Reliable Simulation-based Inference in the Physical Sciences

Particle physics colloquium, KIT

January 12, 2023

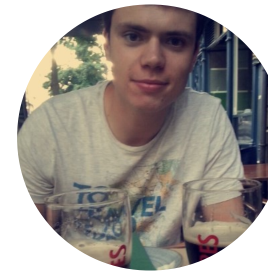
Gilles Louppe
g.louppe@uliege.be

Kyle Cranmer

Johann Brehmer

Joeri Hermans

Antoine Wehenkel



Norman Marlier

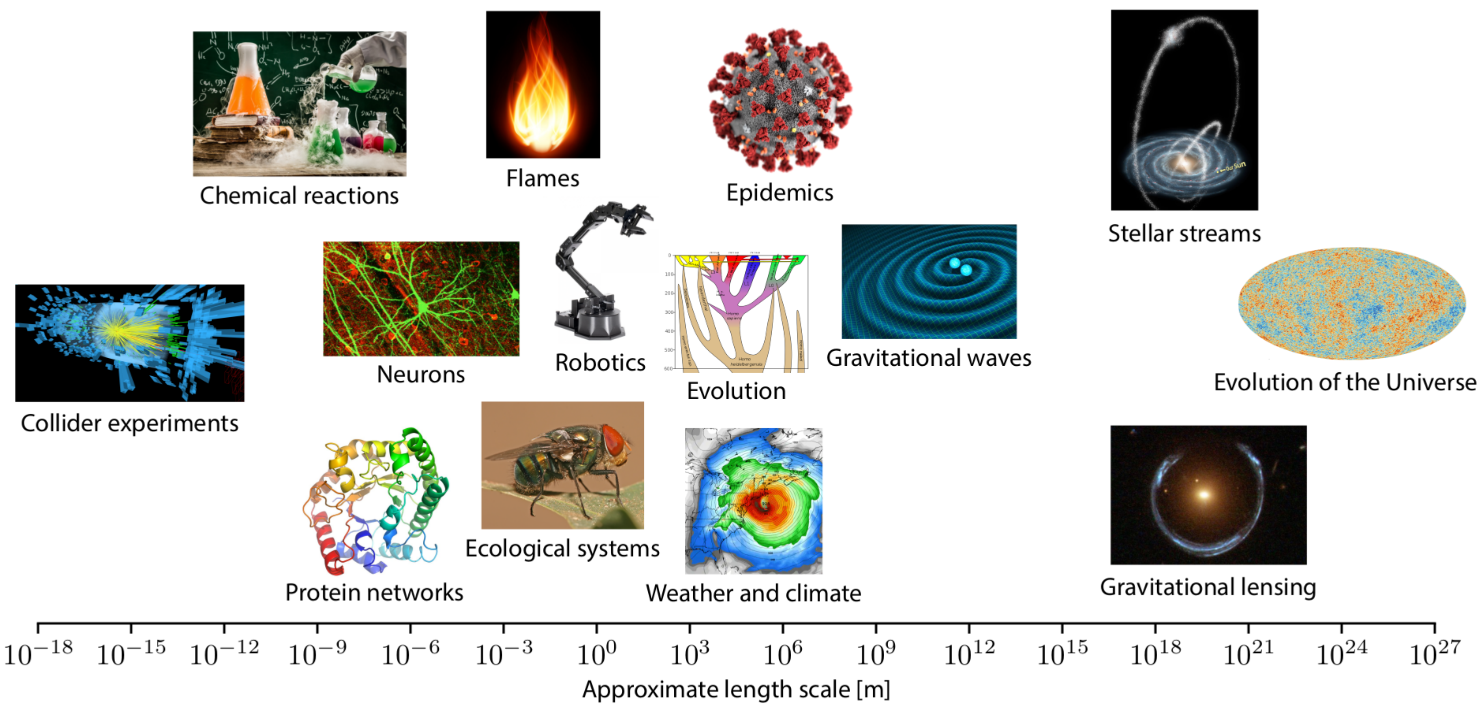
Siddharth Mishra-Sharma

Christoph Weniger

Arnaud Delaunoy

Malavika Vasist

Francois Rozet



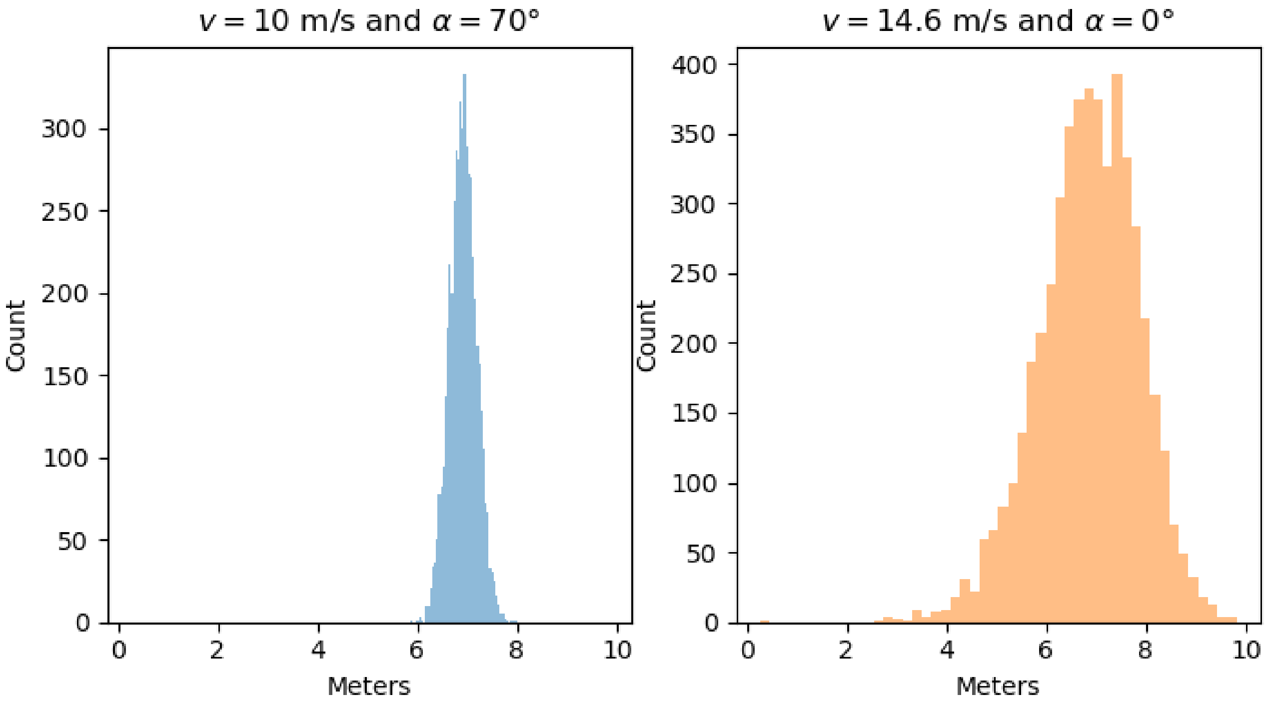
Chemical reactions
Flames
Epidemics
Stellar streams
Collider experiments
Neurons
Robotics
Evolution
Gravitational waves
Evolution of the Universe
Protein networks
Ecological systems
Weather and climate
Gravitational lensing
10^{-18} 10^{-15} 10^{-12} 10^{-9} 10^{-6} 10^{-3} 10^{0} 10^{3} 10^{6} 10^{9} 10^{12} 10^{15} 10^{18} 10^{21} 10^{24} 10^{27}
Approximate length scale [m]

