“Deep” Dive in $b \to c$ Anomalies:

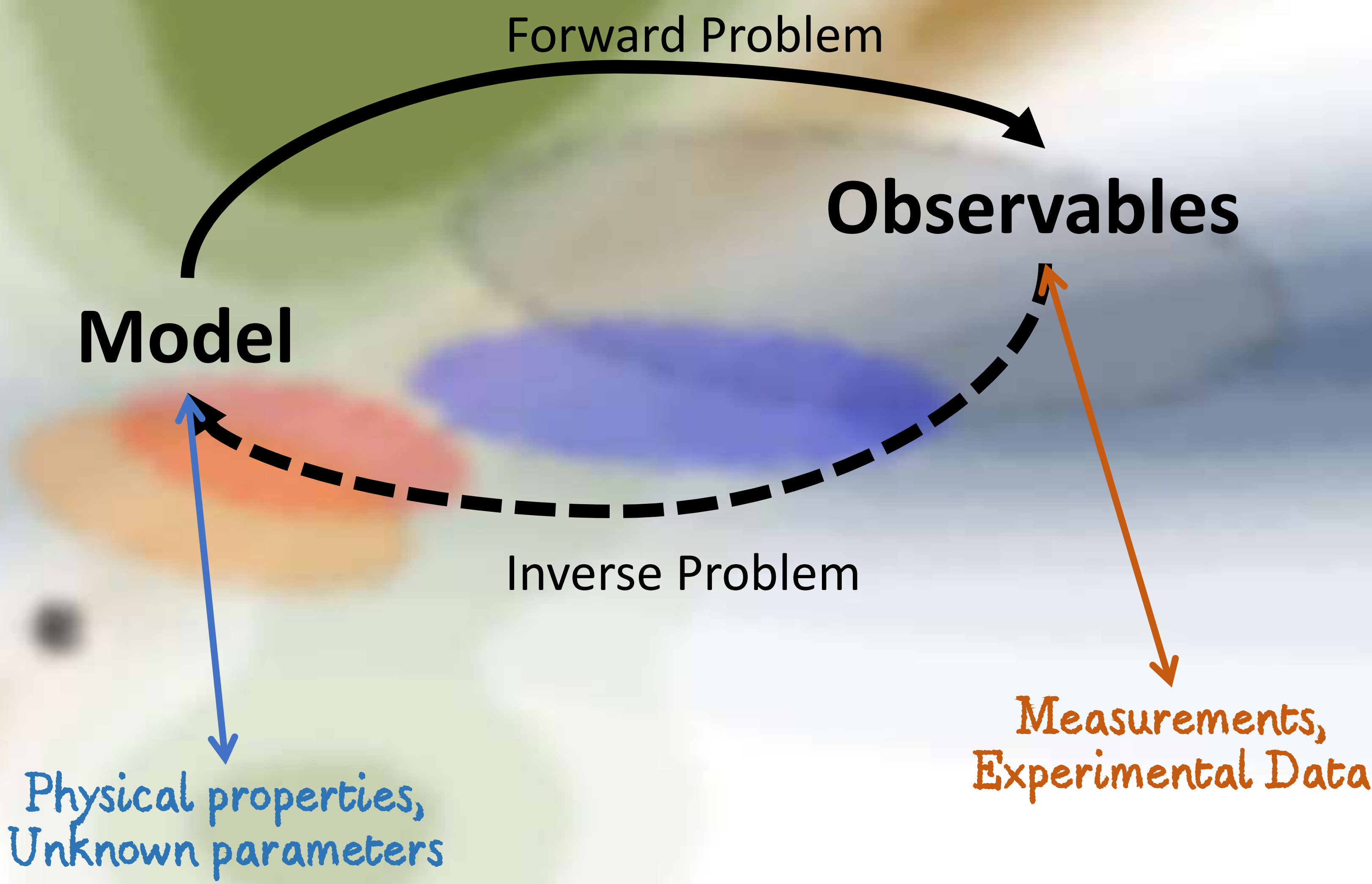
Can we Automatize Model-Selection?

Sunando K. Patra

Bangabasi Evening College, Kolkata

<u>*With:*</u> Soumitra Nandi, Shantanu Sahoo, S. Bhattacharya (*arXiv:**2008.04316***)

Inverse Problem



Inverse Problem

Forward Problem

Observables

Model 1, Model 2, …

Inverse Problem

Physical properties,
Unknown parameters

Measurements,
Experimental Data

Inverse Problem (for Dummies like me)

Observation

?

?

Models

1\.

2\.

Inverse Problem (*for Dummies like me*)

Reality

Image: Diet Wiegman

Model Selection

- *Q: Which is the best model to explain the data?*
  - *Ans: Whichever has minimum prediction error, i.e., optimum Bias and Variance.*
    - *Data → Statistical Inference → Model Selection*
- *Cross-Validation:*
  - *Powerful, reliable but computationally expensive*
  - *Draws on predictive error ⇒ can detect under and overfitting.*
- *Caveats:*
  1. *Error distribution to be known*
  2. *Small data-set → becomes unstable* [Beleites et al, **ACA 760, 25** (2013), G. Varoquaux, (2017), arXiv:**1706.07581** [q-bio.QM].]
  3. *Small data-set + Comparable model size ⇒ No chance of doing anything!*

Alternatives?

- **What to do for $b \rightarrow c\tau\nu$ then?**
  - 5 C_W's (Wilson Coefficients) → $C_{V_1}, C_{V_2}, C_{S_1}, C_{S_2}, C_T$
  - Each one complex → total 10 parameters
  - 4 Observables: 9 Data-points (2018)
- We had used AIC_C arXiv:1805.08222[EPJC 79, 268 (2019)]

Aritra is going to talk about these methods and their shortcomings in more detail… LOOCV, AIC_c etc.

Motivations

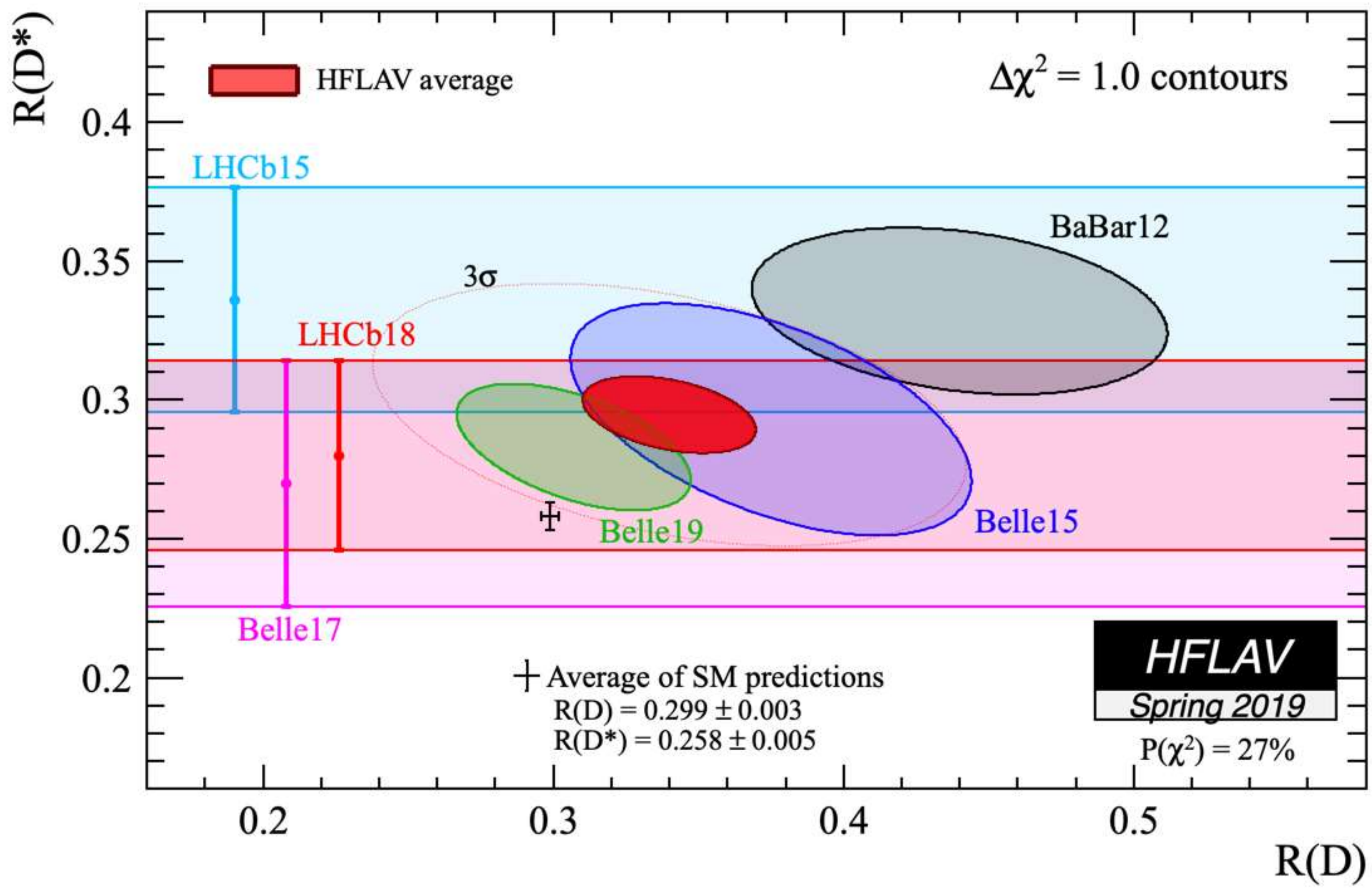
- *No criterion tells the whole truth.*
- *All unstable for small sample sizes*
- *Depends on data at hand, whereas model predictions are available → have to redo everything for each change in data*
- *Depends on the MLE*
  ⇒ no info. about span (uncertainty) of the data-dist.
  ⇒ Proclivity to select simpler models
- **Bayesian Model Selection?**

(Ad Hoc) Bayesian Model Selection

- *No 'true' model exists → data-distribution is 'true'*
- *Information lost to approximate a posterior P with a prior Q*
  - *This loss can be quantified by some '**divergence**' between P and Q*
- *Popular choice:* ***Kullback–Leibler divergence*** $(\boldsymbol{D_{KL}})$
  - *Related to MLE, AIC, and cross-validation: more details in paper…*

1. *Find Bayesian parameter-space for all models*
2. *Find the predicted distributions of observables*
3. *Calculate D_{KL} for them*
4. *Lowest D_{KL} → Best Model*

