

Misura differenziale della produzione del bosone di Higgs nel canale di decadimento WW con reti neurali indipendenti dal modello

BENEDETTA CAMAIANI – UNIVERSITÀ DEGLI STUDI DI FIRENZE

Introduzione

- Esperimento CMS@LHC

Run 2
2016 – 2018

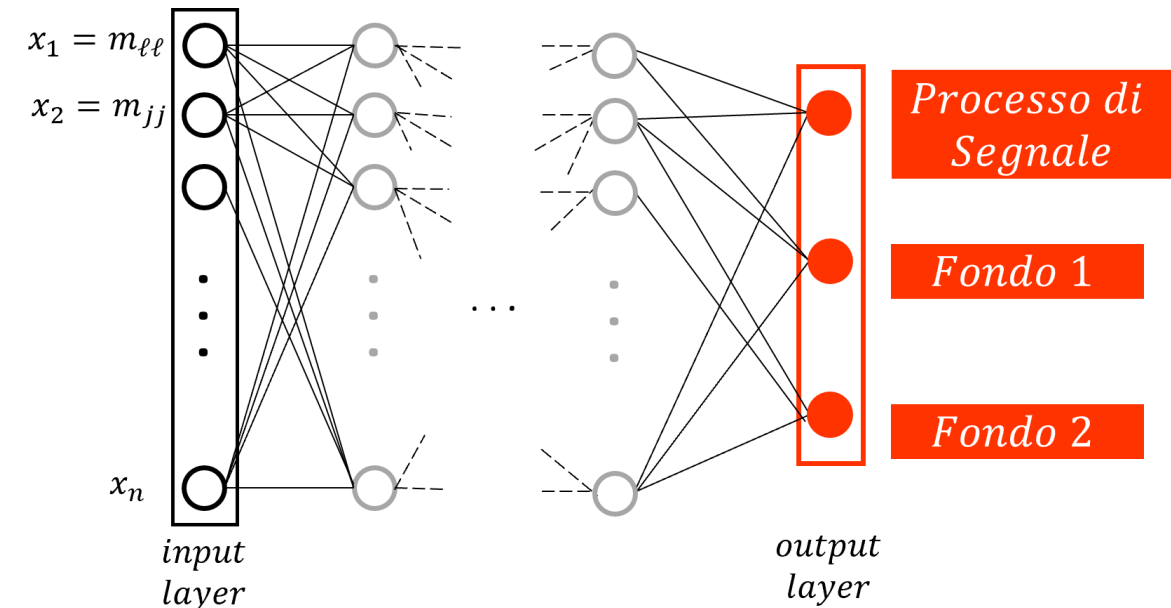
Run 3
2022 – 2024

- Stato dell'arte: utilizzo di reti neurali per la discriminazione dei processi di segnale da quelli di fondo

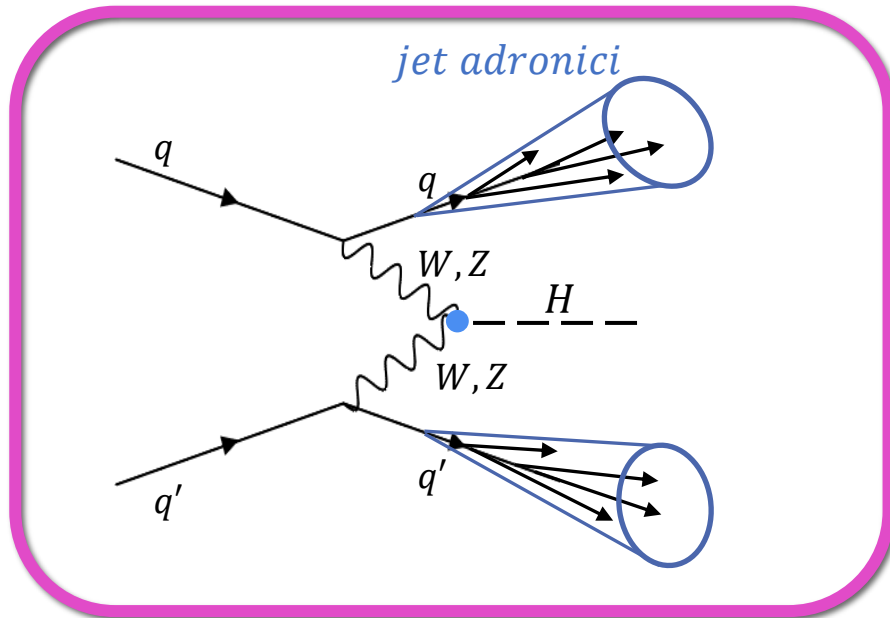
- L'addestramento della rete neurale richiede l'assunzione di un modello di segnale

Possibile «bias» verso il modello assunto nel risultato finale

└───┬───> Domain Adaptation



Fusione di Bosoni Vettori (VBF)



- Meccanismo di produzione raro
- Sensibile a processi di Nuova Fisica:
Modelli di Fisica *Beyond the Standard Model* (BSM) predicono strutture diverse degli accoppiamenti HVV

Misurare la sezione d'urto in modo indipendente dal modello fisico che descrive l'accoppiamento HVV

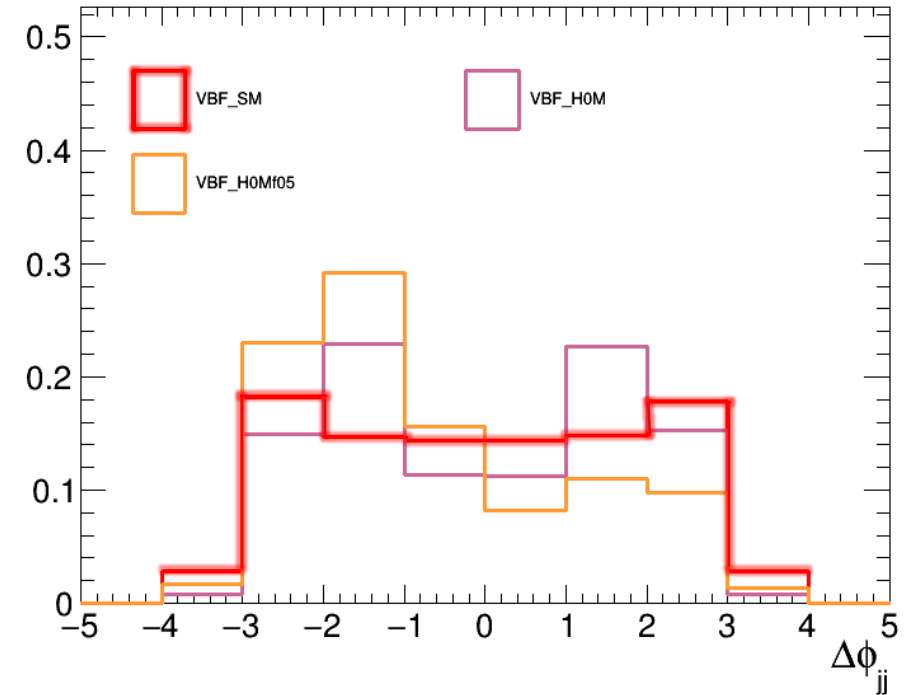
- Misura differenziale

Misura di sezione d'urto differenziale

- Differenza in angolo azimutale ϕ tra i due jet adronici dello stato finale ($\Delta\Phi_{jj}$) particolarmente sensibile a fisica oltre il Modello Standard (BSM)

$$\Delta\Phi_{jj} = \phi_{jk} - \phi_{jl} \quad \text{con} \quad \eta_{jk} > \eta_{jl}$$