$$v_x = v\cos(\alpha), \quad v_y = v\sin(\alpha),$$

$$\frac{dx}{dt} = v_x\,, \quad \frac{dy}{dt} = v_y, \frac{dv_y}{dt} = -G.$$

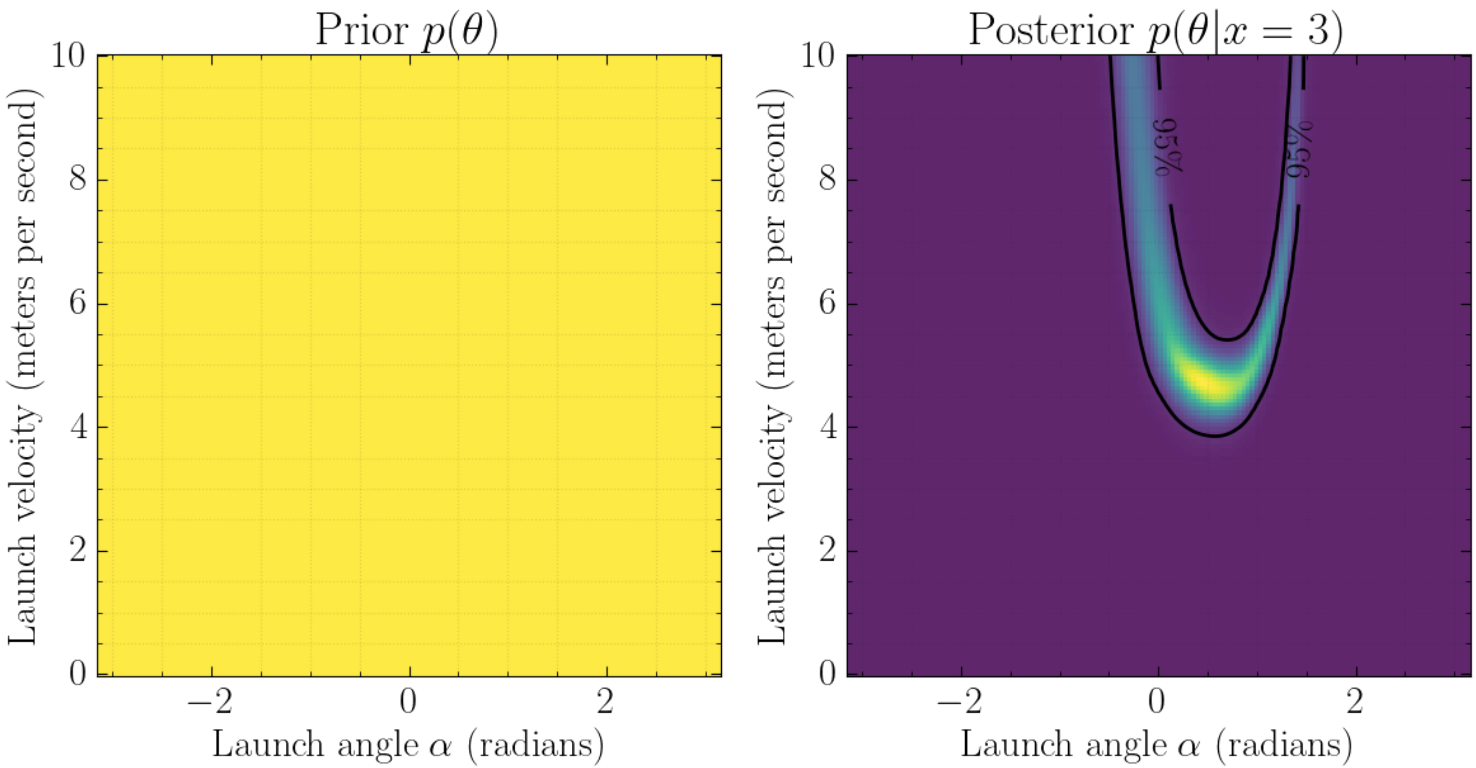
```
def simulate(v, alpha, dt=0.001):
    v_x = v * np.cos(alpha)  # x velocity m/s
    v_y = v * np.sin(alpha)  # y velocity m/s
    y = 1.1 + 0.3 * random.normal()
    x = 0.0

    while y > 0: # simulate until ball hits floor
        v_y += dt * -G  # acceleration due to gravity
        x += dt * v_x
        y += dt * v_y

    return x + 0.25 * random.normal()
```

The computer simulator defines the likelihood function $p(x|\theta)$ implicitly.



What parameter values θ are the most plausible?

$$p(\theta|x_{\mathrm{obs}}) = \frac{p(x_{\mathrm{obs}}|\theta)p(\theta)}{p(x_{\mathrm{obs}})}$$


Prior $p(\theta)$
Launch velocity (meters per second)
Launch angle α (radians)
0
2
4
6
8
10
−2
0
2
Posterior $p(\theta|x = 3)$
Launch velocity (meters per second)
Launch angle α (radians)
95%
95%

The case of particle physics

$$\mathcal{L}_{SM} = -\frac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \frac{1}{4}g_s^2 f^{abc}f^{ade}g^b_\mu g^c_\nu g^d_\mu g^e_\nu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu -$$

$$M^2 W^+_\mu W^-_\mu - \frac{1}{2}\partial_\nu Z^0_\mu\partial_\nu Z^0_\mu - \frac{1}{2c_w^2}M^2 Z^0_\mu Z^0_\mu - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - igc_w(\partial_\nu Z^0_\mu(W^+_\mu W^-_\nu -$$

$$W^+_\nu W^-_\mu) - Z^0_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + Z^0_\mu(W^+_\nu\partial_\nu W^-_\mu - W^-_\nu\partial_\nu W^+_\mu)) -$$

$$igs_w(\partial_\nu A_\mu(W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + A_\mu(W^+_\nu\partial_\nu W^-_\mu -$$

$$W^-_\nu\partial_\nu W^+_\mu)) - \frac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \frac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2c_w^2(Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu -$$

$$Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2(A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w(A_\mu Z^0_\nu(W^+_\mu W^-_\nu -$$

$$W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu) - \frac{1}{2}\partial_\mu H\partial_\mu H - 2M^2\alpha_h H^2 - \partial_\mu\phi^+\partial_\mu\phi^- - \frac{1}{2}\partial_\mu\phi^0\partial_\mu\phi^0 -$$

$$\beta_h\left(\frac{2M^2}{g^2} + \frac{2M}{g}H + \frac{1}{2}(H^2+\phi^0\phi^0+2\phi^+\phi^-)\right) + \frac{2M^4}{g^2}\alpha_h -$$

$$g\alpha_h M\left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) -$$

$$\frac{1}{8}g^2\alpha_h\left(H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2H^2\right) -$$

$$gMW^+_\mu W^-_\mu H - \frac{1}{2}g\frac{M}{c_w^2}Z^0_\mu Z^0_\mu H -$$

$$\frac{1}{2}ig\left(W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)\right) +$$

$$\frac{1}{2}g\left(W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) + W^-_\mu(H\partial_\mu\phi^+ - \phi^+\partial_\mu H)\right) + \frac{1}{2}g\frac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) +$$

$$M\left(\frac{1}{c_w}Z^0_\mu\partial_\mu\phi^0 + W^+_\mu\partial_\mu\phi^- + W^-_\mu\partial_\mu\phi^+\right) - ig\frac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + igs_wMA_\mu(W^+_\mu\phi^- -$$

$$W^-_\mu\phi^+) - ig\frac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + igs_wA_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) -$$

$$\frac{1}{4}g^2W^+_\mu W^-_\mu\left(H^2 + (\phi^0)^2 + 2\phi^+\phi^-\right) - \frac{1}{8}g^2\frac{1}{c_w^2}Z^0_\mu Z^0_\mu\left(H^2+(\phi^0)^2 + 2(2s_w^2-1)^2\phi^+\phi^-\right) -$$

$$\frac{1}{2}g^2\frac{s_w^2}{c_w}Z^0_\mu\phi^0(W^+_\mu\phi^- + W^-_\mu\phi^+) - \frac{1}{2}ig^2\frac{s_w^2}{c_w}Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \frac{1}{2}g^2s_wA_\mu\phi^0(W^+_\mu\phi^- +$$

$$W^-_\mu\phi^+) + \frac{1}{2}ig^2s_wA_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\frac{s_w}{c_w}(2c_w^2-1)Z^0_\mu A_\mu\phi^+\phi^- -$$

$$g^2s_w^2A_\mu A_\mu\phi^+\phi^- + \frac{1}{2}ig_s\lambda^a_{ij}(\bar q^\sigma_i\gamma^\mu q^\sigma_j)g^a_\mu - \bar e^\lambda(\gamma\partial + m^\lambda_e)e^\lambda - \bar\nu^\lambda(\gamma\partial+m^\lambda_\nu)\nu^\lambda - \bar u^\lambda_j(\gamma\partial +$$

$$m^\lambda_u)u^\lambda_j - \bar d^\lambda_j(\gamma\partial + m^\lambda_d)d^\lambda_j + igs_wA_\mu\left(-(\bar e^\lambda\gamma^\mu e^\lambda) + \frac{2}{3}(\bar u^\lambda_j\gamma^\mu u^\lambda_j) - \frac{1}{3}(\bar d^\lambda_j\gamma^\mu d^\lambda_j)\right) +$$

$$\frac{ig}{4c_w}Z^0_\mu\{(\bar\nu^\lambda\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar e^\lambda\gamma^\mu(4s_w^2-1-\gamma^5)e^\lambda) + (\bar d^\lambda_j\gamma^\mu(\tfrac{4}{3}s_w^2-1-\gamma^5)d^\lambda_j) +$$

$$(\bar u^\lambda_j\gamma^\mu(1-\tfrac{8}{3}s_w^2+\gamma^5)u^\lambda_j)\} + \frac{ig}{2\sqrt2}W^+_\mu\left((\bar\nu^\lambda\gamma^\mu(1+\gamma^5)U^{lep}{}_{\lambda\kappa}e^\kappa) + (\bar u^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa}d^\kappa_j)\right) +$$

$$\frac{ig}{2\sqrt2}W^-_\mu\left((\bar e^\kappa U^{lep\dagger}{}_{\kappa\lambda}\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar d^\kappa_j C^\dagger_{\kappa\lambda}\gamma^\mu(1+\gamma^5)u^\lambda_j)\right) +$$

$$\frac{ig}{2M\sqrt2}\phi^+\left(-m^\kappa_e(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1-\gamma^5)e^\kappa) + m^\lambda_\nu(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1+\gamma^5)e^\kappa\right) +$$

$$\frac{ig}{2M\sqrt2}\phi^-\left(m^\lambda_e(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1+\gamma^5)\nu^\kappa) - m^\kappa_\nu(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1-\gamma^5)\nu^\kappa\right) - \frac{g}{2}\frac{m^\lambda_\nu}{M}H(\bar\nu^\lambda\nu^\lambda) -$$

$$\frac{g}{2}\frac{m^\lambda_e}{M}H(\bar e^\lambda e^\lambda) + \frac{ig}{2}\frac{m^\lambda_\nu}{M}\phi^0(\bar\nu^\lambda\gamma^5\nu^\lambda) - \frac{ig}{2}\frac{m^\lambda_e}{M}\phi^0(\bar e^\lambda\gamma^5 e^\lambda) - \frac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa -$$

