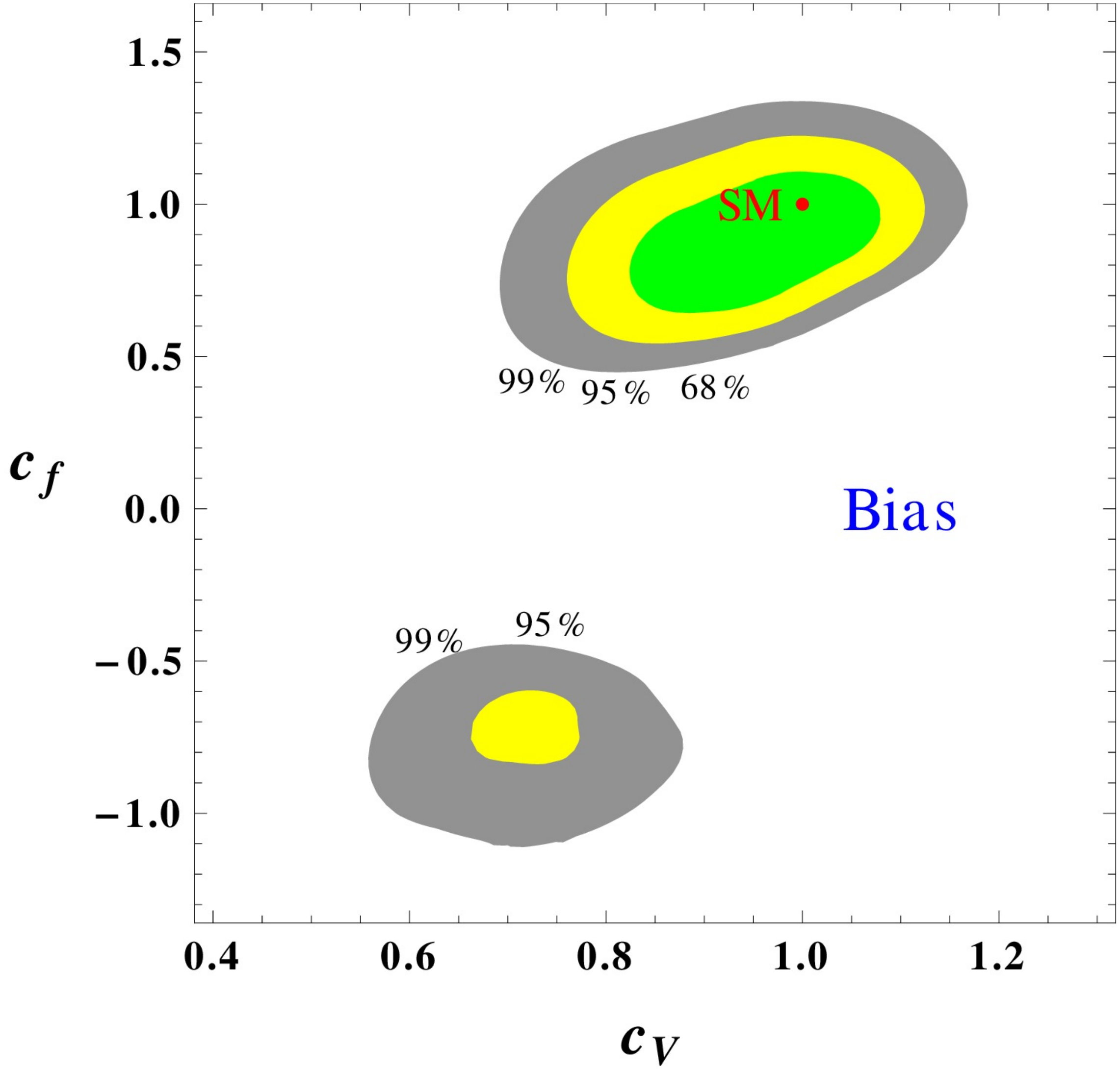
- Under conditions on the δ- dependence of the likelihood which are mostly satisfied in our case, one can show that $\mathcal{C}$ can be obtained from the contours of the likelihood

$$\hat{L}(c_V, c_f) = \max_{\delta \in \mathcal{D}_\delta} [L(c_V, c_f, \delta) / \max_{c \in \mathcal{D}_c} [L(c_V, c_f, \delta)]]$$

- This likelihood looks like a profile likelihood, but with the unprofiled likelihood normalized for any δ, such that δ does NOT contribute to the goodness-of-fit. This implements explicitely the principle of the bias.