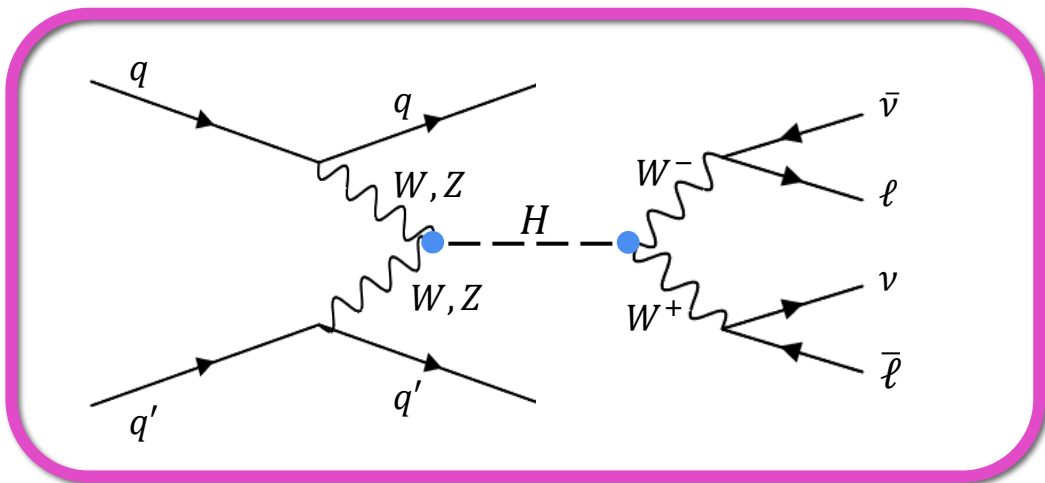
- Le diverse strutture tensoriali dovute ad *accoppiamenti anomali* nel vertice HVV possono modificare la distribuzione di $\Delta\Phi_{jj}$
- La distribuzione di $\sigma(\Delta\Phi_{jj})$ può essere diversa dalle predizioni del MS



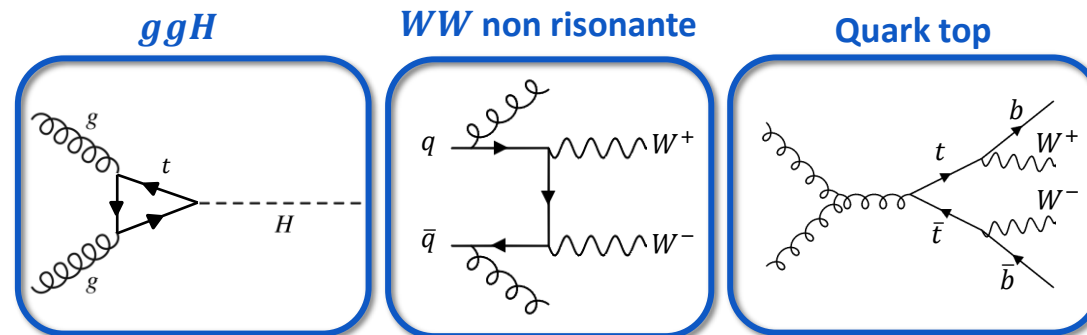
⇒ Strumento efficace per testare ipotesi alternative rispetto al MS

⇒ È opportuno limitare la dipendenza dal modello dell'analisi al fine di evitare che il risultato abbia un «bias» verso la predizione del MS

VBF in $H \rightarrow WW$



- Canale di decadimento considerato: $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$
 - BR elevato in $WW \rightarrow e^\pm \mu^\mp \nu \bar{\nu}$
 - Fondo relativamente basso



- Si riescono a ricostruire bene i leptoni carichi
- I neutrini nello stato finale sfuggono alla rivelazione

→ la massa invariante del sistema non può essere ricostruita

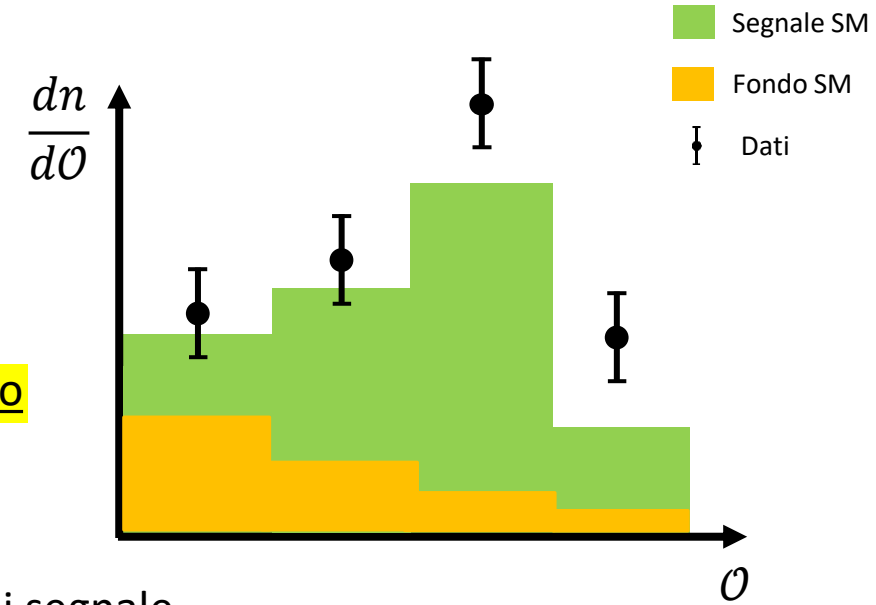
→ Per distinguere il segnale dai fondi si utilizza una combinazione di osservabili cinematiche alternative, di cui in generale non si conosce la distribuzione analitica