$$\frac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa + \frac{ig}{2M\sqrt2}\phi^+\left(-m^\kappa_d(\bar u^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + m^\lambda_u(\bar u^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j\right) +$$

$$\frac{ig}{2M\sqrt2}\phi^-\left(m^\lambda_d(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m^\kappa_u(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1-\gamma^5)u^\kappa_j\right) - \frac{g}{2}\frac{m^\lambda_u}{M}H(\bar u^\lambda_j u^\lambda_j) -$$

$$\frac{g}{2}\frac{m^\lambda_d}{M}H(\bar d^\lambda_j d^\lambda_j) + \frac{ig}{2}\frac{m^\lambda_u}{M}\phi^0(\bar u^\lambda_j\gamma^5 u^\lambda_j) - \frac{ig}{2}\frac{m^\lambda_d}{M}\phi^0(\bar d^\lambda_j\gamma^5 d^\lambda_j) + \bar G^a\partial^2 G^a + g_s f^{abc}\partial_\mu\bar G^a G^b g^c_\mu +$$

$$\bar X^+(\partial^2 - M^2)X^+ + \bar X^-(\partial^2-M^2)X^- + \bar X^0(\partial^2 - \frac{M^2}{c_w^2})X^0 + \bar Y\partial^2 Y + igc_wW^+_\mu(\partial_\mu\bar X^0 X^- -$$

$$\partial_\mu\bar X^+ X^0) + igs_wW^+_\mu(\partial_\mu\bar Y X^- - \partial_\mu\bar X^+ Y) + igc_wW^-_\mu(\partial_\mu\bar X^- X^0 -$$

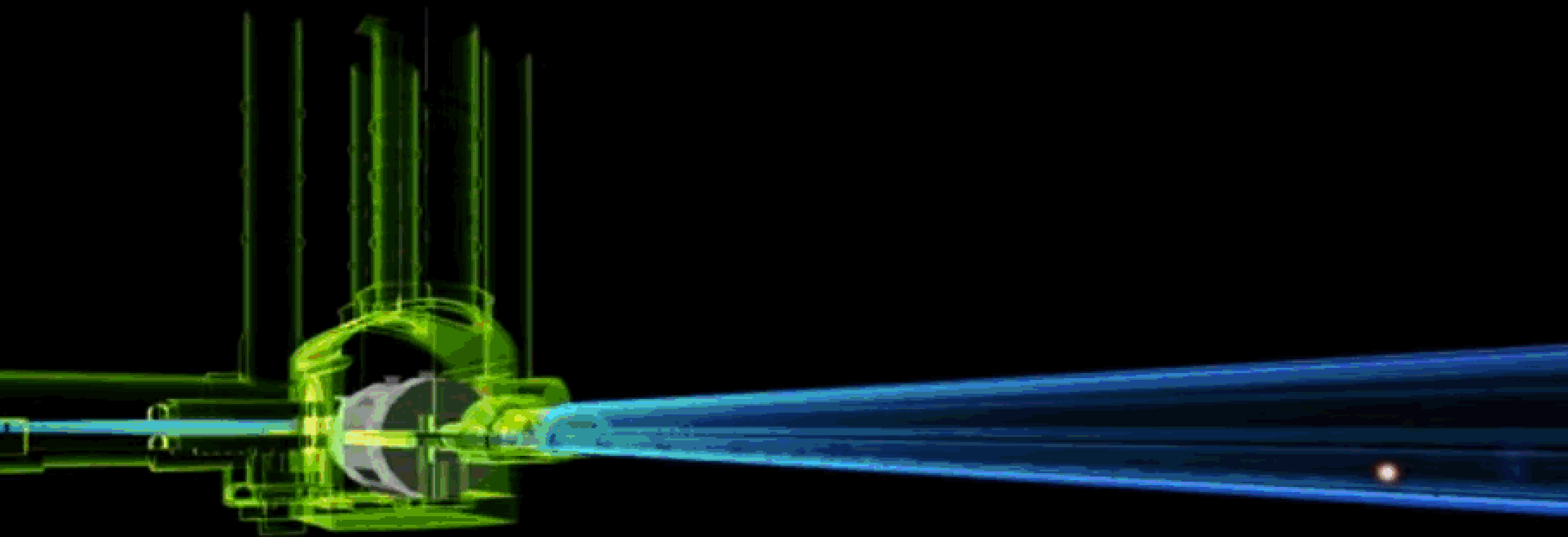
$$\partial_\mu\bar X^0 X^+) + igs_wW^-_\mu(\partial_\mu\bar X^- Y - \partial_\mu\bar Y X^+) + igc_wZ^0_\mu(\partial_\mu\bar X^+X^+ -$$

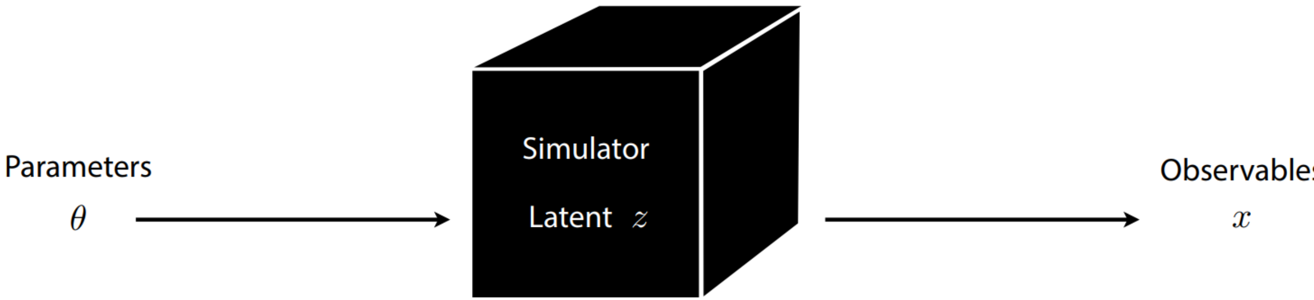
$$\partial_\mu\bar X^- X^-) + igs_wA_\mu(\partial_\mu\bar X^+X^+ -$$

$$\partial_\mu\bar X^-X^-) - \frac{1}{2}gM\left(\bar X^+X^+H + \bar X^-X^-H + \frac{1}{c_w^2}\bar X^0X^0H\right) + \frac{1-2c_w^2}{2c_w}igM\left(\bar X^+X^0\phi^+ - \bar X^-X^0\phi^-\right) +$$

$$\frac{1}{2c_w}igM\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + igMs_w\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) +$$

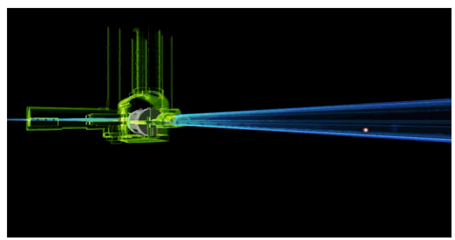
$$\frac{1}{2}igM\left(\bar X^+X^+\phi^0 - \bar X^-X^-\phi^0\right).$$