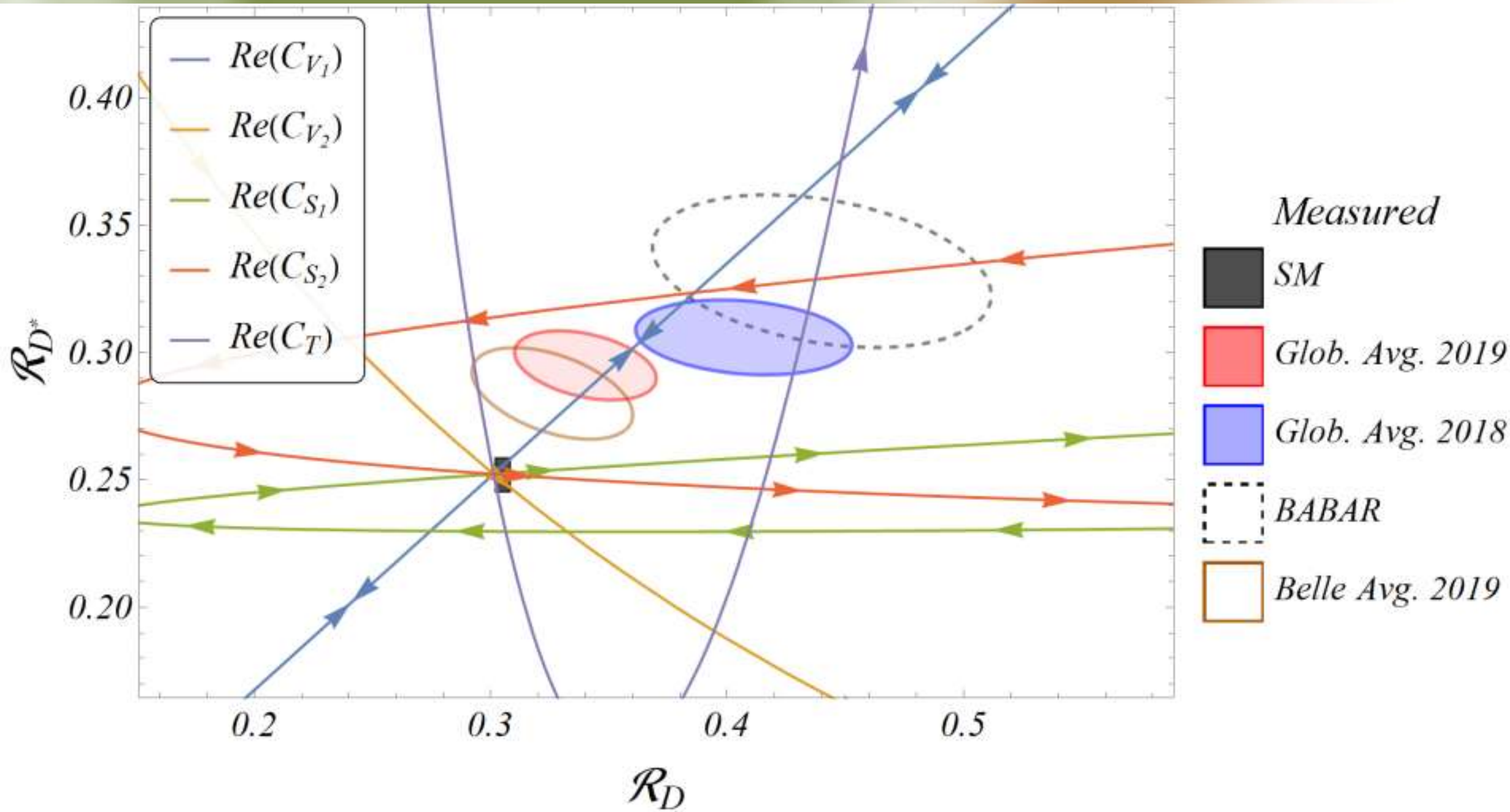
$R(D) - R(D^*)$ Status *(2019)*



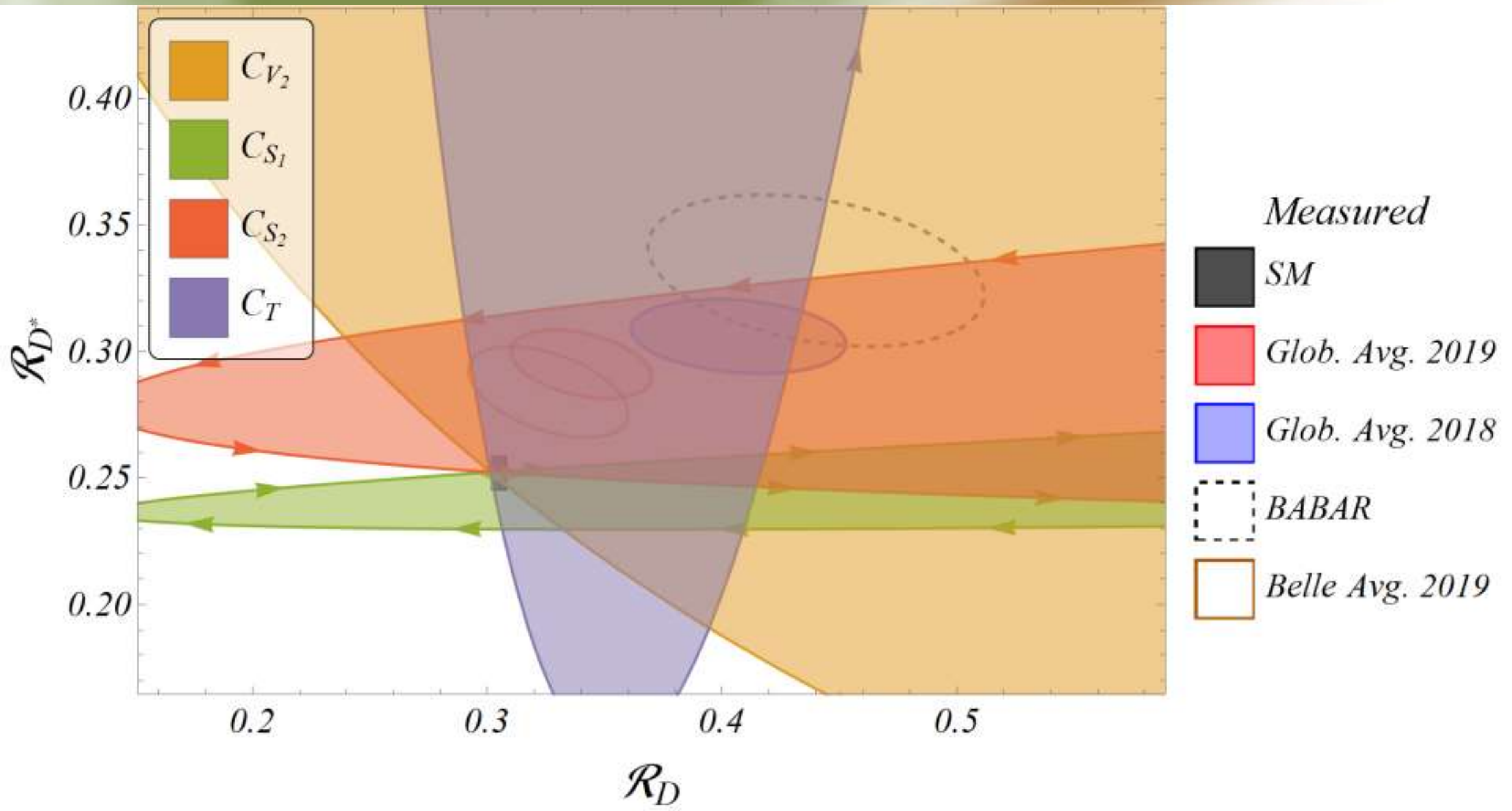
$\boldsymbol{R(D) - R(D^*)}$ Status *(2019)*

Observables		Measurement
$\mathcal{R}_D$ [6]	4-obs. data-set	$0.340(27)(13)$
$\mathcal{R}_{D^*}$ [6]	4-obs. data-set	$0.295(11)(8)$
$P_\tau(D^*)$ [40]	4-obs. data-set	$-0.38(51)(\,^{+0.21}_{-0.16})$
$F_L^{D^*}$ [35]	4-obs. data-set; *New Lattice: arXiv:2007.06956*	$0.60(8)(4)$
$\mathcal{R}_{J/\psi}$ [12]	SM: 0.2601 $\pm$0.0036	$0.71(17)(18)$