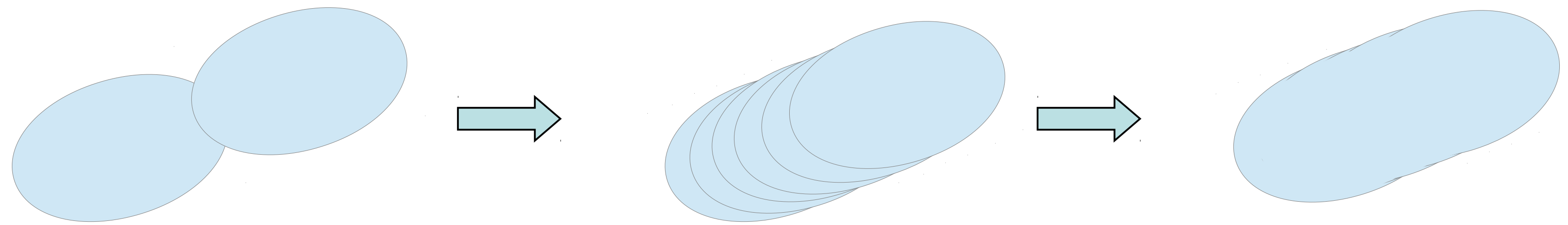
Results: Frequentist bias

- Independent δ_1, δ_2
- Obtained using interpolated likelihoods



Frequentist bias

- Instead of taking the extrema, let us consider the union of regions for N different values of δ, and let $N \to \infty$. The boundary defines a contour $\mathcal{C}$.



- Under conditions on the δ- dependence of the likelihood which are mostly satisfied in our case, one can show that $\mathcal{C}$ can be obtained from the contours of the likelihood

$$\hat{L}(c_V, c_f) = \max_{\delta \in \mathcal{D}_\delta}[L(c_V, c_f, \delta) / \max_{c \in \mathcal{D}_c}[L(c_V, c_f, \delta)]]$$

- This likelihood looks like a profile likelihood, but with the unprofiled likelihood normalized for any δ, such that δ does NOT contribute to the goodness-of-fit. This implements explicitely the principle of the bias.

- This continuous approach is better, however it is difficult to give a statistical meaning to the contours.

Let's try Bayesian

- The Bayesian version of the continuous bias can be derived directly from the starting principle. It translates into demanding independence of the posterior $p(\delta|d)$ on δ,

$$\frac{\partial}{\partial \delta_i} p(\delta_i|d) \equiv 0$$

- This enforces the prior to be

$$p(\delta_i) \propto 1/p(\delta_i|d) = 1/\int_{\mathcal{D}_c} dc_V dc_f \, L(c_V, c_f, \delta_i) p(c_V, c_F)$$