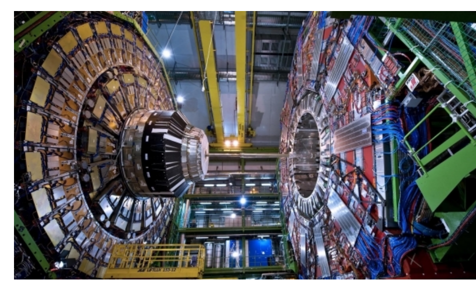


SM with parameters θ

Simulated observables x



Real observations x_{obs}



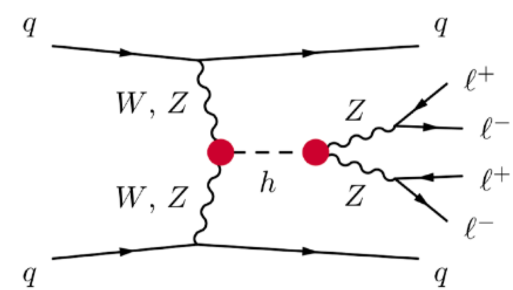
Latent variables

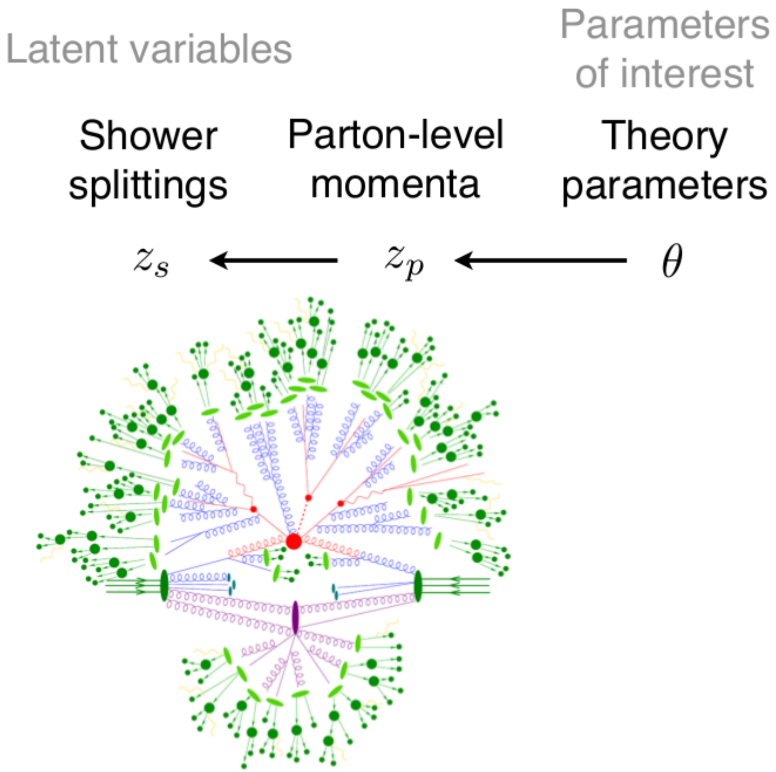
Parameters of interest

Parton-level momenta

Theory parameters

$$z_p \longleftarrow \theta$$





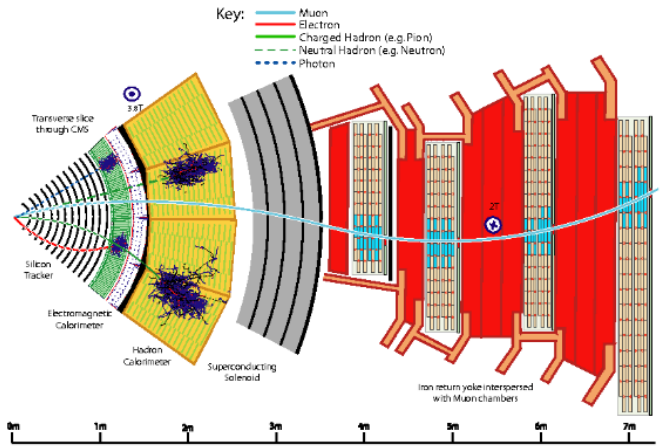
Latent variables
Parameters of interest
Shower splittings
Parton-level momenta
Theory parameters
z_s
z_p
θ

Latent variables

Parameters of interest

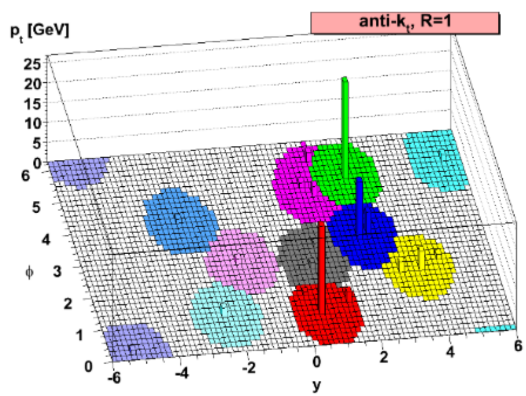
Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

$$z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$



Features		Latent variables		Parameters of interest
Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
x	z_d	z_s	z_p	θ

$$x \longleftarrow z_d \longleftarrow z_s \longleftarrow z_p \longleftarrow \theta$$



[Image source: M. Cacciari,
G. Salam, G. Soyez 0802.1189]

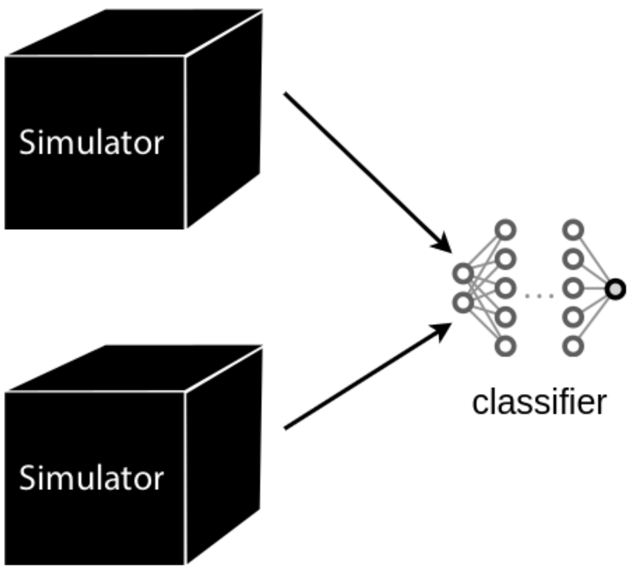
$$p(x|\theta) = \underbrace{\iiint}_{\text{yikes!}} p(z_p|\theta)p(z_s|z_p)p(z_d|z_s)p(x|z_d)dz_p dz_s dz_d$$

Neural ratio estimation (NRE)

The likelihood-to-evidence $r(x|\theta) = \frac{p(x|\theta)}{p(x)} = \frac{p(x,\theta)}{p(x)p(\theta)}$ ratio can be learned, even if neither the likelihood nor the evidence can be evaluated:

$x, \theta \sim p(x, \theta)$

$x, \theta \sim p(x)p(\theta)$



$\hat{r}(x|\theta)$

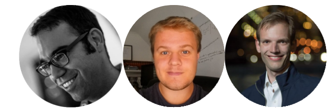
—
Credits: Cranmer et al, 2015; Hermans et al, 2020.

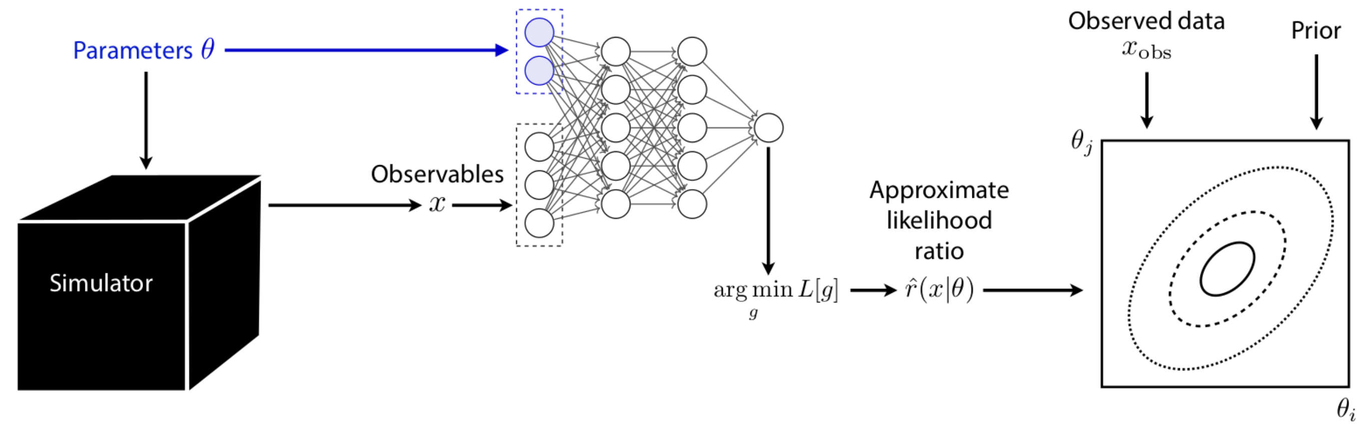
The solution d found after training approximates the optimal classifier

$$d(x, \theta) \approx d^*(x, \theta) = \frac{p(x, \theta)}{p(x, \theta) + p(x)p(\theta)}.$$

Therefore,

$$r(x|\theta) = \frac{p(x|\theta)}{p(x)} = \frac{p(x, \theta)}{p(x)p(\theta)} \approx \frac{d(x, \theta)}{1 - d(x, \theta)} = \hat{r}(x|\theta).$$





1. Simulation

Run simulator and save data

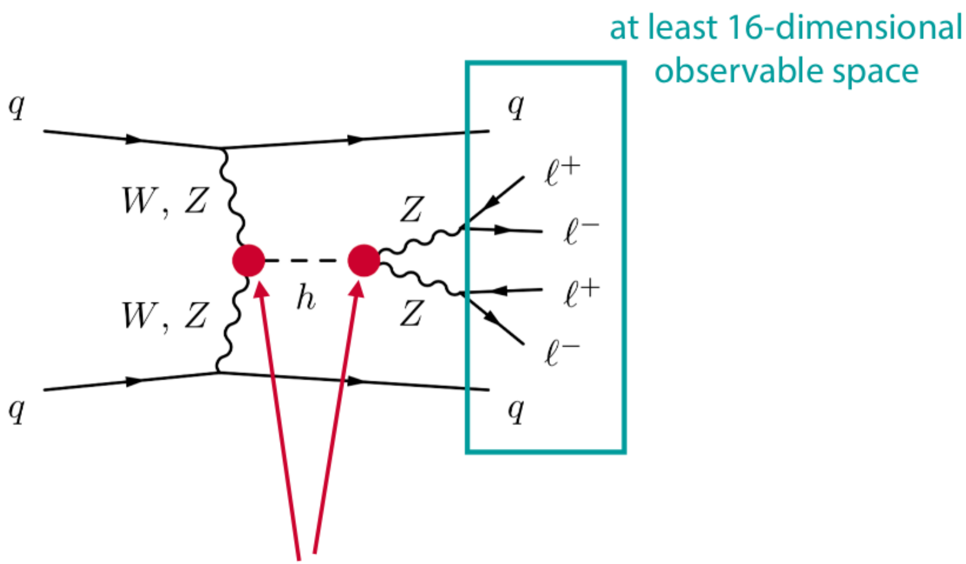
2. Machine Learning

Train NN classifier, interpret as likelihood ratio estimator

3. Inference

Amortized: cheap to repeat for new data

Hunting new physics at particle colliders (Frequentist)