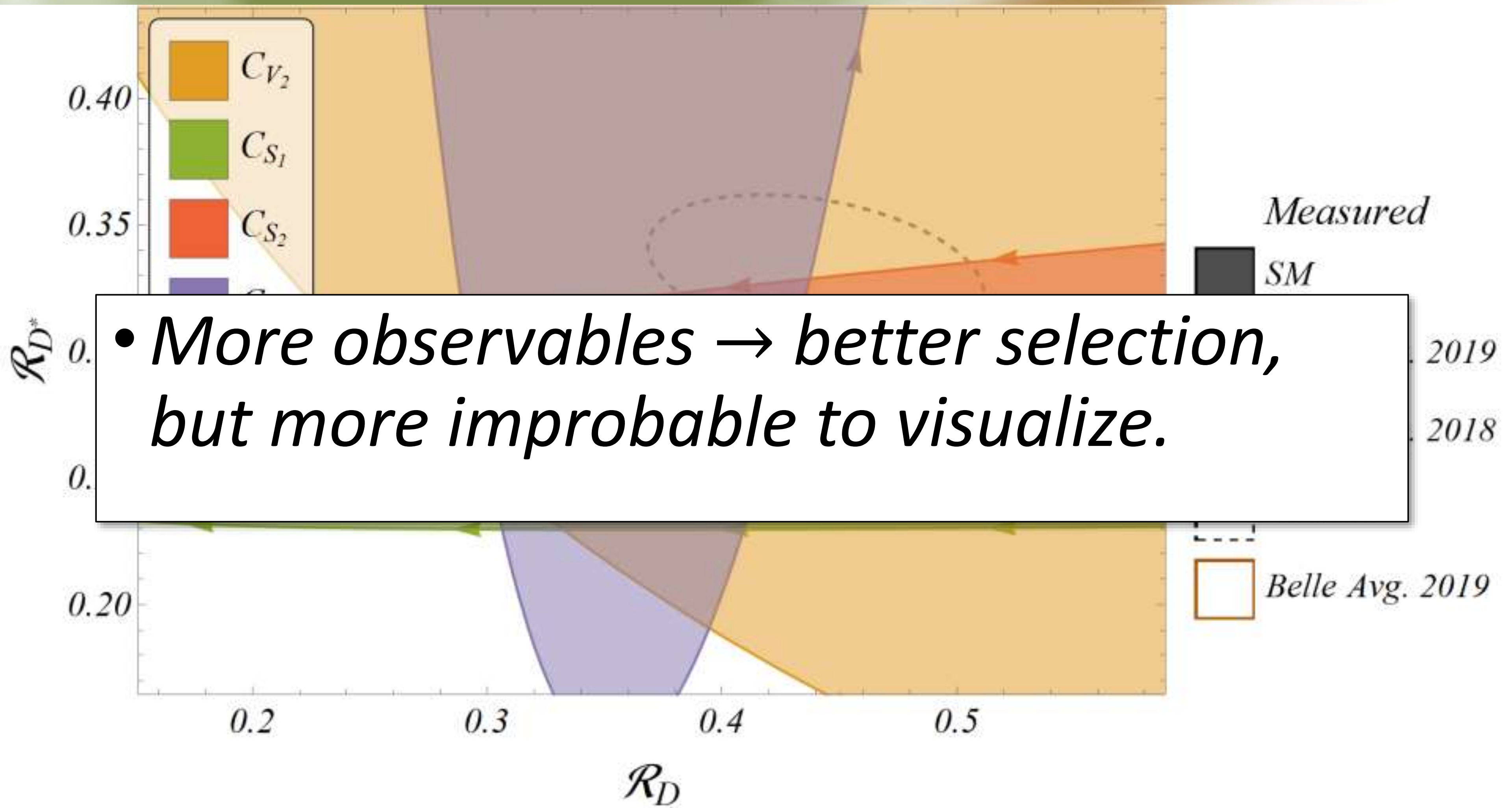
Complexity of Inverse Problem



Complexity of Inverse Problem

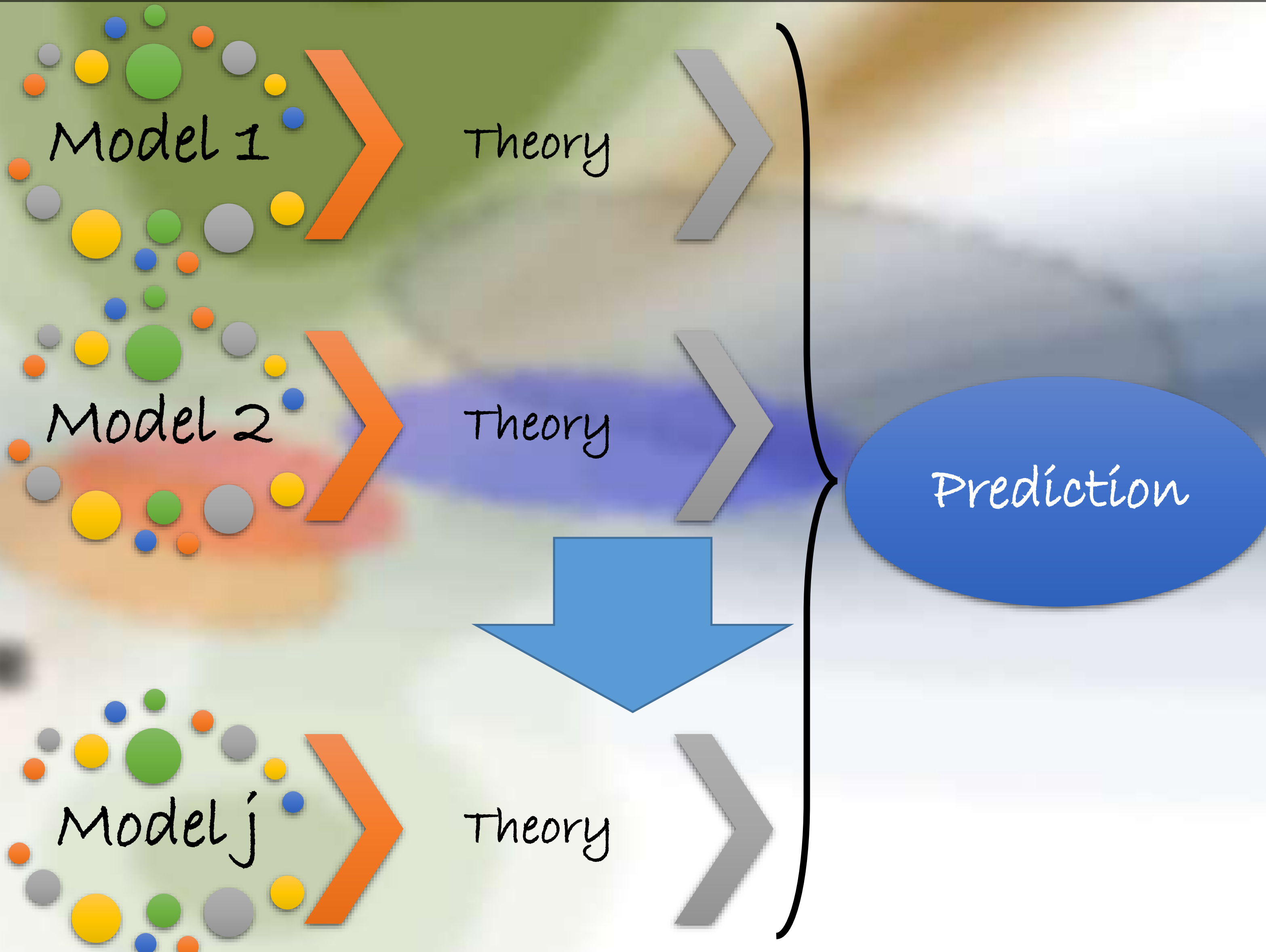


Complexity of Inverse Problem

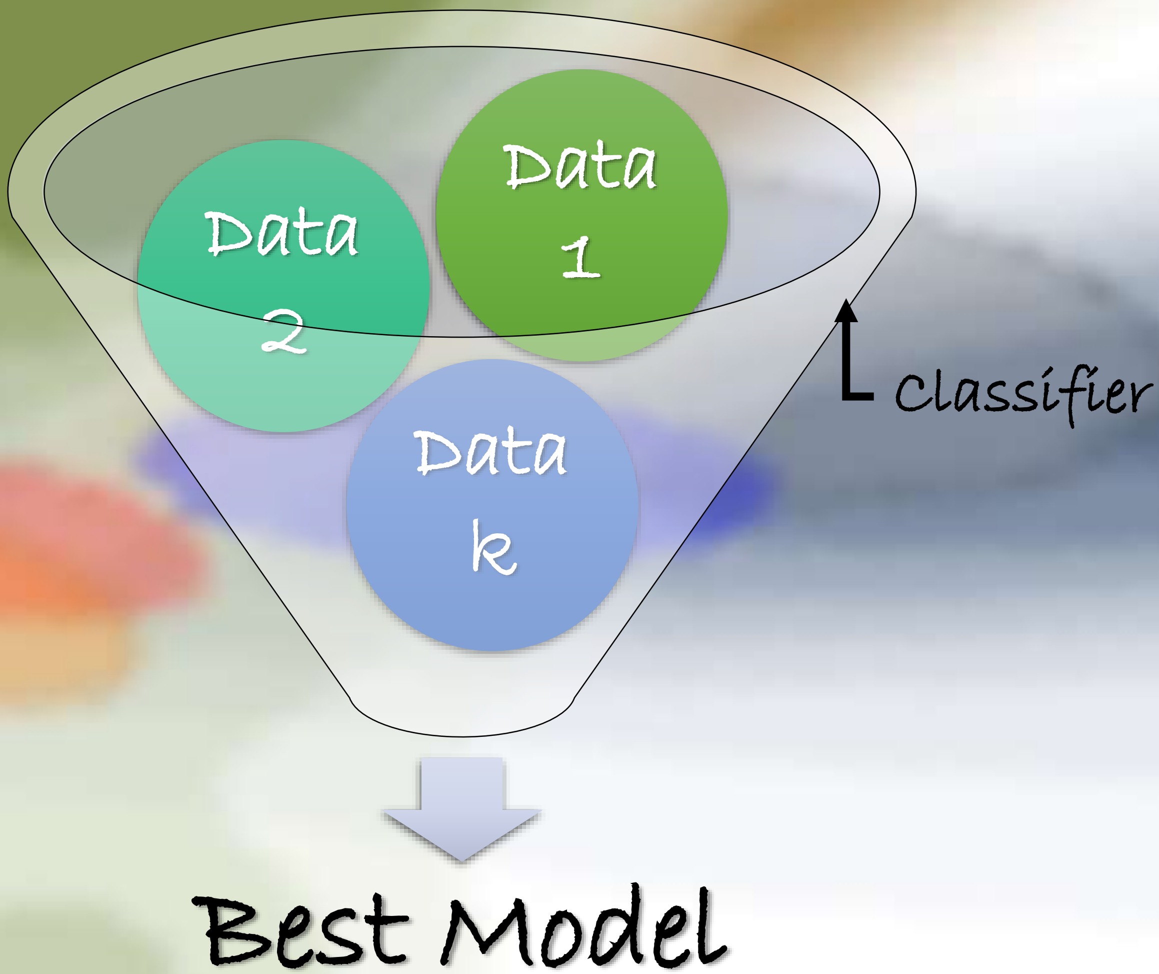


- *More observables → better selection, but more improbable to visualize.*

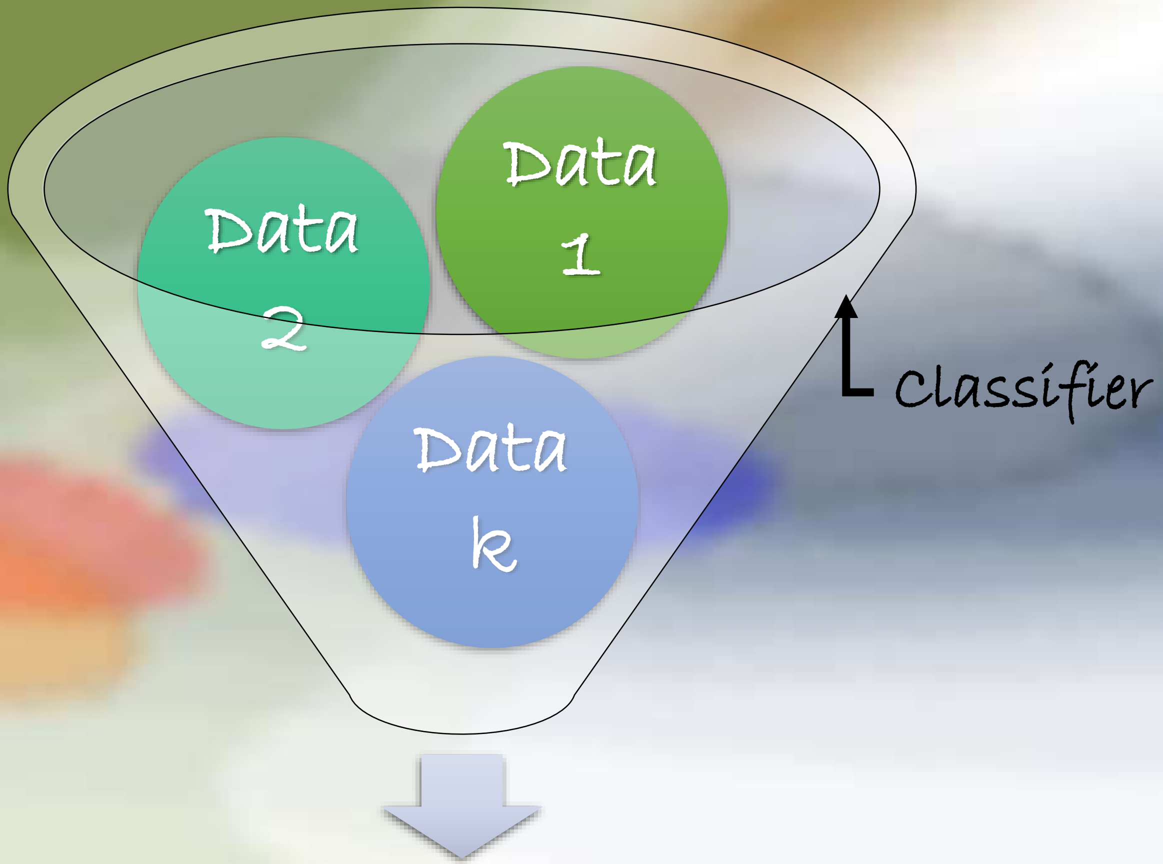
Proposal



Probable Inverse Functions

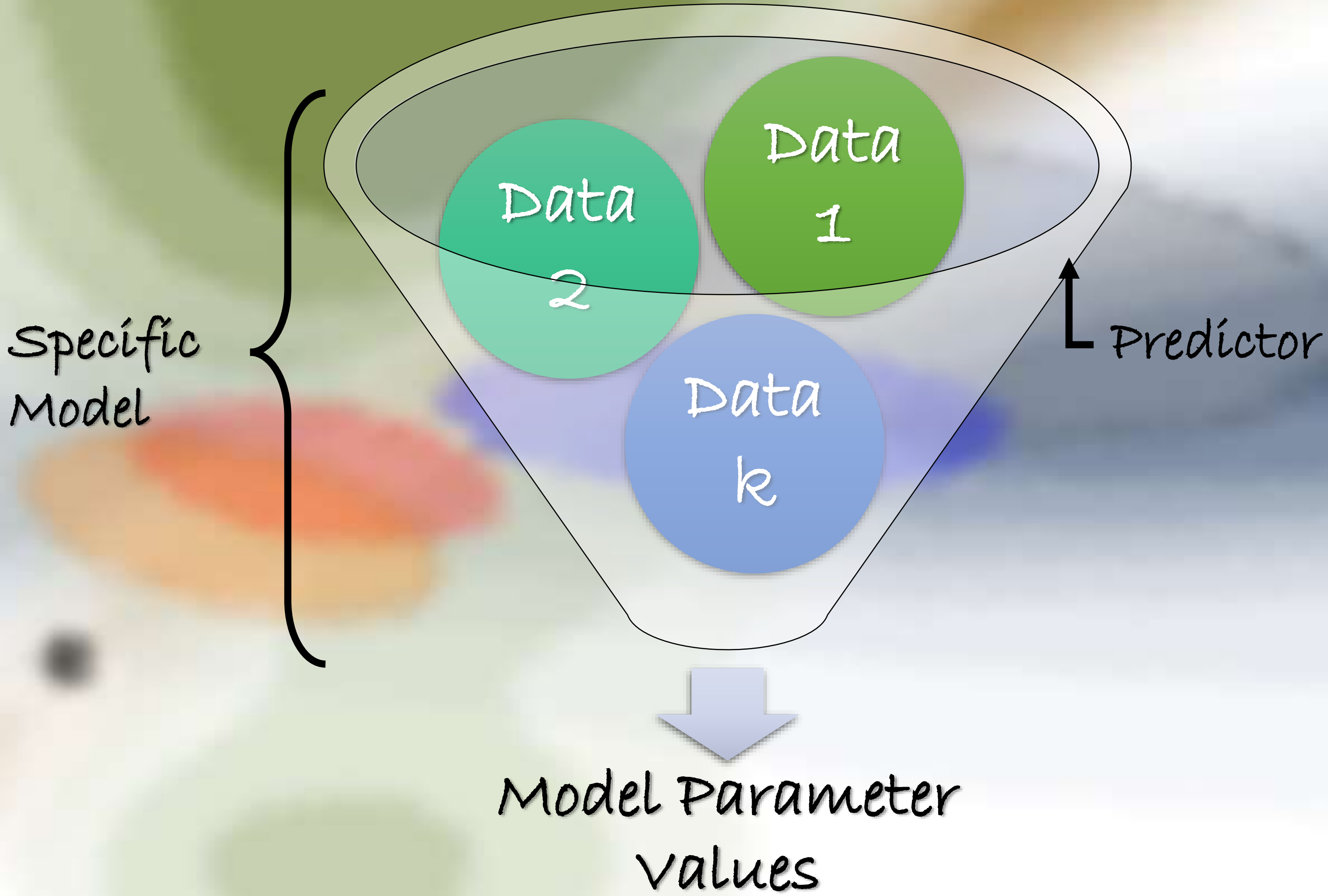


Probable Inverse Functions



Model p (% prob) > Model q (% prob) > Model r (% prob),
…

Probable Inverse Functions



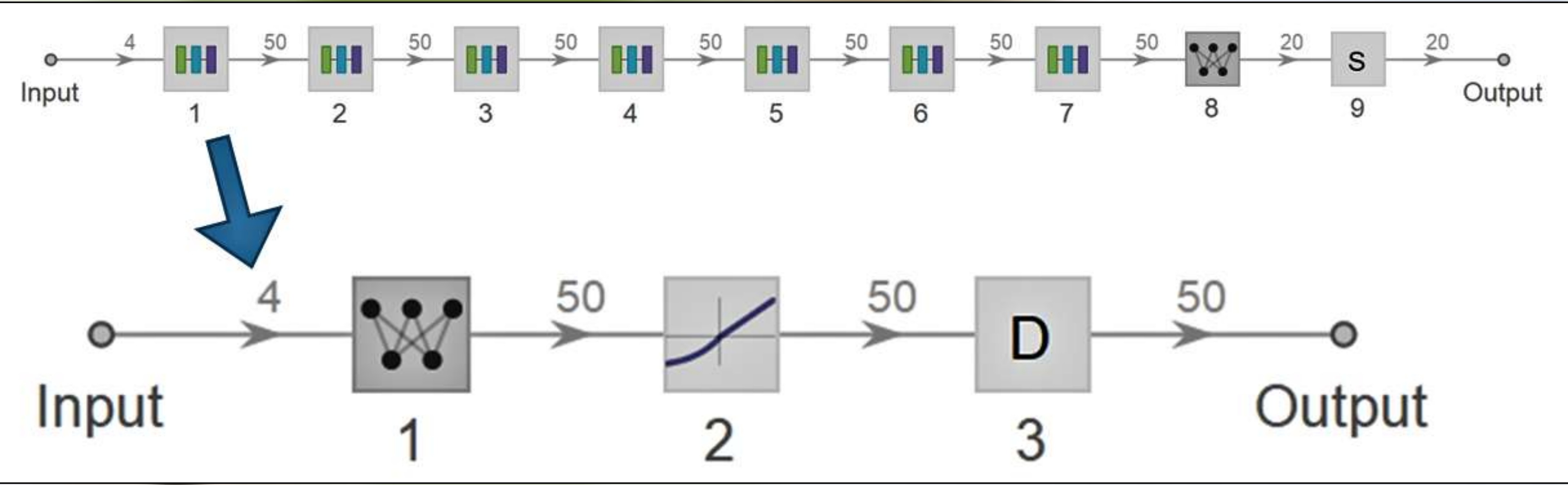
Neural Network: SNN

- *Self-Normalizing Neural-networks: proposed in 2017 [Klambauer et al, arXiv:**1706.02515**]*