TEMPLATE FIT

Template fit

- Si costruisce un'osservabile \mathcal{O} con buona discriminazione segnale/fondo
- La p.d.f non è nota a priori e viene sostituita da istogrammi (*template*)
- Si genera la simulazione Monte Carlo di \mathcal{O} , assumendo un certo modello fisico
- Si considera la distribuzione di \mathcal{O} nei dati
- Si fitta la simulazione MC del segnale ai dati per ricavare il numero di eventi di segnale

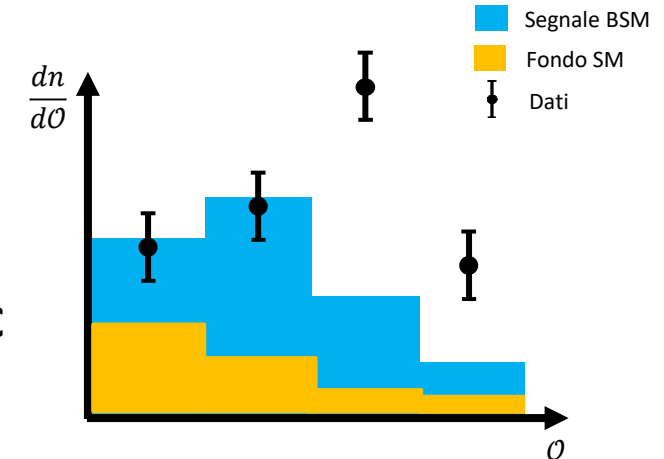


La distribuzione di \mathcal{O} in generale dipende dal modello teorico che si assume!

**Modelli di Fisica BSM
predicono distribuzioni
diverse di \mathcal{O}**



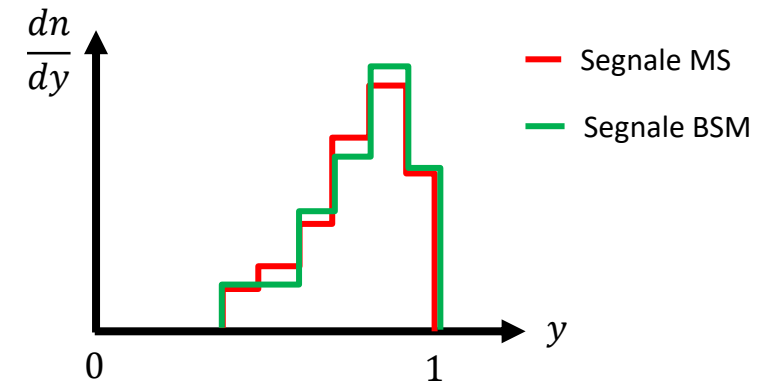
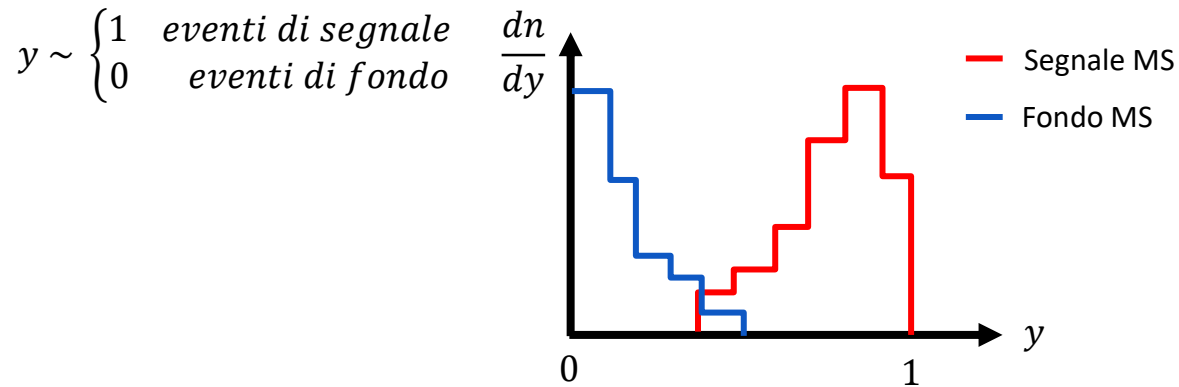
**Il risultato ha un «bias» verso la predizione del
modello utilizzato per generare la simulazione MC**



Obiettivo

Implementare una nuova variabile discriminante y che sia indipendente dai diversi modelli di fisica considerati, cioè:

- Discrimini bene tra il processo di segnale e i fondi
- **Non** discrimini tra il Modello Standard e i modelli BSM (ed in questo senso sia «model independent»)