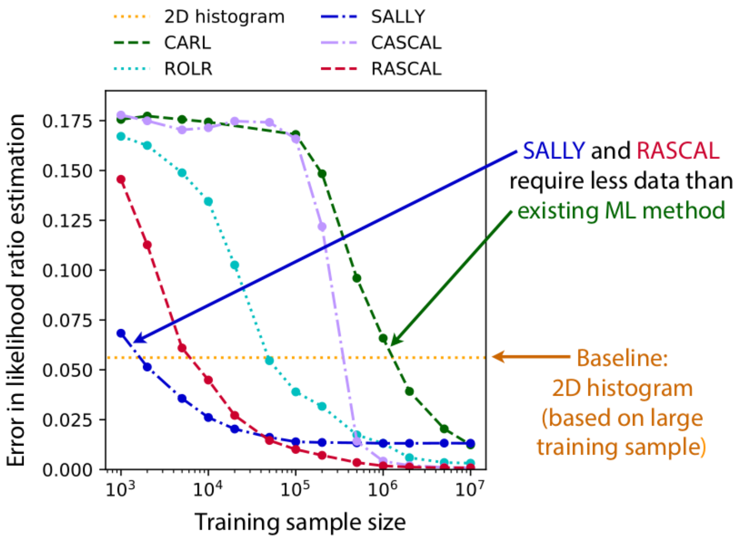


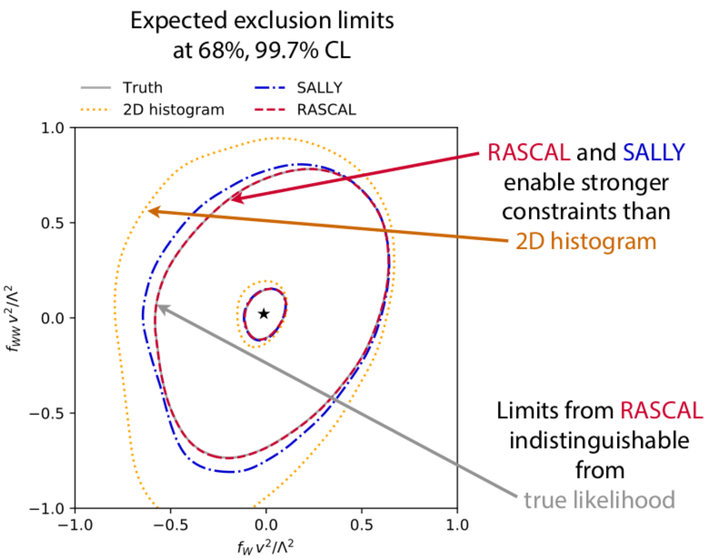
Exciting new physics might hide here!
We parameterize it with two EFT coefficients:

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \boxed{\frac{f_W}{\Lambda^2}}\,\underbrace{\frac{\mathrm{i}g}{2}\,(D^\mu\phi)^\dagger\,\sigma^a\,D^\nu\phi\,W^a_{\mu\nu}}_{\mathcal{O}_W} - \boxed{\frac{f_{WW}}{\Lambda^2}}\,\underbrace{\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^a_{\mu\nu}\,W^{\mu\nu\,a}}_{\mathcal{O}_{WW}}$$

—

Credits: Brehmer et al, 2018.

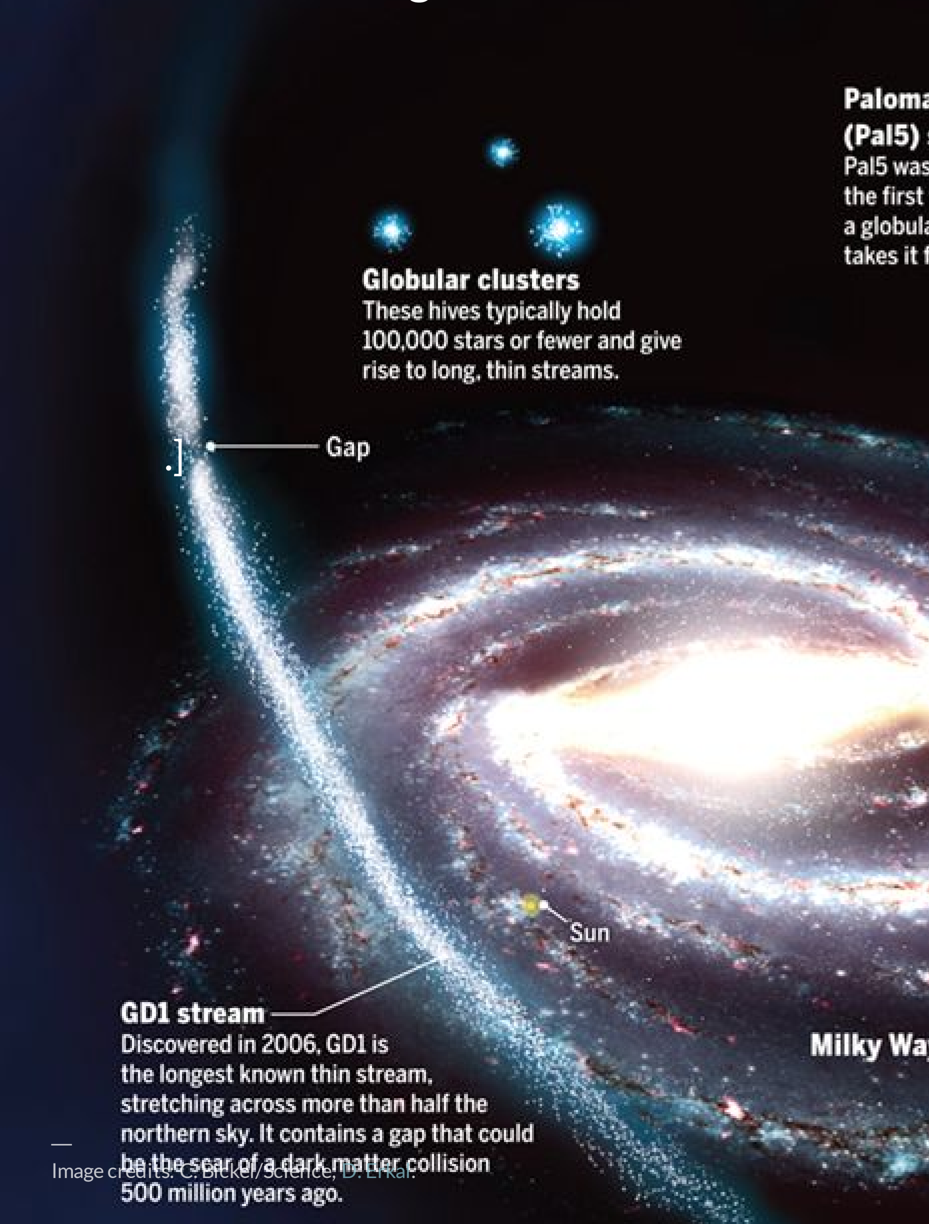





—

Credits: Brehmer et al, 2018.

Constraining dark matter with stellar streams (Bayesian)