- *Solves gradient-vanishing/exploding problem of traditional Fully-Connected-Networks*
- *Enables 'deep' networks*
  ⇒ Performance $\geq$ shallow algorithms
- *Special activation function: **SELU***
- *Variation remains normalized over the whole network.*
- *(Modified) Dropout layers possible*
  *⇒ Enables Regularization, *Bayesian Networks**

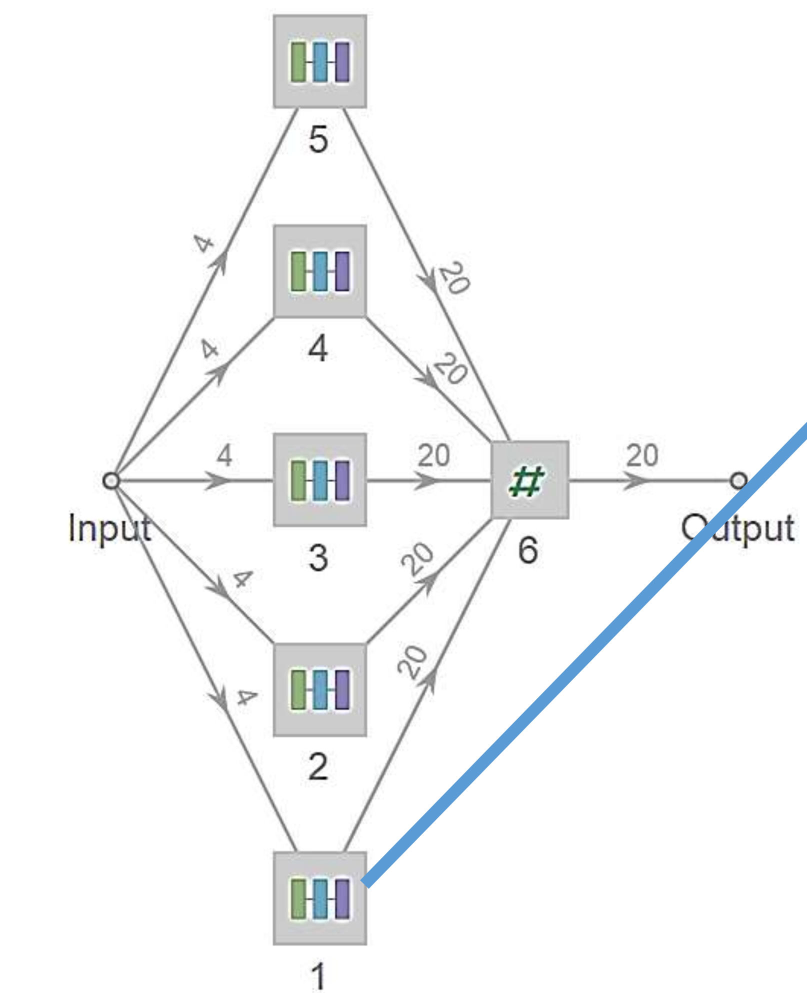
Neural Network: SNN

Classifier	Accuracy
ensemble-SNN	**89.08%**
Random Forest	82.70%
Decision Tree	78.27%
Nearest Neighbors	74.22%
Gradient Boosted Trees	74.10%
Support Vector Machine	63.82%
Markov Process	50.76%
Naive Bayes	46.64%
Logistic Regression	29.80%