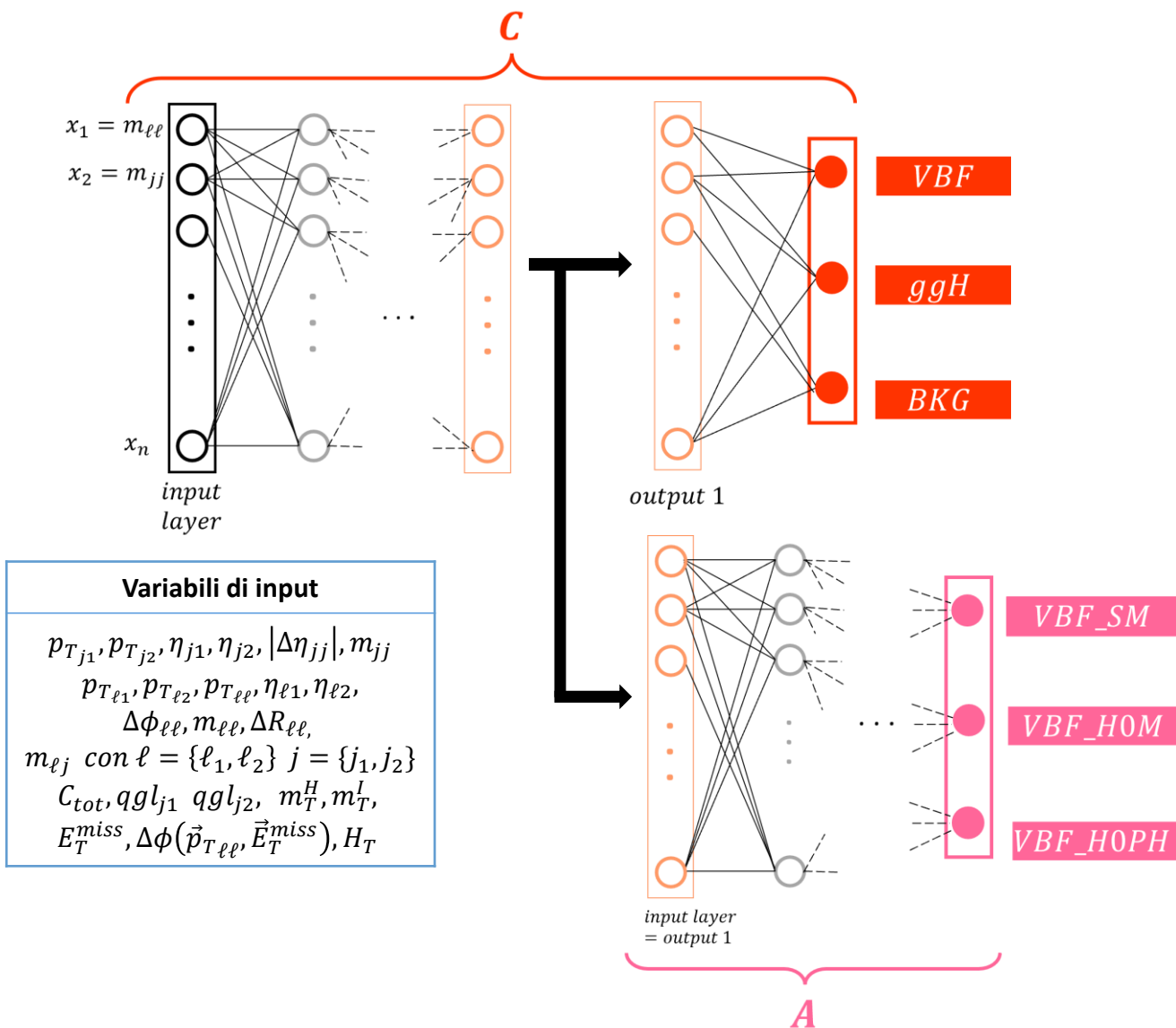


y non introduce un bias nel risultato del fit dal momento che la forma della sua distribuzione è \sim la stessa qualunque sia il modello teorico che descrive i dati



Domain Adaptation

Rete Neurale Avversaria



Classificatore

- Ogni output rappresenta la probabilità che un evento corrisponda ad uno dei processi fisici considerati
- È allenato in diversi «domini» corrispondenti a diversi modelli di segnale

Avversario

- È allenato solo su eventi di segnale
- Identifica il modello fisico di un evento di segnale

Training competitivo

- Il classificatore viene penalizzato se l'avversario riesce a risalire al modello fisico di un evento di segnale
- $\mathcal{L}oss(A)$
- $\mathcal{L}oss(C + A) = \mathcal{L}oss(C) - \alpha \cdot \mathcal{L}oss(A)$

se C riesce a prevenire che A identifichi il modello di segnale, allora la classificazione è indipendente dai domini di provenienza degli eventi

- Campioni di eventi usati per il training:
 - VBF nell'ipotesi SM
 - VBF nell'ipotesi H0M
 - VBF nell'ipotesi H0PH
 - ggH
 - WW non risonante
 - Eventi di quark top

La procedura di training è stato implementato utilizzando le librerie Keras e Tensorflow ed sfruttando le GPU di Google

Ottimizzazione

- Gli iperparametri che definiscono la ANN sono stati ottimizzati utilizzando il software [Optuna](#)
- L'ottimizzazione è stata eseguita in modo tale da:

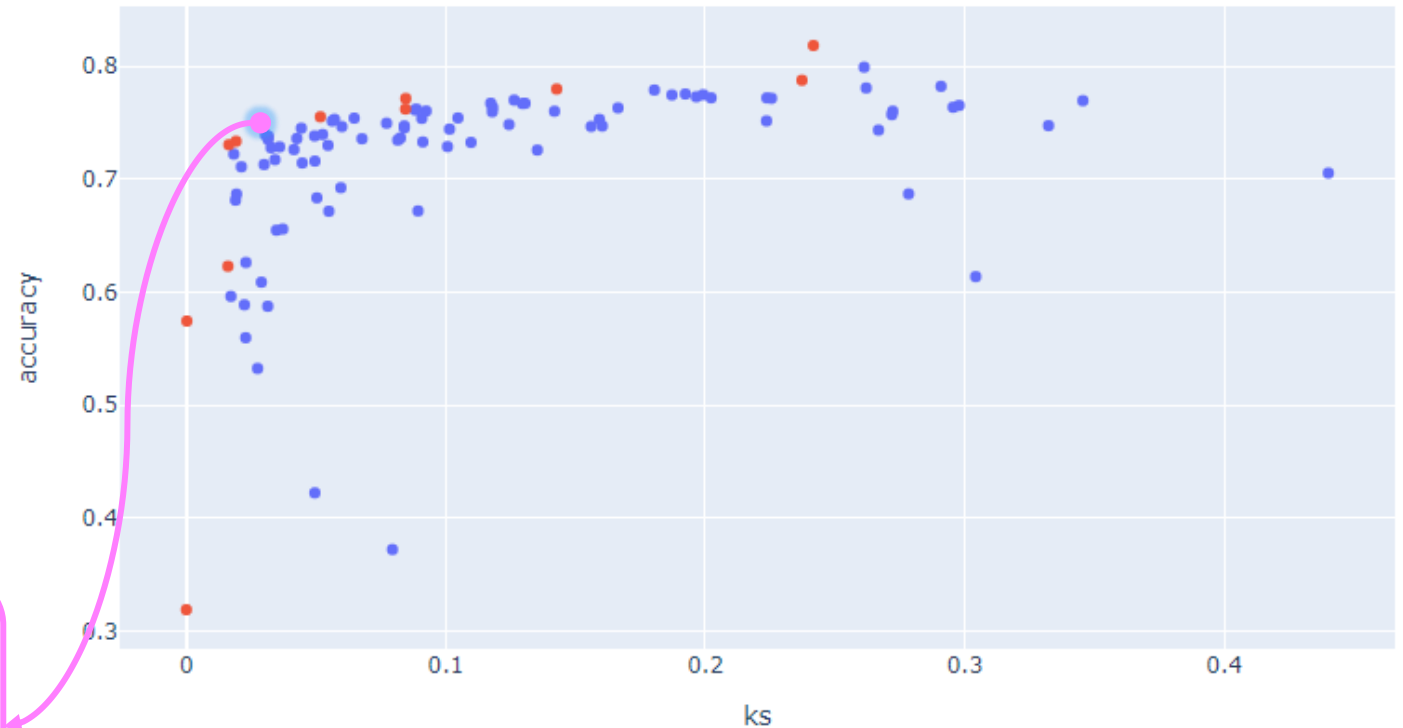
- Massimizzare l'**accuracy** del classificatore