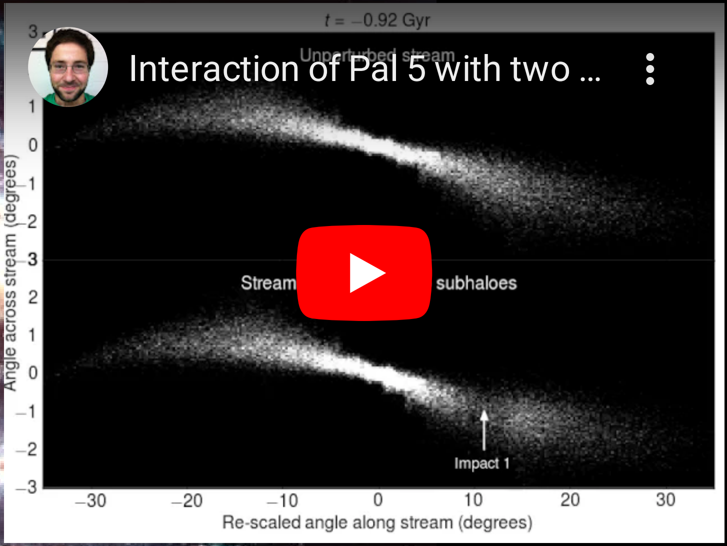
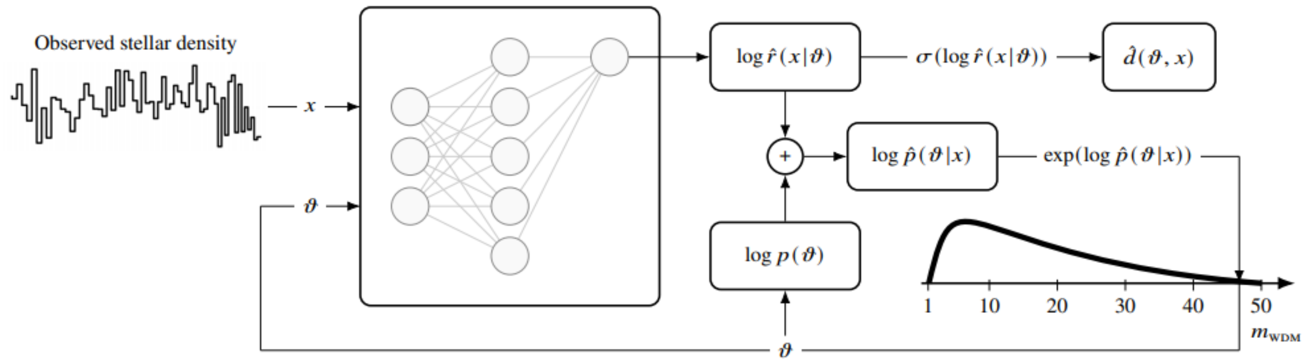
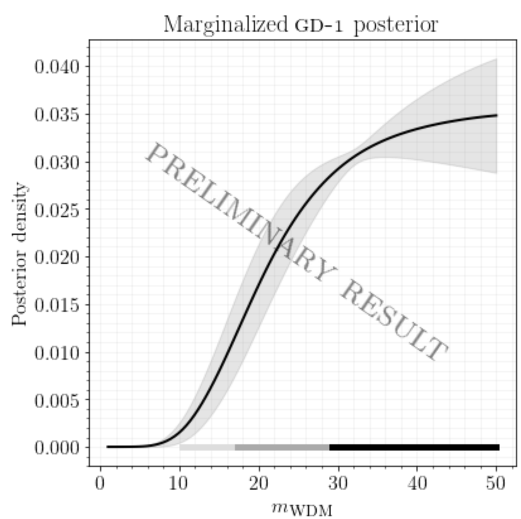
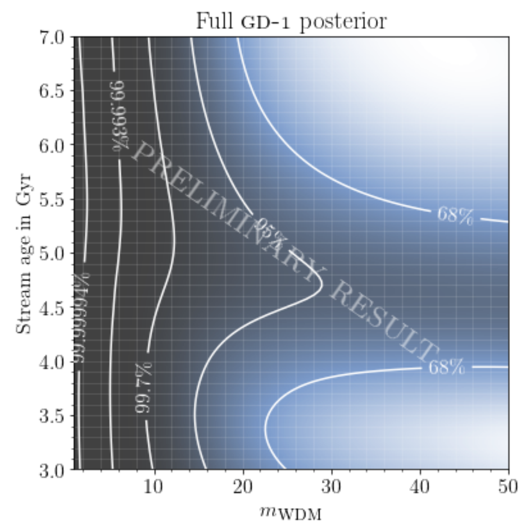


Image credits: C. Bickel/Science, D. Erkal.







—
Credits: Hermans et al, 2021.

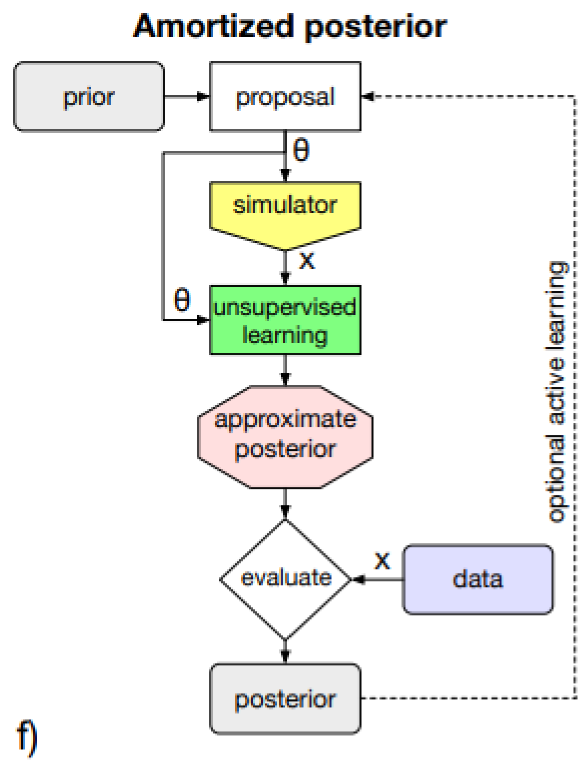
Preliminary results for GD-1 suggest a **preference for CDM over WDM**.

Neural Posterior Estimation (NPE)

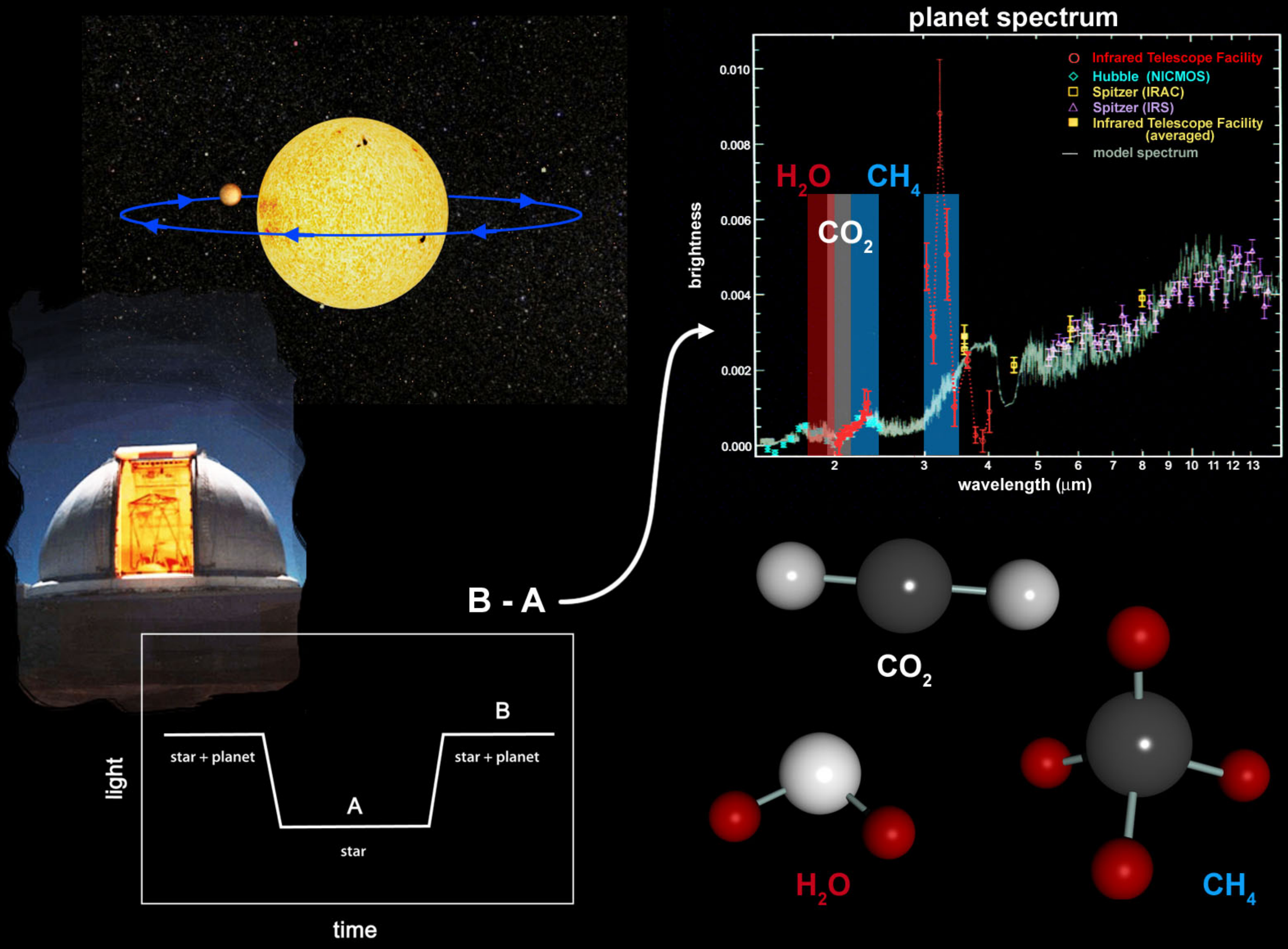
Use variational inference to directly estimate the posterior, by solving

$$\min_{q_\phi} \mathbb{E}_{p(x)} \left[\mathrm{KL}(p(\theta|x) || q_\phi(\theta|x)) \right]$$

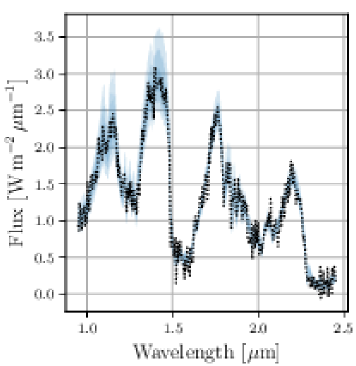
where q_ϕ is a neural density estimator, such as a normalizing flow.