SNN → Ensemble SNN



SNN → *Ensemble SNN*



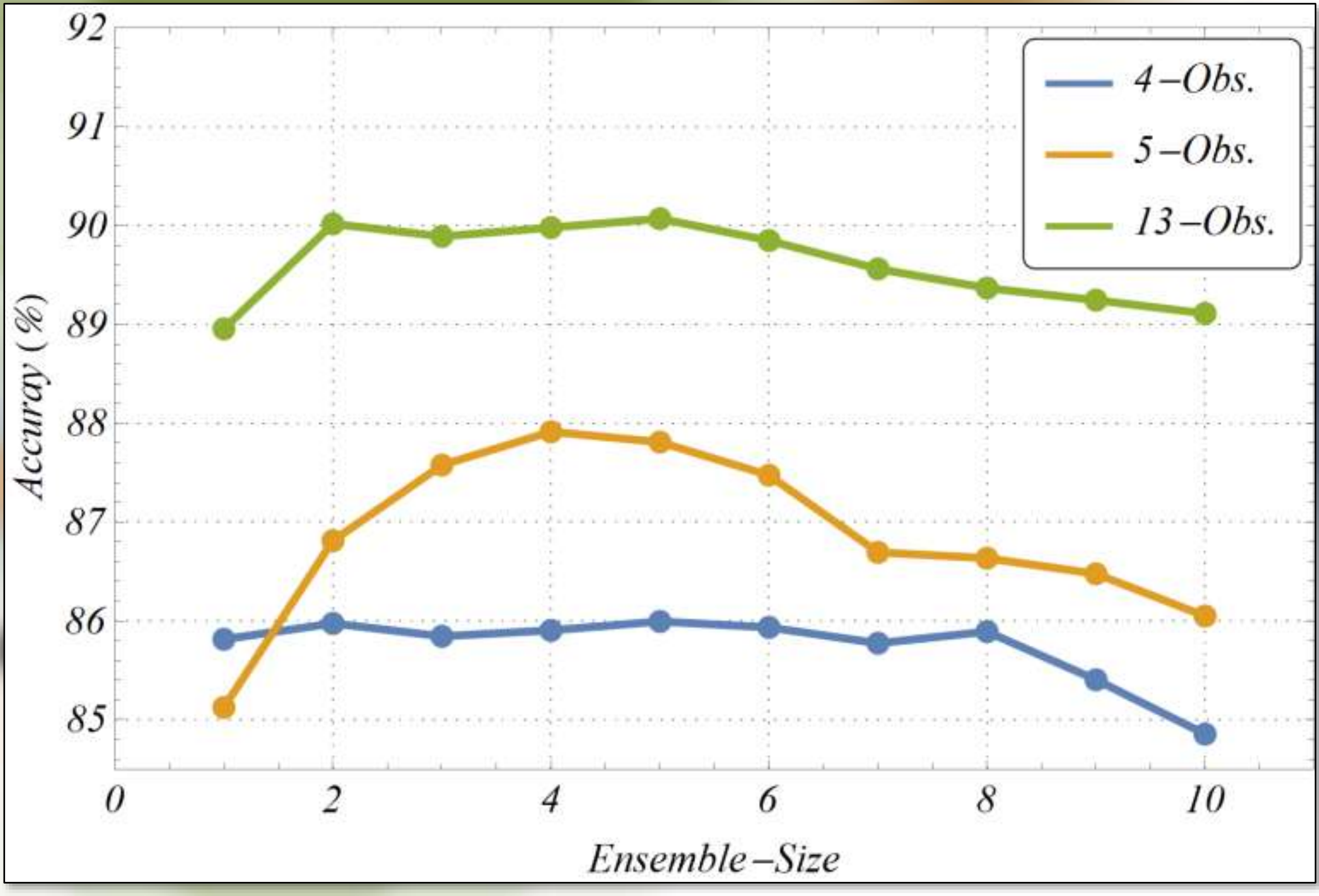
Inputs Input: vector (size: 4)

Outputs Output: class

1:	**NetChain**	
	Input	vector (size: 4)
1	NetChain (3 nodes)	vector (size: 50)
2	NetChain (3 nodes)	vector (size: 50)
3	NetChain (3 nodes)	vector (size: 50)
4	NetChain (3 nodes)	vector (size: 50)
5	NetChain (3 nodes)	vector (size: 50)
6	NetChain (3 nodes)	vector (size: 50)
7	NetChain (3 nodes)	vector (size: 50)
8	LinearLayer	vector (size: 20)
9	SoftmaxLayer	vector (size: 20)
	Output	vector (size: 20)

1:	**NetChain**	
	Input	vector (size: 4)
1	LinearLayer	vector (size: 50)
2	ScaledExponentialLinearUnit[x]	vector (size: 50)
3	DropoutLayer	vector (size: 50)
	Output	vector (size: 50)