- Minimizzare il **Kolmogorov-Smirnov test** tra le distribuzioni dell'output *VBF* ottenute su eventi di segnale SM e BSM:

$$ks = \frac{ks(SM, HOM) + ks(SM, HOPH)}{2}$$

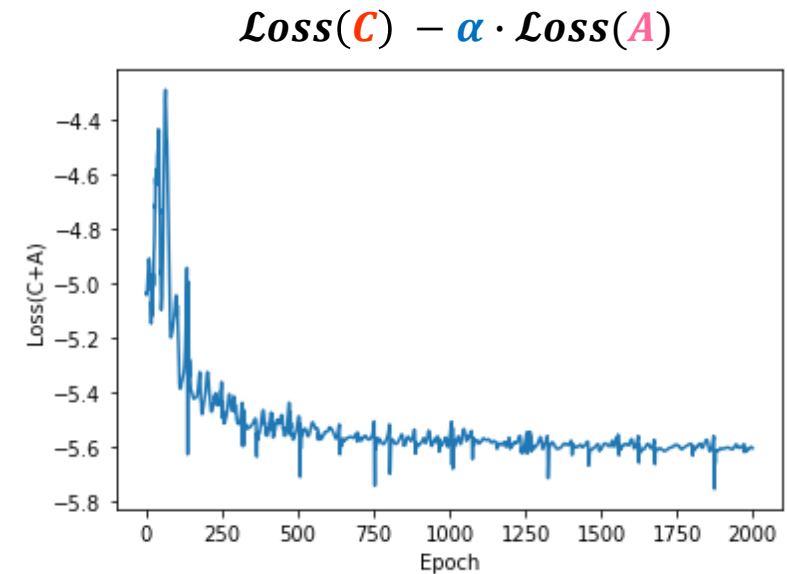
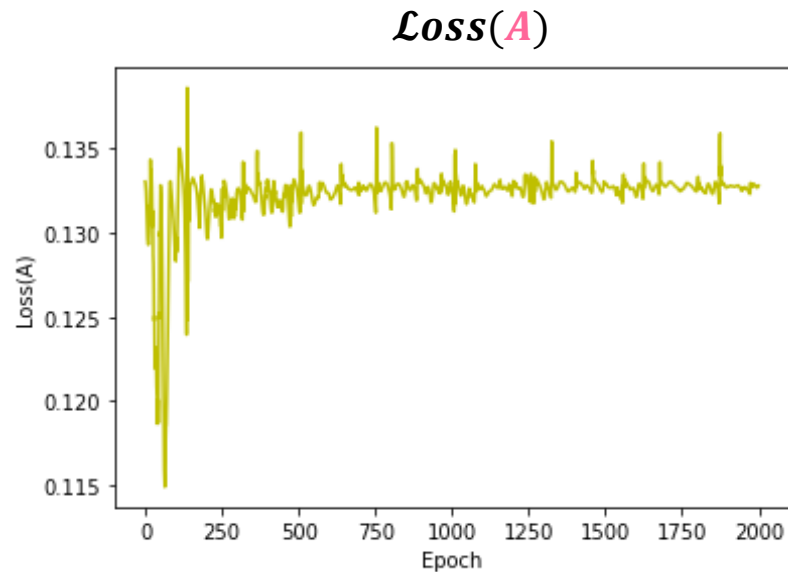
- 100 trial da 500 epoche ciascuno

- $\alpha = 46.2$
- # layers di C = 7
- # layers di A = 18
- # nodi in ogni layer = 94
- learning rate di C = 0.0003
- learning rate di A = 0.0016



Training ottimizzato

- Training da 2000 epoche
- $\mathcal{Loss}(A)$ e $\mathcal{Loss}(C)$ sono **Categorical Cross-Entropy**



- Tempo impiegato: **CPU** $\sim 1\ h\ 30\ minuti$ **GPU** $\sim 5\ minuti$

Prestazioni

- Lo spazio delle fasi è stato suddiviso in bin di $\Delta\Phi_{jj}$:

<i>Bin 0</i>	<i>Bin 1</i>	<i>Bin 2</i>	<i>Bin 3</i>
$-\pi < \Delta\Phi_{jj} < -\frac{\pi}{2}$	$-\frac{\pi}{2} < \Delta\Phi_{jj} < 0$	$0 < \Delta\Phi_{jj} < \frac{\pi}{2}$	$\frac{\pi}{2} < \Delta\Phi_{jj} < \pi$