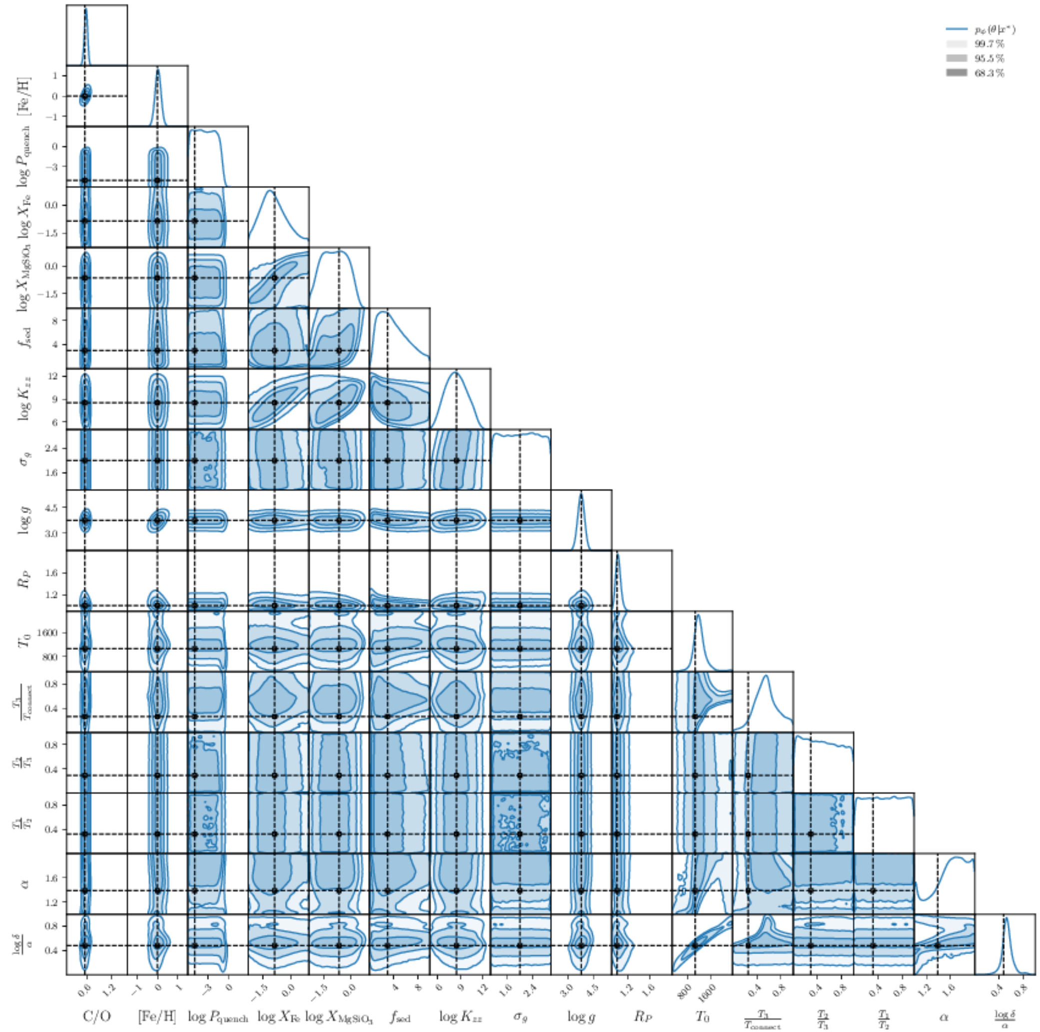
Exoplanet atmosphere characterization (Bayesian)



—
Credits: NSA/JPL-Caltech, 2010.



Flux [W m^-2 μm^-1]
Wavelength [μm]

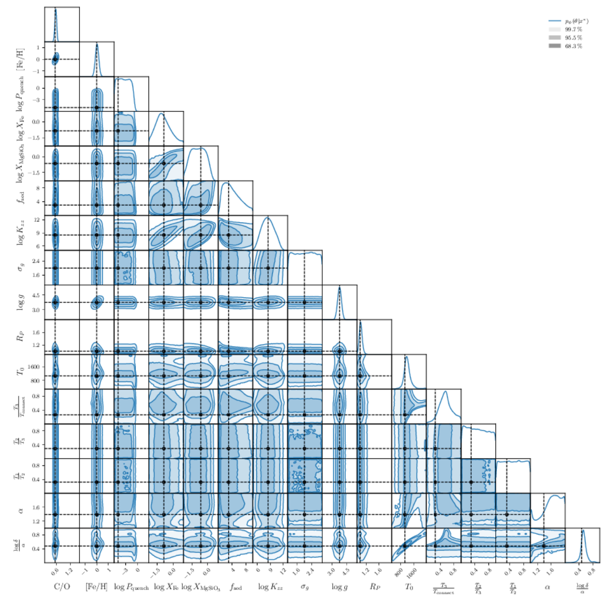


$p_\phi(\theta|x^*)$
99.7 %
95.5 %
68.3 %
C/O
[Fe/H]
log $P_{\rm quench}$
log $X_{\rm Fe}$
log $X_{\rm MgSiO_3}$
$f_{\rm sed}$
log K_{zz}
σ_g
log g
R_P
T_0
$\frac{T_3}{T_{\rm connect}}$
$\frac{T_2}{T_3}$
$\frac{T_1}{T_2}$
α
$\frac{\log \delta}{\alpha}$

Computational faithfulness

$$\hat{p}(\theta|x) = \text{sbi}(p(x|\theta), p(\theta), x)$$

We must make sure our approximate simulation-based inference algorithms can (at least) actually realize faithful inferences on the (expected) observations.

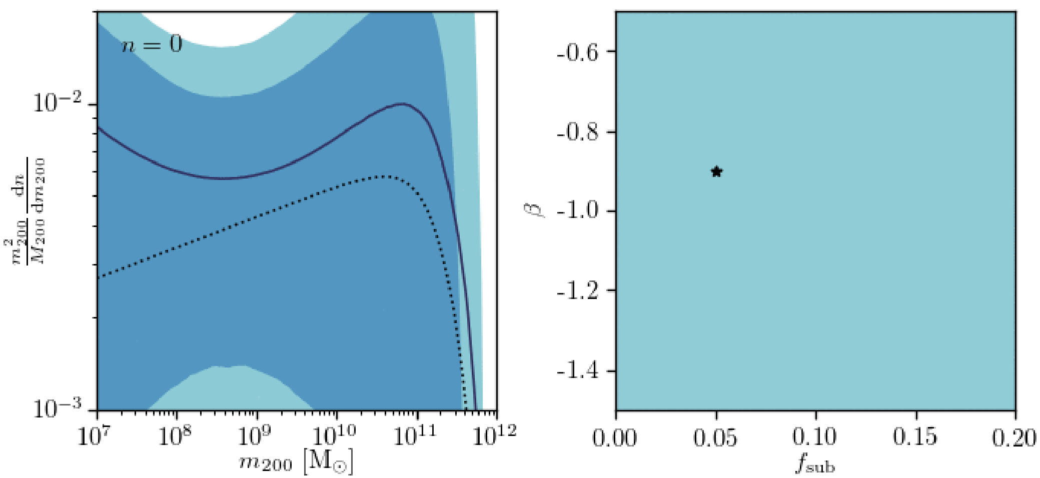


How do we know this is good enough?

Mode convergence:

The maximum a posteriori estimate converges towards the nominal value θ^* for an increasing number of independent and identically distributed observables $x_i \sim p(x|\theta^*)$:

$$\lim_{N\to\infty} \arg\max_\theta p(\theta|\{x_i\}_{i=1}^N)$$
$$= \lim_{N\to\infty} \arg\max_\theta p(\theta) \prod_{x_i} r(x_i|\theta) = \theta^*$$



—

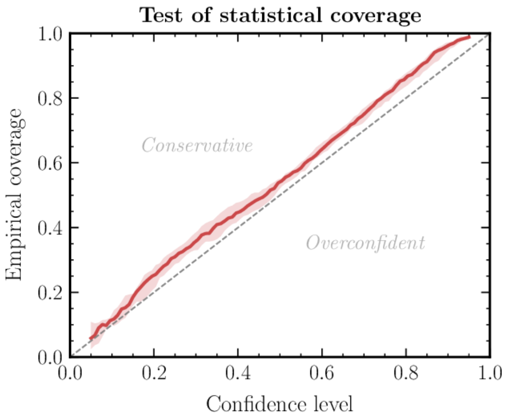
Credits: Brehmer et al, 2019.

A common observation at the root of several other diagnostics is to check for the **self-consistency** of the Bayesian joint distribution,