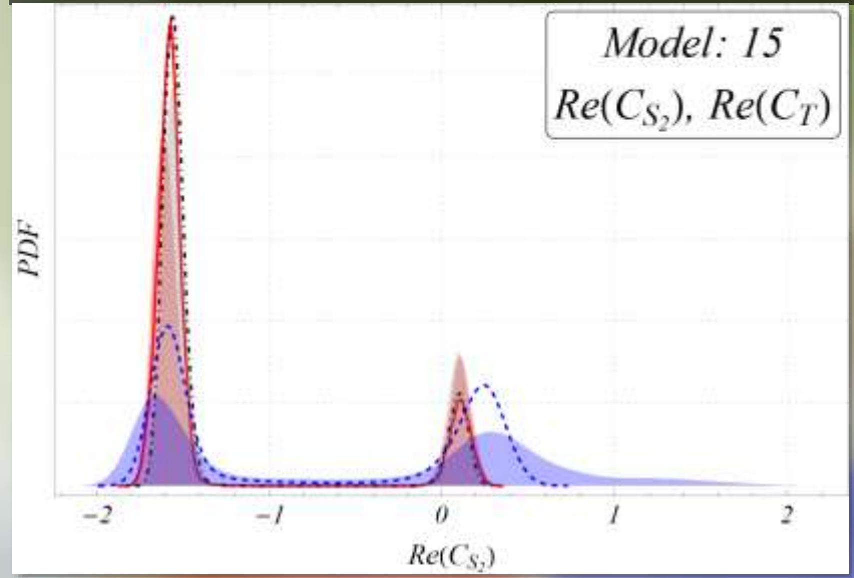
SNN → Ensemble SNN

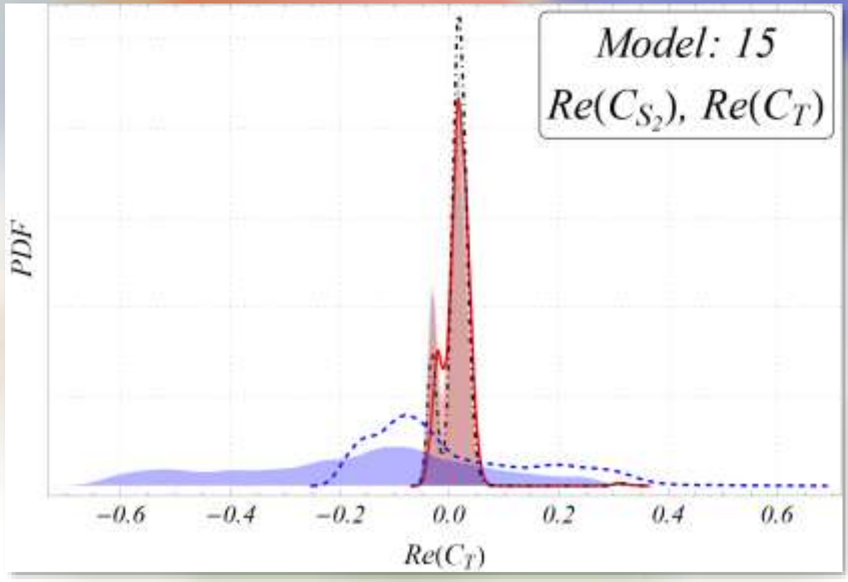


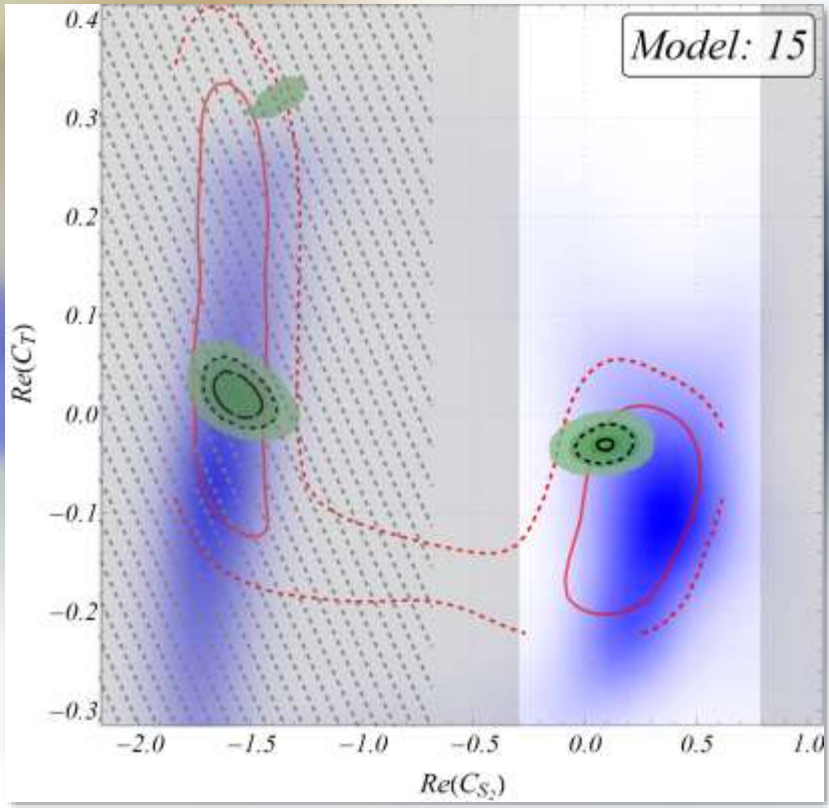
Classification Results

Data-Set	Models (SNN-Aggregate)	Parameters	Aggregate Prob. (%)	D_{KL} Serial	SNN-Central Serial	ΔAIC_c Serial	$w^{\Delta\text{AIC}_c}$ (%)
4-Obs.	**12**	$Re(C_{V_2}),\ Re(C_T)$	**38.48**	**2**	**1**	18	0.07
	15	$Re(C_{S_2}),\ Re(C_T)$	**26.30**	**3**	**5**	**6**	**0.32**
	13	$Re(C_{S_1}),\ Re(C_{S_2})$	11.53	9	7	9	0.29
	14	$Re(C_{S_1}),\ Re(C_T)$	8.04	**1**	15	17	0.11
	8	$Re(C_{V_1}),\ Re(C_{S_2})$	5.01	18	**3**	**7**	**0.32**
	10	$Re(C_{V_2}),\ Re(C_{S_1})$	2.80	**6**	**2**	11	0.16
	6	$Re(C_{V_1}),\ Re(C_{V_2})$	2.19	14	9	12	0.12
	11	$Re(C_{V_2}),\ Re(C_{S_2})$	1.74	**4**	**4**	**8**	**0.32**
	19	C_{S_2}	1.18	17	6	10	0.29
5-Obs.	**12**	$Re(C_{V_2}),\ Re(C_T)$	**42.62**	**1**	**2**	17	0.69
	13	$Re(C_{S_1}),\ Re(C_{S_2})$	15.71	19	**6**	**7**	**3.19**
	15	$Re(C_{S_2}),\ Re(C_T)$	8.56	**4**	10	**4**	**3.37**
	6	$Re(C_{V_1}),\ Re(C_{V_2})$	6.7	16	12	11	1.33
	8	$Re(C_{V_1}),\ Re(C_{S_2})$	6.54	18	**4**	**5**	**3.31**
	14	$Re(C_{S_1}),\ Re(C_T)$	6.09	**2**	7	15	1.14
	7	$Re(C_{V_1}),\ Re(C_{S_1})$	4.63	8	11	12	1.29
	11	$Re(C_{V_2}),\ Re(C_{S_2})$	3.7	**3**	**3**	**6**	**3.3**
	10	$Re(C_{V_2}),\ Re(C_{S_1})$	2.66	**6**	**1**	9	1.67
	17	C_{V_2}	1.45	15	9	10	1.33

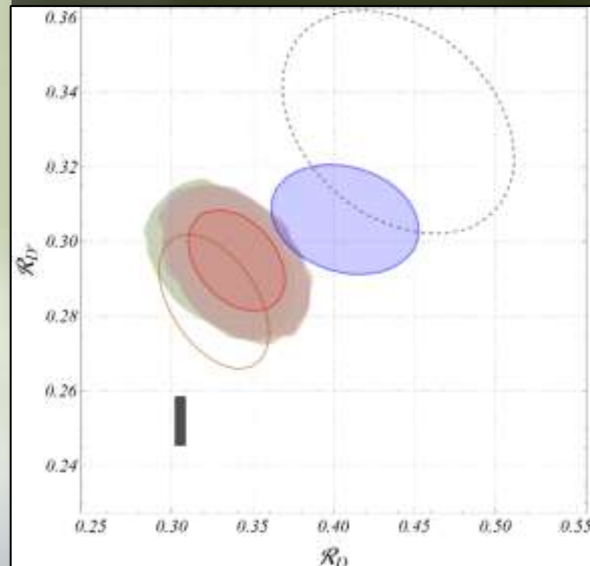
Predictions: Test Model - 15

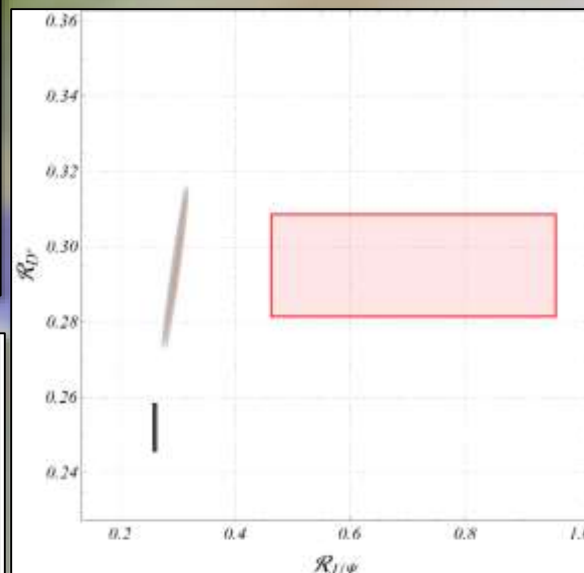


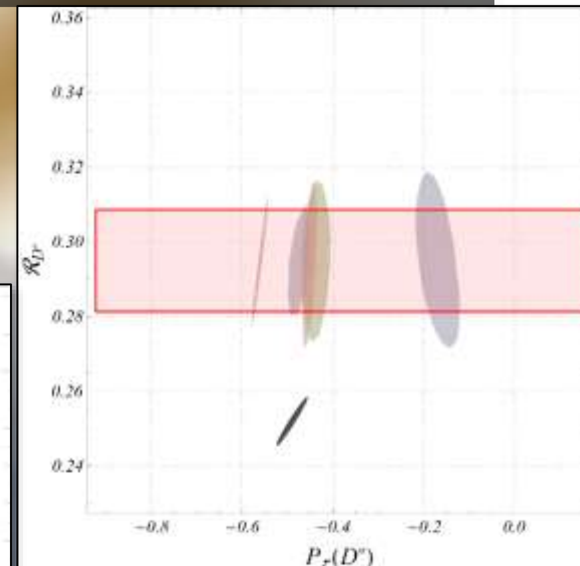


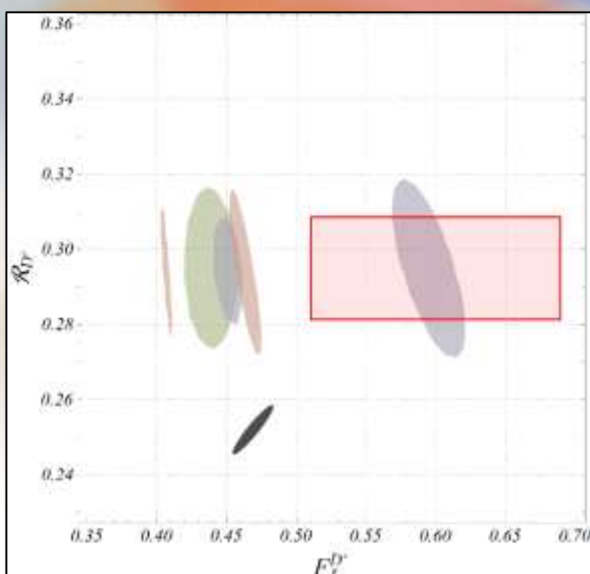


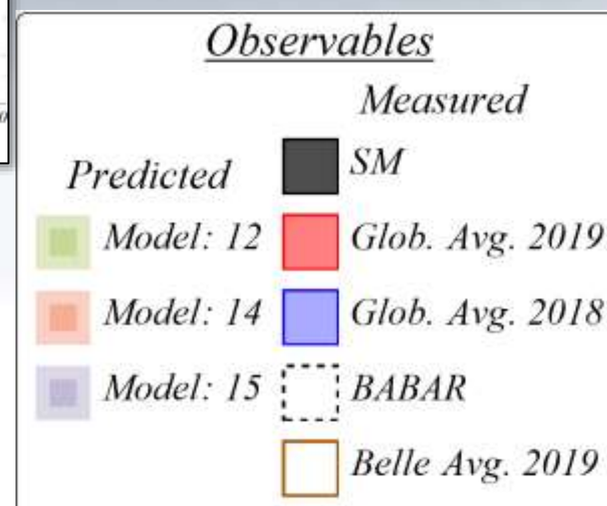
Observables:











Predictions + Classification

Dataset	Model Index	D_{KL}	
		(Net)	(Bayes)
4-Obs.	15	3.75	31.78
	14	10.43	1.15×10^5
	10	14.53	1.21×10^6
	12	18.20	6.53×10^4
	11	29.67	5.49×10^5
5-Obs.	12	95.60	6.29×10^6
	15	1146.50	1.92×10^6
	14	1631.02	9.60×10^7
	10	3232.05	5.20×10^6
	11	3905.8	2.41×10^6

Predictions + Classification

Dataset	Model Index	Net Performance (%)		
		σ	R^2	MSE
4-Obs.	15	1.80	99.93	0.03
	14	2.03	99.91	0.04
	10	3.08	99.84	0.10
	12	1.30	99.93	0.02
	11	2.34	99.91	0.05
5-Obs.	12	1.58	99.90	0.02
	15	2.19	99.89	0.05
	14	2.33	99.88	0.05
	10	3.28	99.82	0.11
	11	2.97	99.86	0.09