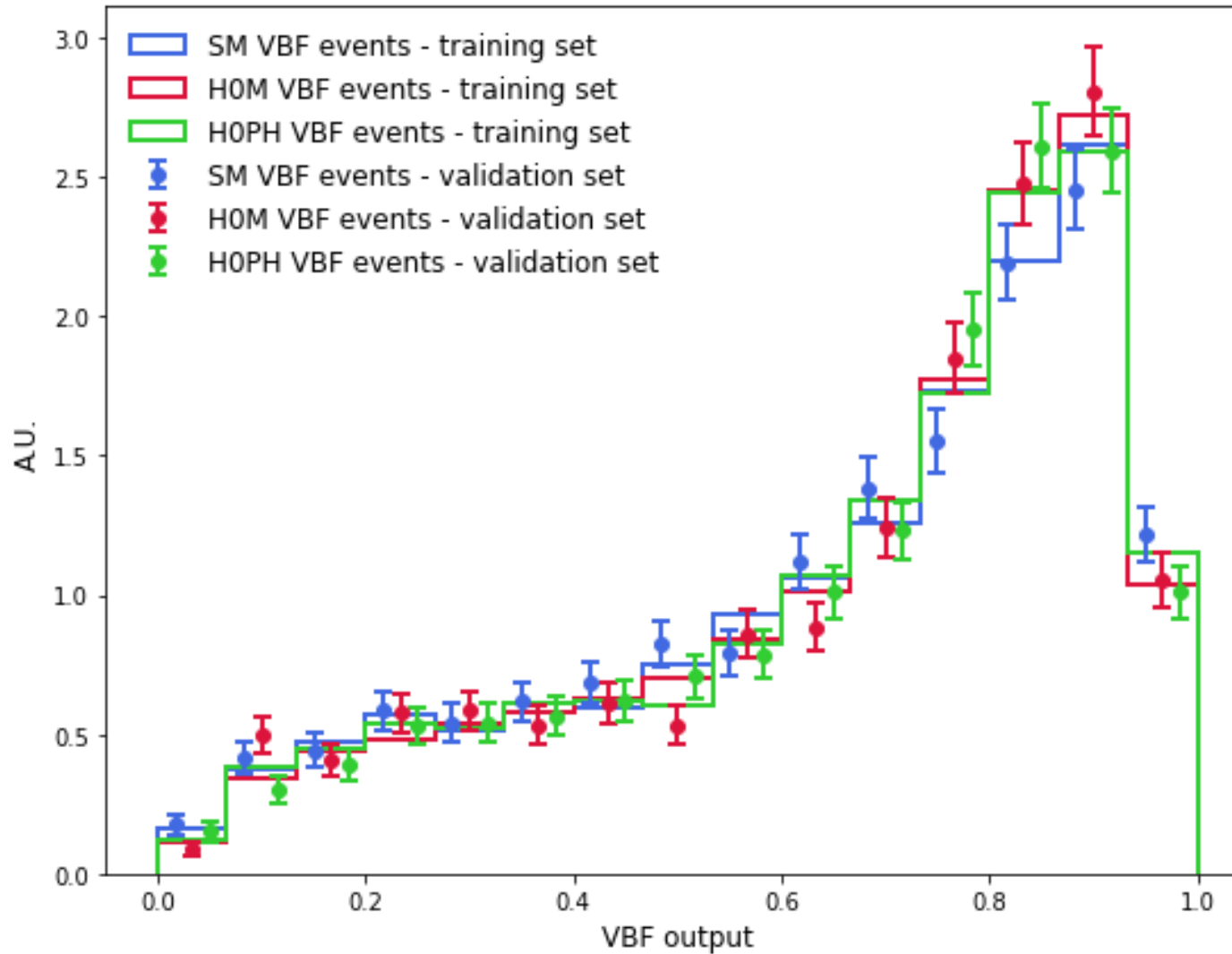
- Sono state allenate due ANN

- $|\Delta\Phi_{jj}| < \frac{\pi}{2}$

- $\frac{\pi}{2} < |\Delta\Phi_{jj}| < \pi$

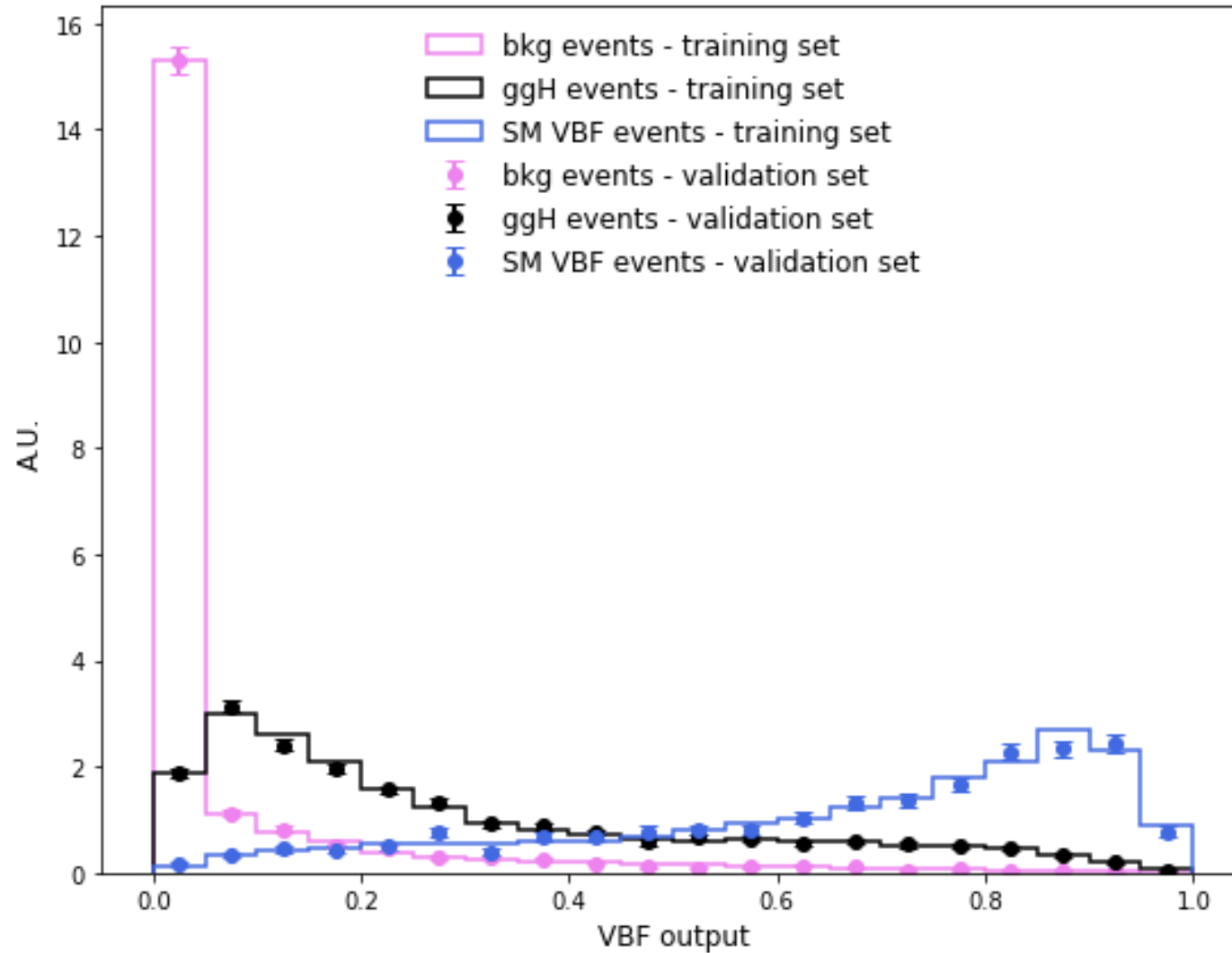
Prestazioni

- $|\Delta\Phi_{jj}| < \frac{\pi}{2}$



Prestazioni

- $|\Delta\Phi_{jj}| < \frac{\pi}{2}$



Prestazioni

- $|\Delta\Phi_{jj}| < \frac{\pi}{2}$

True label \ Predicted label	VBF	ggH	BKG
VBF	0.78	0.221	0.019
ggH	0.23	0.697	0.049
BKG	0.045	0.132	0.827