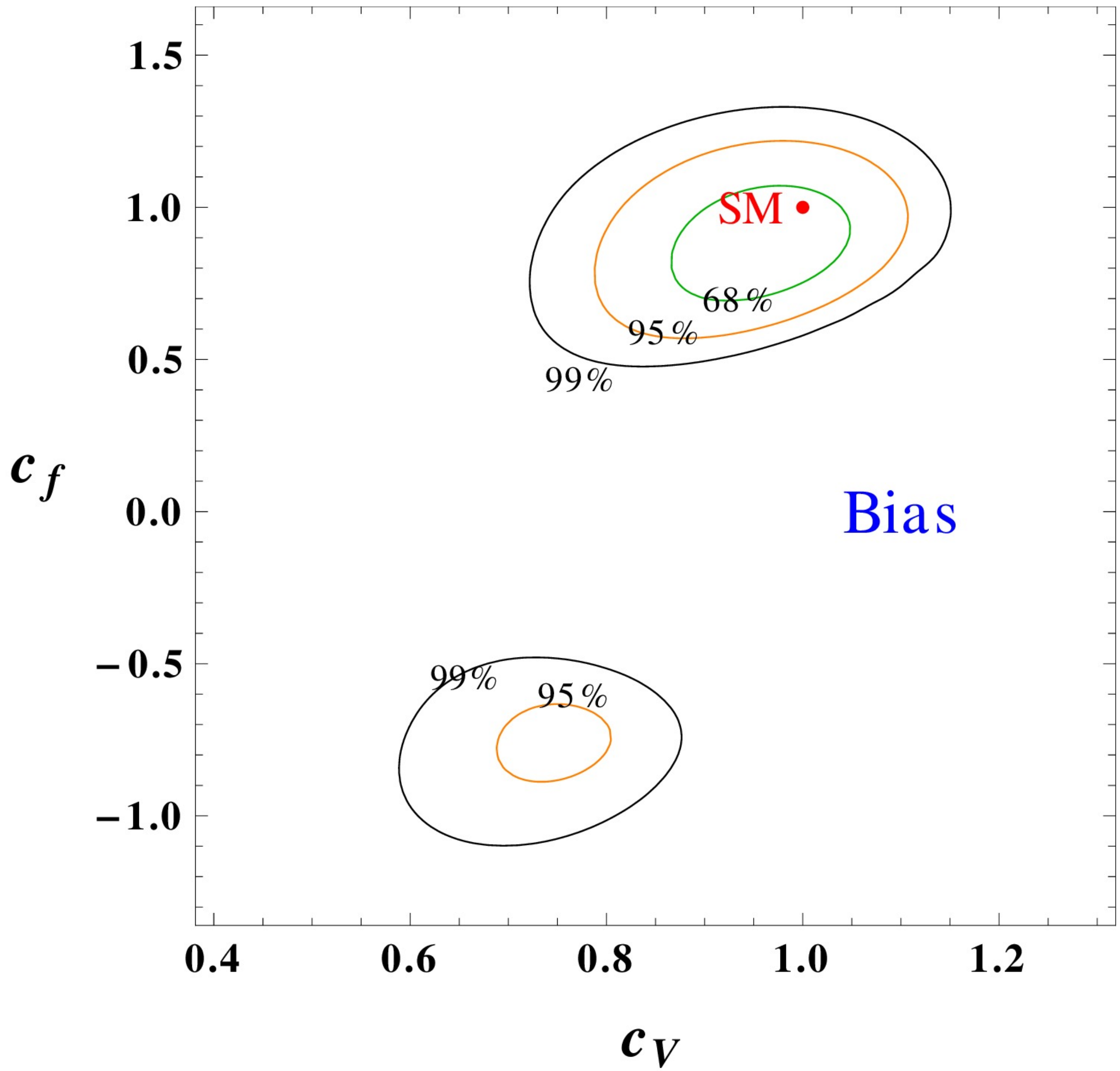
The posterior in the (c_V, c_f) plane is therefore