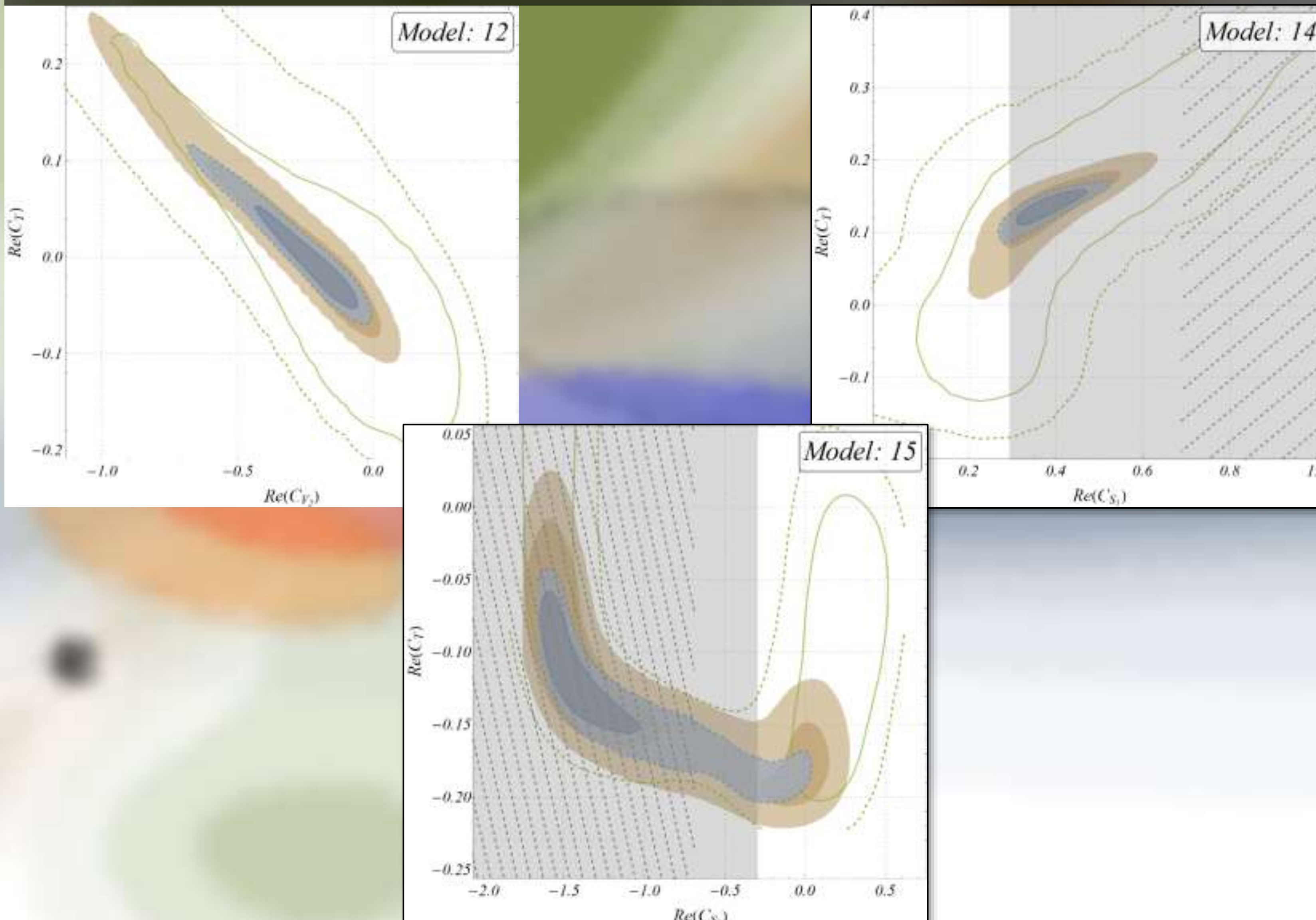
Future: This mode

- *There are other, unmeasured observables:*
  1. $B \to D^{(*)}\tau\nu$: $\mathcal{A}_{FB}^{(*)}$, $P_\tau(D)$
  2. $\Lambda_b \to \Lambda_c\tau\nu_\tau$: $\mathcal{R}_\Lambda^\ell$, $\mathcal{A}_{FB}^\Lambda$
  3. $B_c \to J/\Psi\tau\nu_\tau$: $\mathcal{A}_{FB}^{J/\Psi}$, $P_\tau(J/\Psi)$, $F_L(J/\Psi)$
  4. *And updates on all present ones…*

13 in Total

- *We trained another full set of classifier and predictor SNNs for all 'models'… because we can!*
  - *Their tests will only happen when data appears…*
- *In the mean time, let's see what happens to the predictions with future luminosities:*
  (central values kept same, uncertainties decreased)

Future: This mode



Future: Other observables?

- *Global* $\boldsymbol{b \to s\,\ell\ell}$ *?*
  - $9\ C_W$*'s (Wilson Coefficients)* $\to$
    $$C_7', \Delta C_9, C_9', \Delta C_{10}, C_{10}', C_S, C_S', C_P, C_P'$$
  - *If all are Complex* $\to$ *1022 combinations*
  - > 200 *obs.* $\Rightarrow$ *Cross Validation, Bayesian Model Selection possible*
  - *Enough data for full classification and precise prediction: Preliminary tests: > 99% accuracy*
  - *Varying network structure, calculating prediction error – much more important.*
- *Electroweak Data + Higgs Decays* $\to$ *SMEFT model independent predictions?*
- *Why stop there? Compare classes of BSM models…*

For Now:

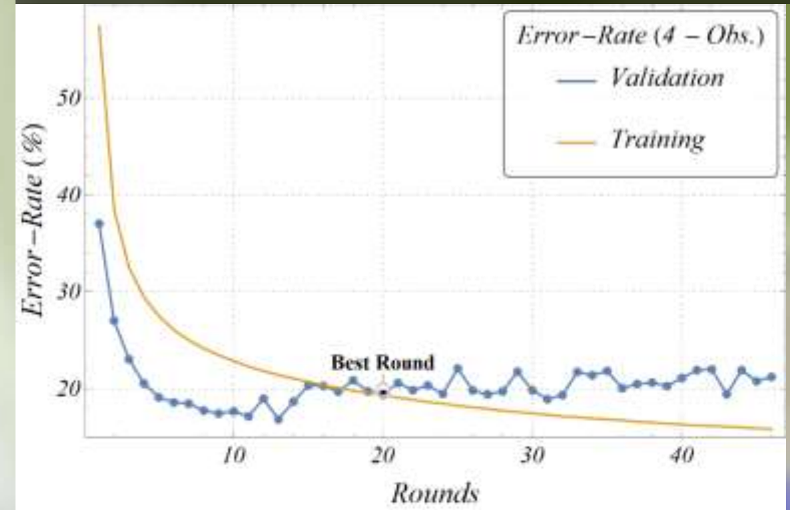
- *Play around with our trained networks*
  $\rightarrow$ classifiers, predictors and Bayesian Results…

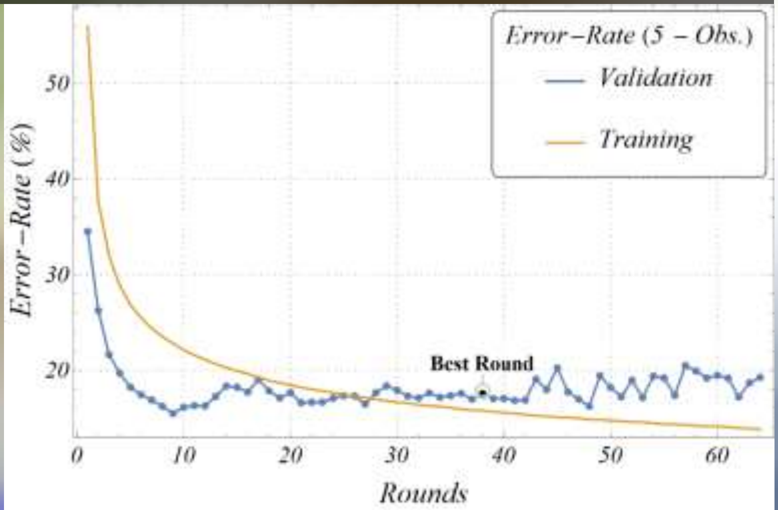
  https://github.com/FlavorIITG/MLResourcesForSemileptonic-b2c

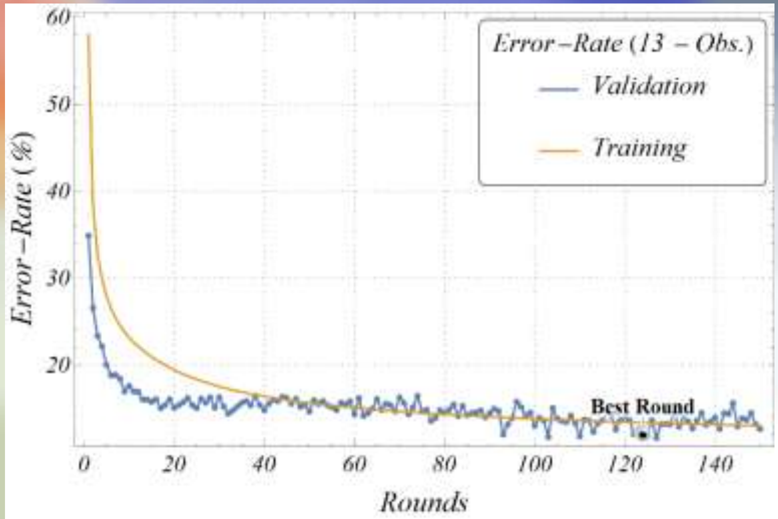
- *An interactive web applet will come soon …*

THANK YOU!