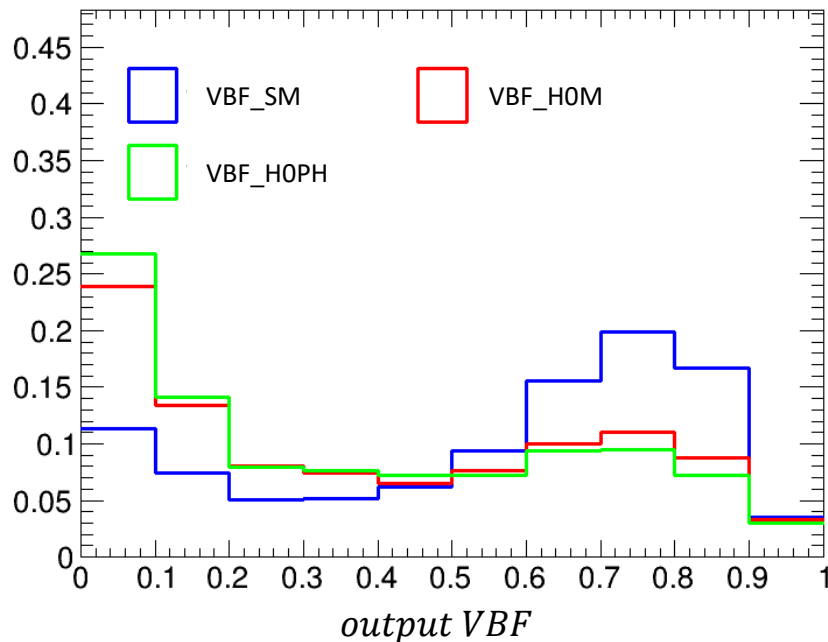
Estrazione del segnale

- La sezione d'urto è stata misurata eseguendo un fit multidimensionale:
 - Gli output del classificatore (*VBF*, *ggH* e *BKG*) sono stati utilizzati come variabili di fit
 - È stato generato il template di ogni variabile in ciascun bin di $\Delta\Phi_{jj}$ ed è stato fittato ai dati
- È stato utilizzato uno pseudo-dataset (*Asimov dataset*) definito in modo tale che il valore misurato della sezione d'urto sia uguale alla predizione del modello ipotizzato per generare la simulazione MC.
- Questa procedura permette di stimare le incertezze attese sulla misura con i dati veri

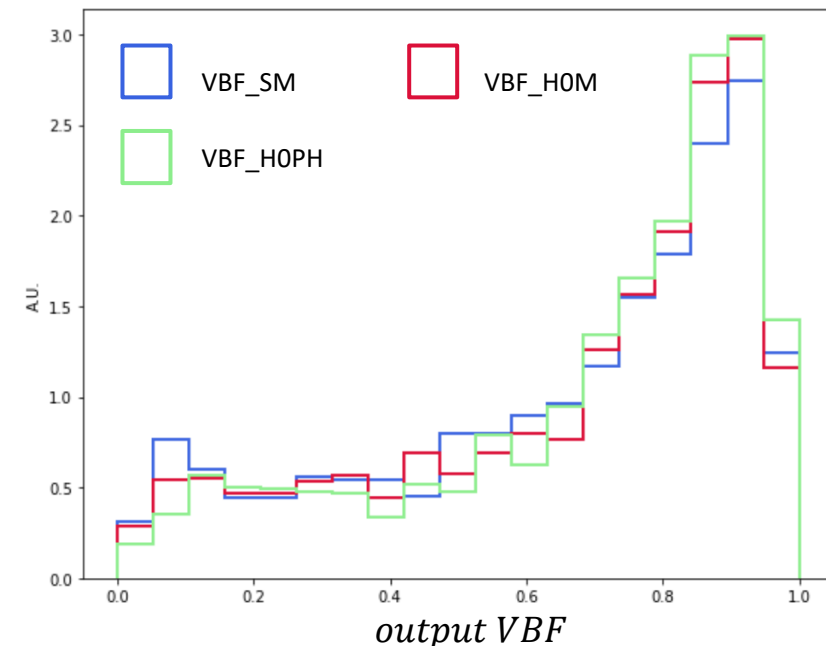
Studio della dipendenza dal modello

- Ho quantificato il «bias» del risultato della misura introdotto dalla procedura di fit
- Per la stima ho generato un Asimov dataset utilizzando l'ipotesi BSM ed ho eseguito il fit utilizzando la predizione del MS
- In questo caso il «bias» è dovuto al fatto che il fit viene eseguito con un'ipotesi fisica diversa da quella che realmente descrive i dati

DNN



ANN

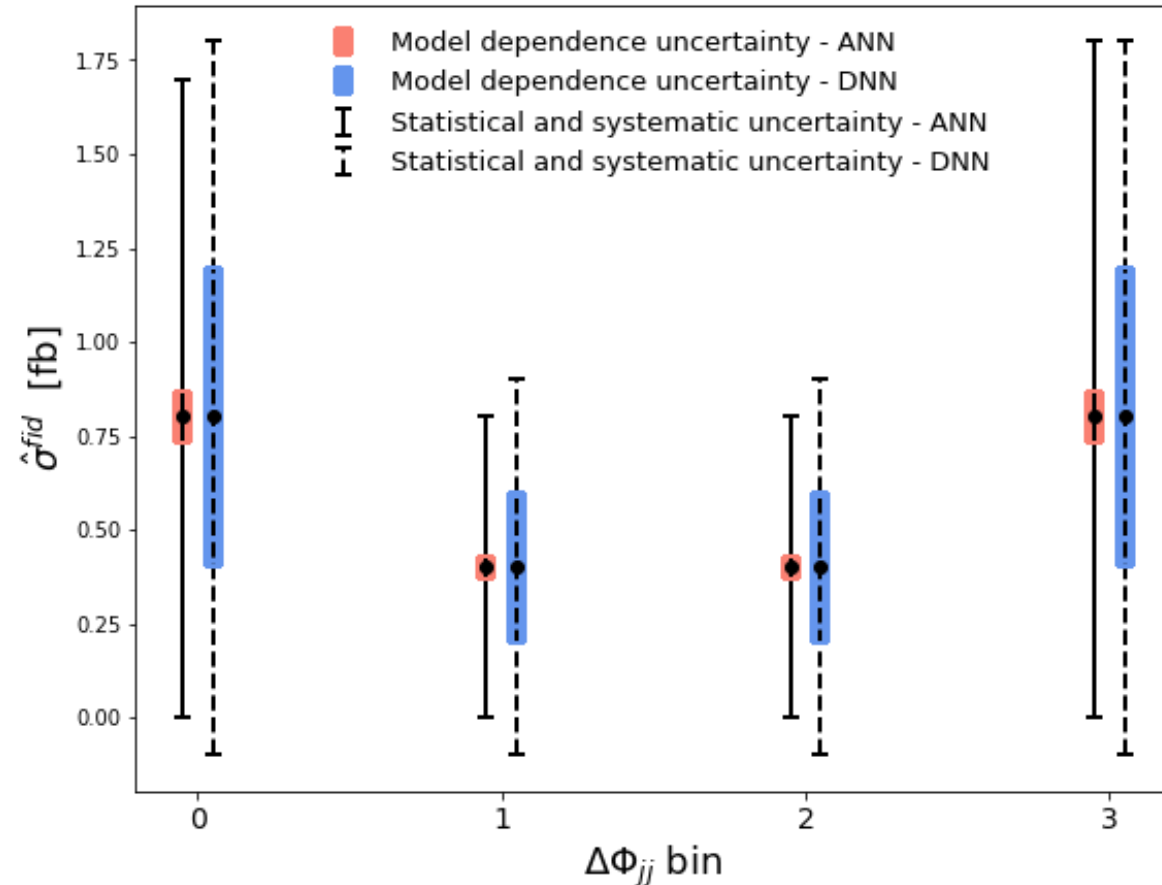


Studio della dipendenza dal modello

- Ho stimato un'incertezza da associare alla sezione d'urto misurata che tenga conto della dipendenza dal modello del risultato