$$p(\theta) = \int p(\theta')p(x|\theta')p(\theta|x)d\theta'\, dx.$$

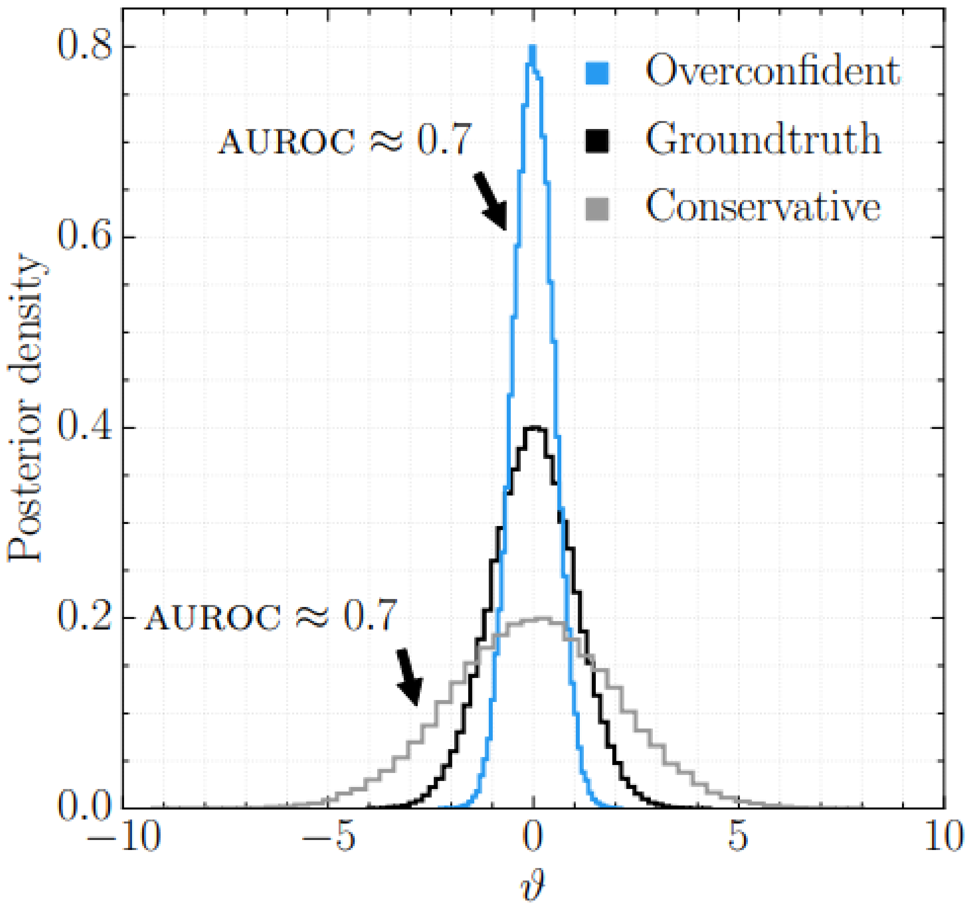
Coverage diagnostic:

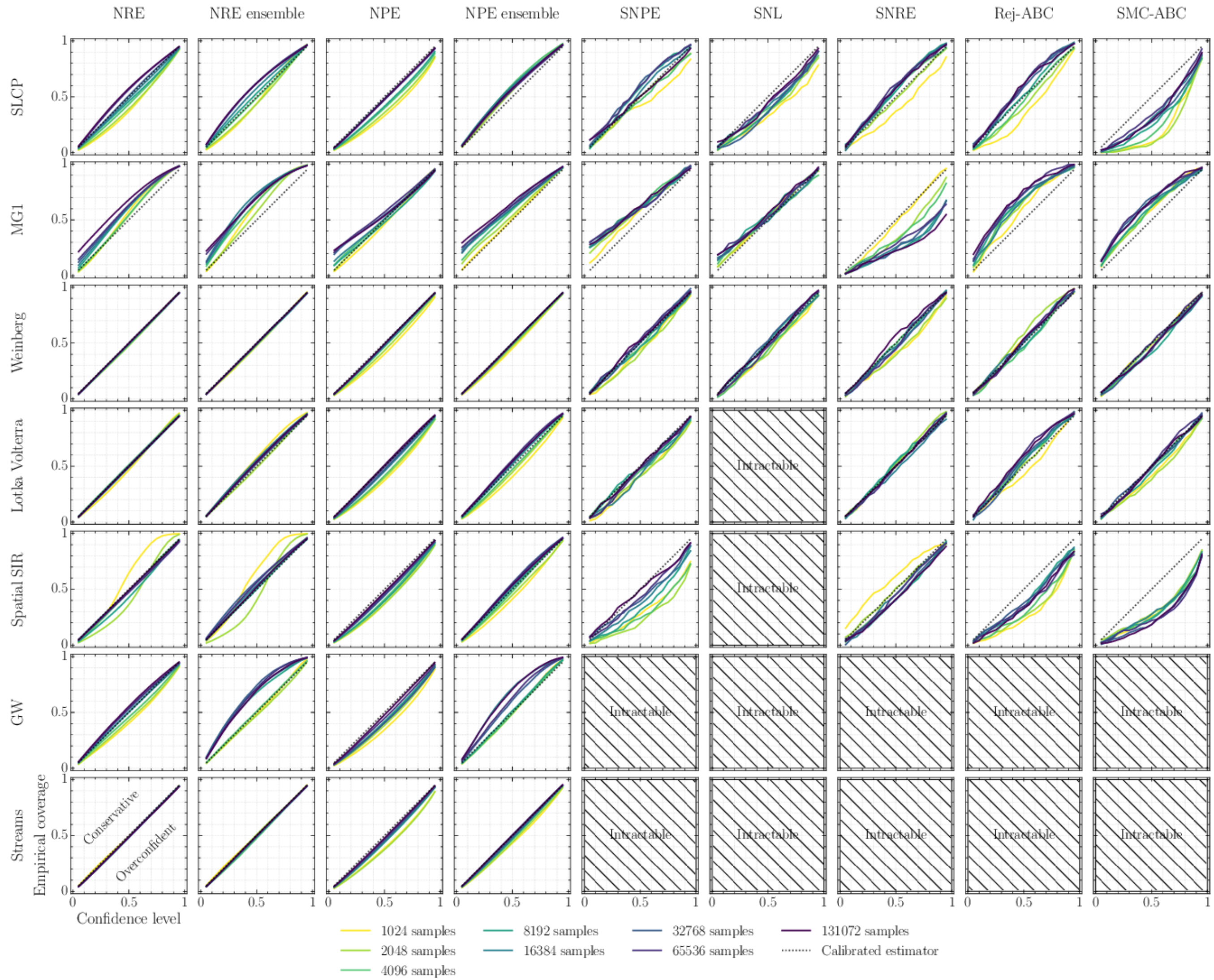
- For $x, \theta \sim p(x, \theta)$, compute the $1-\alpha$ credible interval based on $\hat{p}\left(\theta|x\right)$.
- If the fraction of samples for which θ is contained within the interval is larger than the nominal coverage probability $1-\alpha$, then the approximate posterior $\hat{p}\left(\theta|x\right)$ has coverage.



—

Credits: Hermans et al, 2021; Siddharth Mishra-Sharma, 2021.


Overconfident
Groundtruth
Conservative
AUROC ≈ 0.7
AUROC ≈ 0.7
Posterior density
0.8
0.6
0.4
0.2
0.0
−10
−5
0
5
10
ϑ



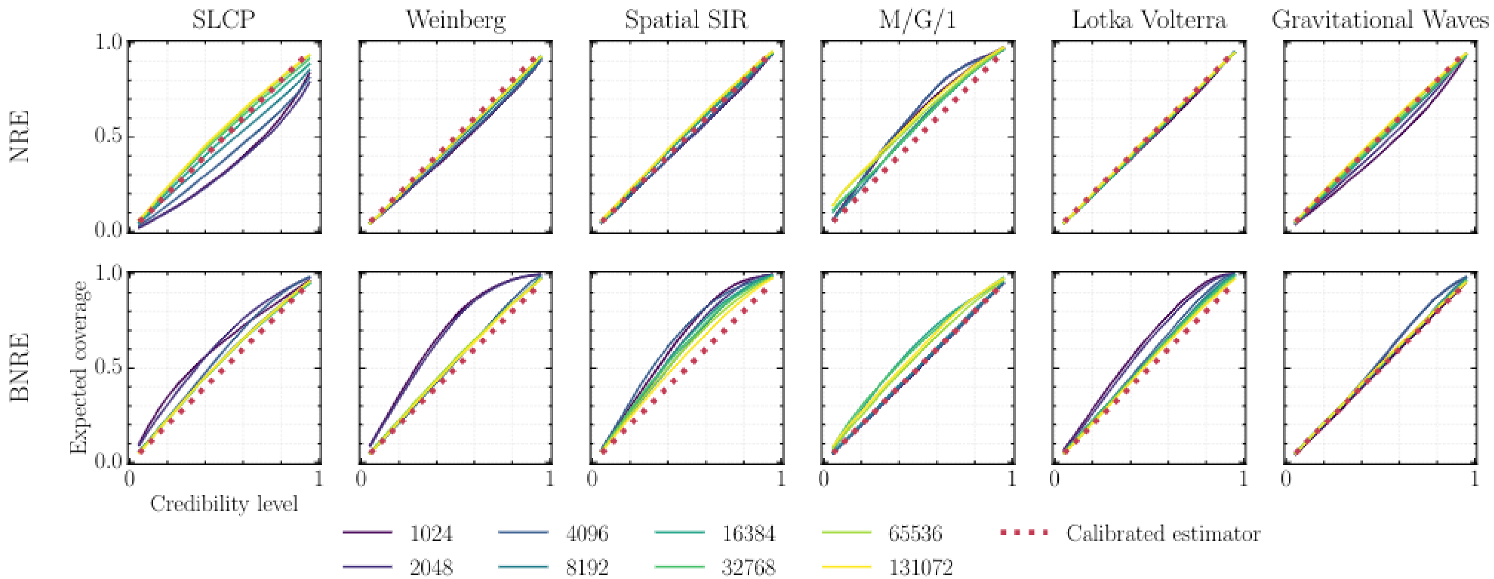
—
Credits: Hermans et al, 2021.

What if diagnostics fail?

Balanced NRE

Enforce neural ratio estimation to be **conservative** by using binary classifiers $\hat{d}$ that are balanced, i.e. such that

$$\mathbb{E}_{p(\theta,x)}\left[\hat{d}\left(\theta,x\right)\right] = \mathbb{E}_{p(\theta)p(x)}\left[1 - \hat{d}\left(\theta,x\right)\right].$$



—
Credits: Delaunoy et al, 2022.

Summary

Simulation-based inference is a major evolution in the statistical capabilities for science, enabled by advances in machine learning.

Need to reliably and efficiently evaluate the quality of the posterior approximations.

The end.