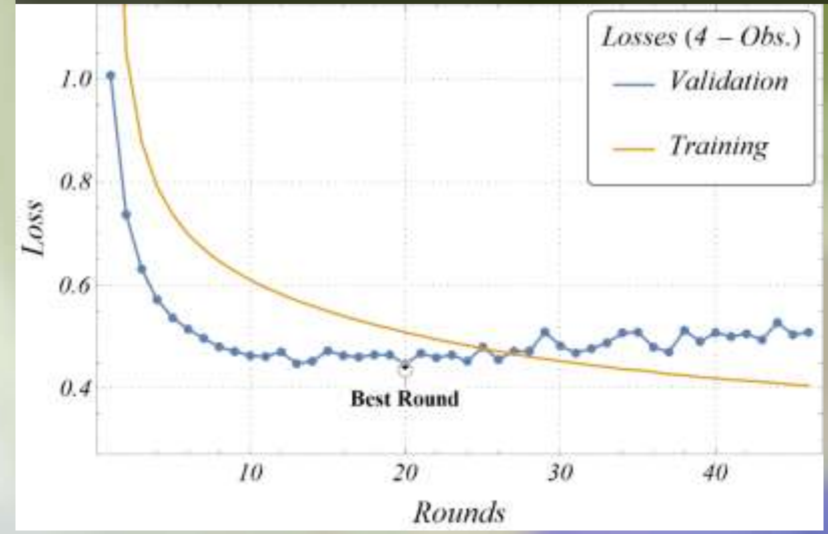
Back-Up: Error-Rate Plots

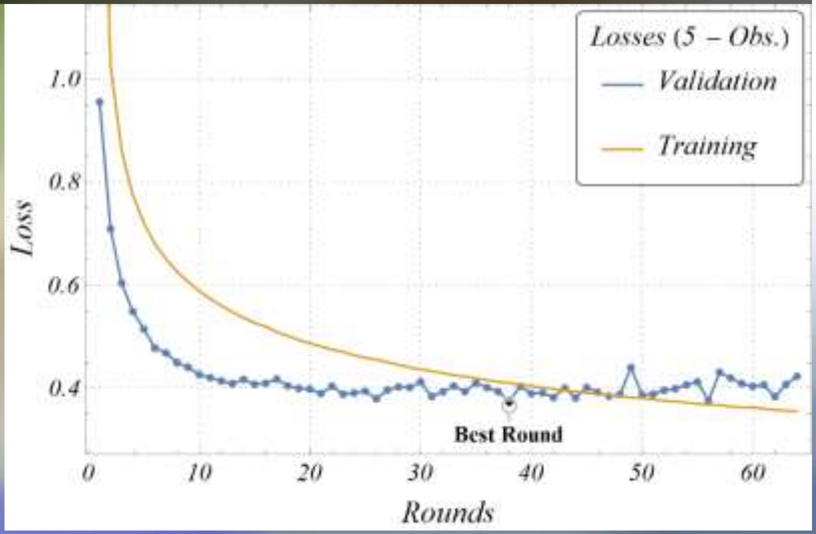


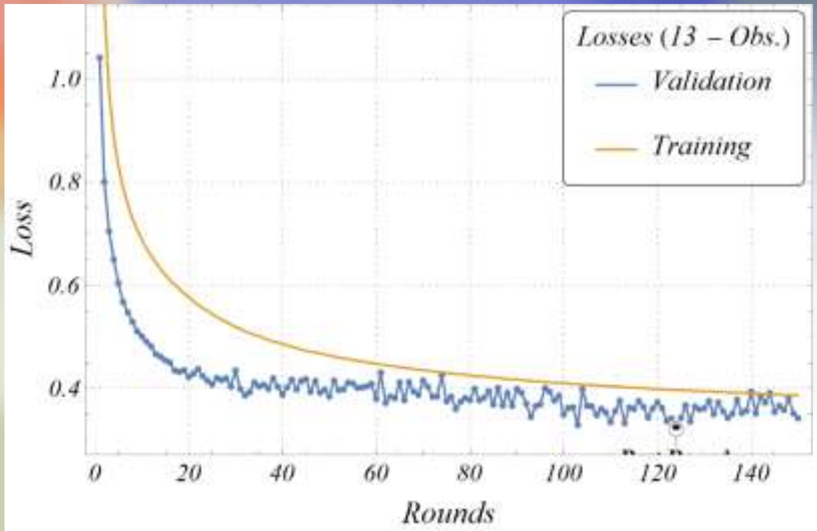




Back-Up: Loss-Rate Plots







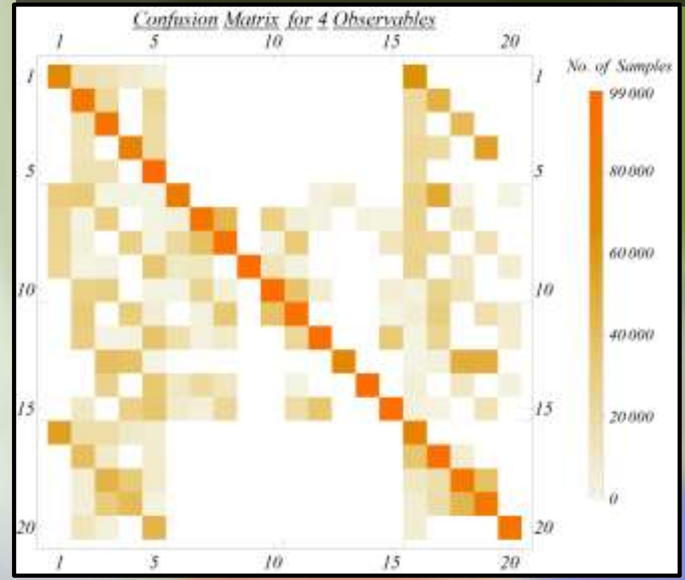
Back-Up: Other Properties

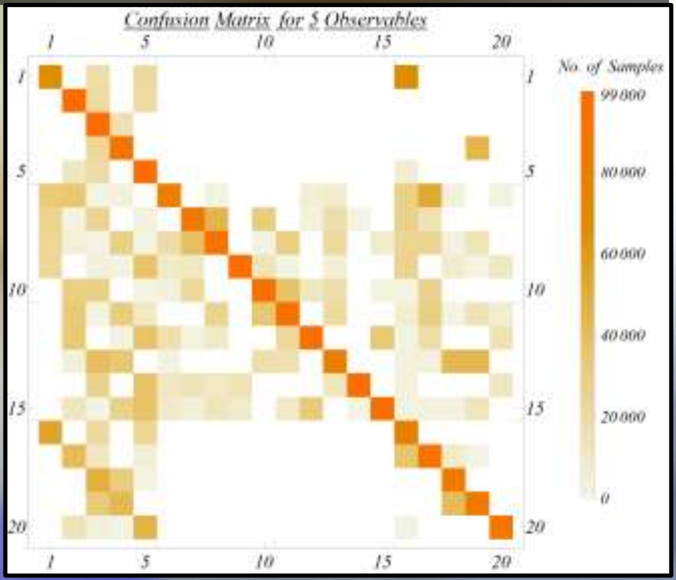
Training Info.	Details
Batches/Round	170
Batch Size	10000
Best Valid. Round	20
Final ℓ	0.0009998
Initial ℓ	0.001
Avg. Batches/Second	81.65
Final Round Loss	0.406
Final Round Error	14.48%
Total Batches	7820
Total Rounds	46
Validation Loss	0.51
Validation Error-Rate	21%

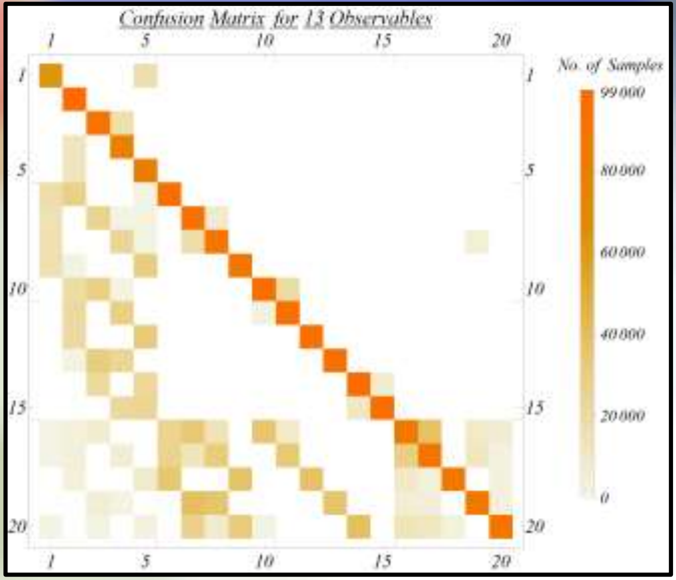
Back-Up: Other Properties

Measure	4 Obs.	5 Obs.	13 Obs.
Accuracy	85.88%	88.67%	89.68%
Cohen's κ	85.14%	88.08%	89.42%
Error	14.12%	11.33%	10.32%
Geometric Mean Prob.	69.54%	71.97%	71.75%
Mean Cross-Entropy	0.363	0.329	0.332
Mean Decision Utility	85.88%	88.67%	89.68%
Perplexity	1.44	1.39	1.39
Scott's π	85.13%	88.07%	89.42%

Back-Up: Confusion Matrices







Back-Up: $\boldsymbol{b \to c}$

$$\mathcal{H}_{eff} = \frac{4G_F}{\sqrt{2}} V_{cb} \Big[(\delta_{\ell\tau} + C^{\ell}_{V_1}) \mathcal{O}^{\ell}_{V_1} + C^{\ell}_{V_2} \mathcal{O}^{\ell}_{V_2} + C^{\ell}_{S_1} \mathcal{O}^{\ell}_{S_1} + C^{\ell}_{S_2} \mathcal{O}^{\ell}_{S_2} + C^{\ell}_{T} \mathcal{O}^{\ell}_{T} \Big],$$

$$\begin{aligned}
\mathcal{O}^{\ell}_{V_1} &= (\bar{c}_L \gamma^{\mu} b_L)(\bar{\tau}_L \gamma_{\mu} \nu_{\ell L}), \\
\mathcal{O}^{\ell}_{V_2} &= (\bar{c}_R \gamma^{\mu} b_R)(\bar{\tau}_L \gamma_{\mu} \nu_{\ell L}), \\
\mathcal{O}^{\ell}_{S_1} &= (\bar{c}_L b_R)(\bar{\tau}_R \nu_{\ell L}), \\
\mathcal{O}^{\ell}_{S_2} &= (\bar{c}_R b_L)(\bar{\tau}_R \nu_{\ell L}), \\
\mathcal{O}^{\ell}_{T} &= (\bar{c}_R \sigma^{\mu\nu} b_L)(\bar{\tau}_R \sigma_{\mu\nu} \nu_{\ell L}),
\end{aligned}$$

Wilson coefficients $\to C^{\ell}_{W}$

Back-Up: Models

Index	Parameters	Index	Parameters
1(1)	$\mathrm{Re}(C_{V_1})$	16	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$
2(2)	$\mathrm{Re}(C_{V_2})$	17	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_2})$
3(3)	$\mathrm{Re}(C_{S_1})$	18	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_T)$
4(4)	$\mathrm{Re}(C_{S_2})$	19	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$
5(5)	$\mathrm{Re}(C_T)$	20	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_T)$
6(6)	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$	21	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$
7(7)	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_1})$	22	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$
8(8)	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_2})$	23	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_T)$
9(9)	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_T)$	24	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$
10(10)	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$	25	$\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$
11(11)	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_2})$	26	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$
12(12)	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_T)$	27	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_T)$
13(13)	$\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$	28	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$
14(14)	$\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_T)$	29	$\mathrm{Re}(C_{V_1})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$
15(15)	$\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$	30	$\mathrm{Re}(C_{V_2})$, $\mathrm{Re}(C_{S_1})$, $\mathrm{Re}(C_{S_2})$, $\mathrm{Re}(C_T)$

Back-Up: *Models*

Index	Parameters
31(16)	C_{V_1}
32(17)	C_{V_2}
33(18)	C_{S_1}
34(19)	C_{S_2}
35(20)	C_T
36	$C_{V_1},\ C_{V_2}$
37	$C_{V_1},\ C_{S_1}$
38	$C_{V_1},\ C_{S_2}$
39	$C_{V_1},\ C_T$
40	$C_{V_2},\ C_{S_1}$
41	$C_{V_2},\ C_{S_2}$
42	$C_{V_2},\ C_T$
43	$C_{S_1},\ C_{S_2}$
44	$C_{S_1},\ C_T$
45	$C_{S_2},\ C_T$