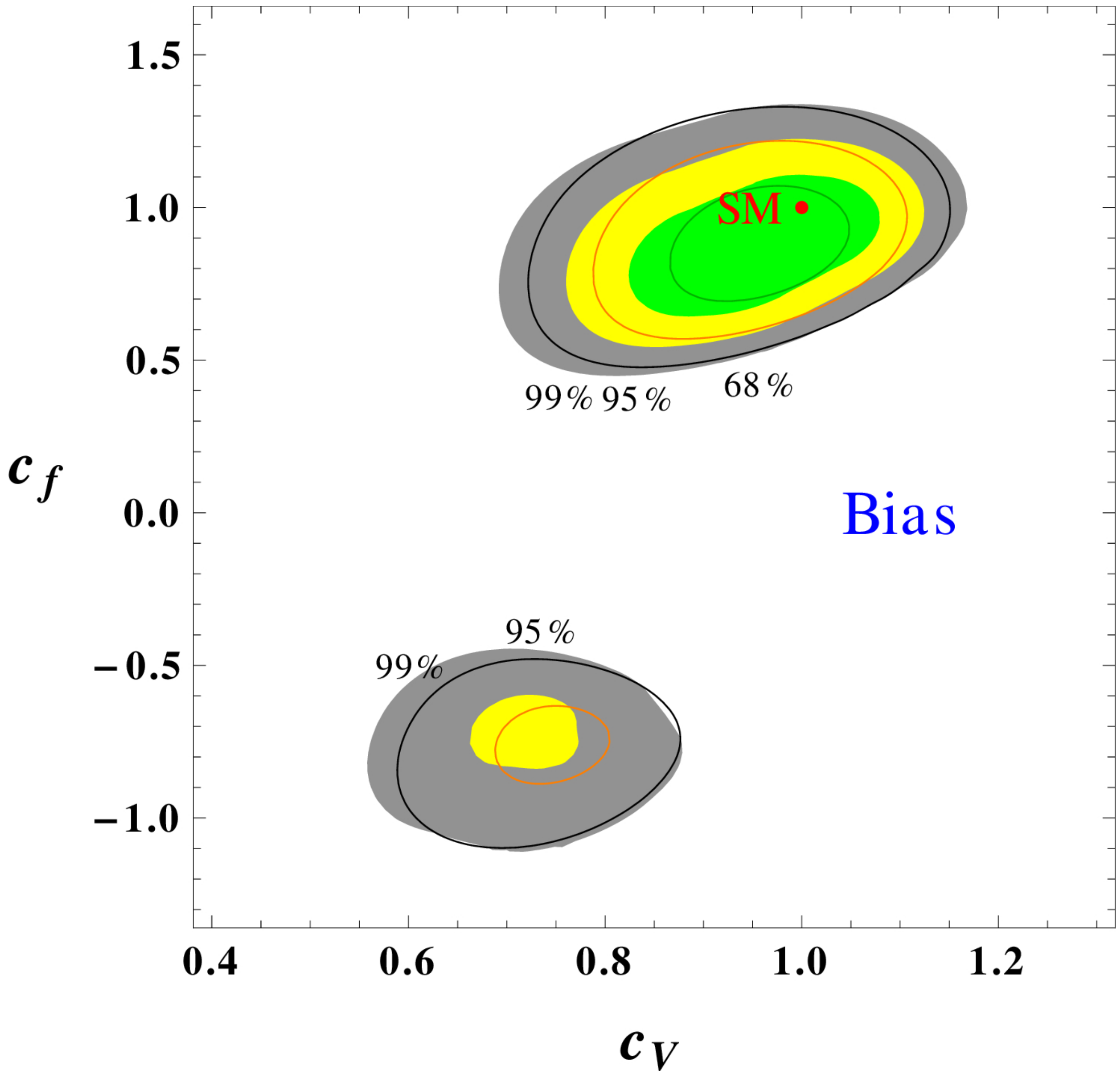
$$p(d|c_V, c_f) \propto \int_{\mathcal{D}_\delta} d\delta_i \left[L(c_V, c_f, \delta) p(c_V, c_f) / \underbrace{\int_{\mathcal{D}_c} dc_V dc_f \, L(c_V, c_f, \delta_i) p(c_V, c_F)}_{p(\delta_i)} \right]$$

- This prior is such that it cancels the contribution to goodness of fit from δ.

- The Bayesian Credible regions are well-defined as always.

- Independent δ_1, δ_2
- Obtained using interpolated likelihoods




1.5
1.0
0.5
0.0
−0.5
−1.0
0.4
0.6
0.8
1.0
1.2
c_f
c_V
SM
99 %
95 %
68 %
99 %
95 %
Bias

Summary

Summary

- The Higgs precision physics era begins, New Physics may show up as small anomalies in Higgs properties. Theoretical uncertainties are relevant, and will more and more dominate. A good understanding and treatment of these TUs is necessary.

- Combining TUs with EUs in quadrature is inadequate. $\sigma_i = \sqrt{(\sigma_i^{exp})^2 + (\sigma_i^{th})^2}$

- Using marginalization, we provide analytic formulas accounting for all the correlations in both Freq. and Bayes. Their use requires same CPU power as the usual combination in quadrature.

- We develop the bias approach in both Freq. and Bayes., complementary from marginalization. Formulas are heavier to use, but still manageable using interpolation methods.

- All these methods we apply to Higgs fit in the (c_V, c_f) plane using latest available data. Substantial changes appear because correlations are taken into account.

Summary

$$\sigma_i = \sqrt{(\sigma_i^{exp})^2 + (\sigma_i^{th})^2}$$

	Freq. Marg.	Bayes. Marg.	Freq. Bias.	Bayes. Bias.
Well-defined contours	✓	✓	~	✓
CPU	+	+	-	-
Conservative	-	-	++	+

Thank you for your attention !!!