<i>ANN</i>		
$\Delta\Phi_{jj}$ bin	$\hat{\sigma}^{fid}$ [fb]	$\Delta\sigma^{bias}$ [fb]
0	$0.8^{+0.9}_{-0.8}$	0.07
1	$0.4^{+0.4}_{-0.4}$	0.03
2	$0.4^{+0.4}_{-0.4}$	0.03
3	$0.8^{+1.0}_{-0.8}$	0.07

<i>DNN</i>		
$\Delta\Phi_{jj}$ bin	$\hat{\sigma}^{fid}$ [fb]	$\Delta\sigma^{bias}$ [fb]
0	$0.8^{+1.0}_{-0.9}$	0.4
1	$0.4^{+0.5}_{-0.5}$	0.2
2	$0.4^{+0.5}_{-0.5}$	0.2
3	$0.8^{+1.0}_{-0.9}$	0.5



NB: la misura è stata eseguita su un campione di dati simulati (Asimov dataset)!

Conclusioni

- Ho eseguito una misura di sezione d'urto di produzione del bosone di Higgs in funzione della differenza in angolo azimutale tra i due jet adronici dello stato finale ($\Delta\Phi_{jj}$)
- Per ottenere un risultato confrontabile direttamente con la predizione di qualunque modello teorico, ho implementato una **Rete Neurale Avversaria**
- Questo mi ha permesso di **selezionare il processo d'interesse in modo indipendente dal modello di segnale ipotizzato** e di **ridurre il «bias»** introdotto dalla procedura di estrazione del segnale nel risultato finale
- **L'incertezza della sezione d'urto misurata associata alla dipendenza dal modello è un ordine di grandezza inferiore rispetto al caso in cui si utilizzi una semplice Rete Neurale Profonda**

Sviluppi futuri

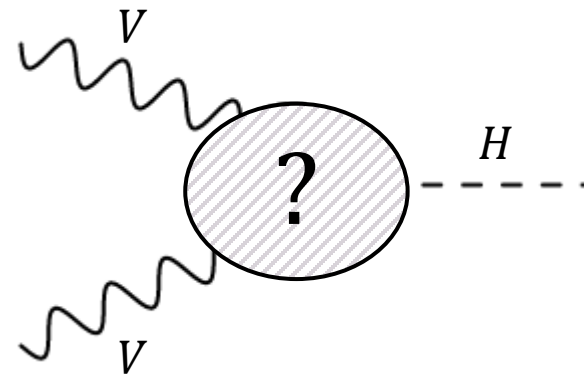
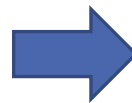
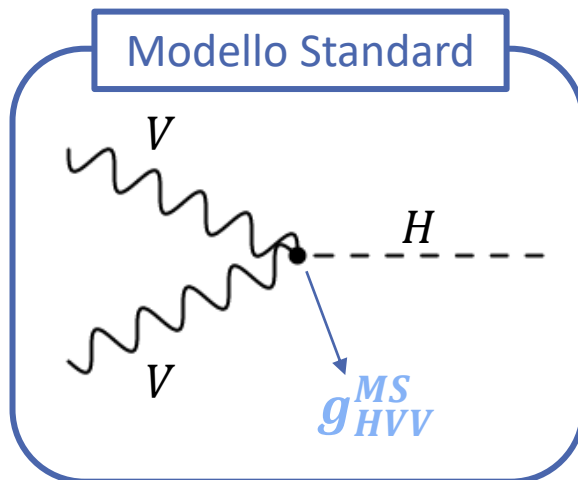
- Estendere l'analisi a tutti i dati raccolti da CMS durante il Run 2
- Utilizzare questo approccio per ridurre l'impatto delle incertezze sistematiche nelle misure future

Grazie per l'attenzione

Back up

Teorie di campo effettive

- Il Modello Standard è assunto come limite di basse energie di una teoria più completa (completamento UV) sconosciuta e valida alla scala di energia Λ
- L'impatto che questo nuovo settore fisico ha sulle osservabili misurabili a $E \ll \Lambda$ può essere parametrizzato senza specificare nessuna proprietà del modello UV



- Vertice *effettivo*, di cui si ignora la struttura interna
- Accoppiamento effettivo

$$g_{HVV}^{eff}$$

Accoppiamenti Anomali in HVV

Ampiezza di scattering tra un bosone di Higgs (spin 0) e due bosoni V (spin 1)

$$A(HVV) \sim \underbrace{\left[a_1^{VV} + \frac{k_1^{VV} q_{V1}^2 + k_2^{VV} q_{V2}^2}{(\Lambda_1^{VV})^2} \right]}_{L_1} m_{V1}^2 \epsilon_{V1}^* \epsilon_{V2}^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2)\mu\nu}$$

Accoppiamento	CP	Modello
$a_1 = 1$ e $a_2, a_3, L_1 = 0$	+	SM
$a_2 = 1, a_1 = 0$	+	H0PH
$L_1 = 1, a_1 = 0$	+	H0L1
$a_3 = 1, a_1 = 0$	-	H0M

H0PHf05	H0L1f05	H0Mf05
$a_2 = 0.5,$ $a_1 = 0.5$	$L_1 = 0.5,$ $a_1 = 0.5$	$a_3 = 0.5,$ $a_1 = 0.5$

Stima del «bias»

- \mathcal{O} corrisponde a *outputVBF*
- Ho generato un **campioni di dati distribuiti secondo la predizione del modello BSM**
- Ho generato la simulazione MC di \mathcal{O} , assumendo il MS
- Ho fittato la simulazione MC del MS ai dati BSM

- Ho ricavato $\mu^{fit} = \frac{N_{BSM}^{fit}}{N_{SM}^{exp}}$

- A parte, ho calcolato il valore che mi aspettavo $\tilde{\mu} = \frac{N_{BSM}}{N_{SM}^{exp}}$

- Ho definito il «**bias**» come:

$$b = \mu^{fit} - \tilde{\mu}$$

