The impact of model realism on interpretations of the Galactic Center Excess

Speaker: Christopher ECKNER

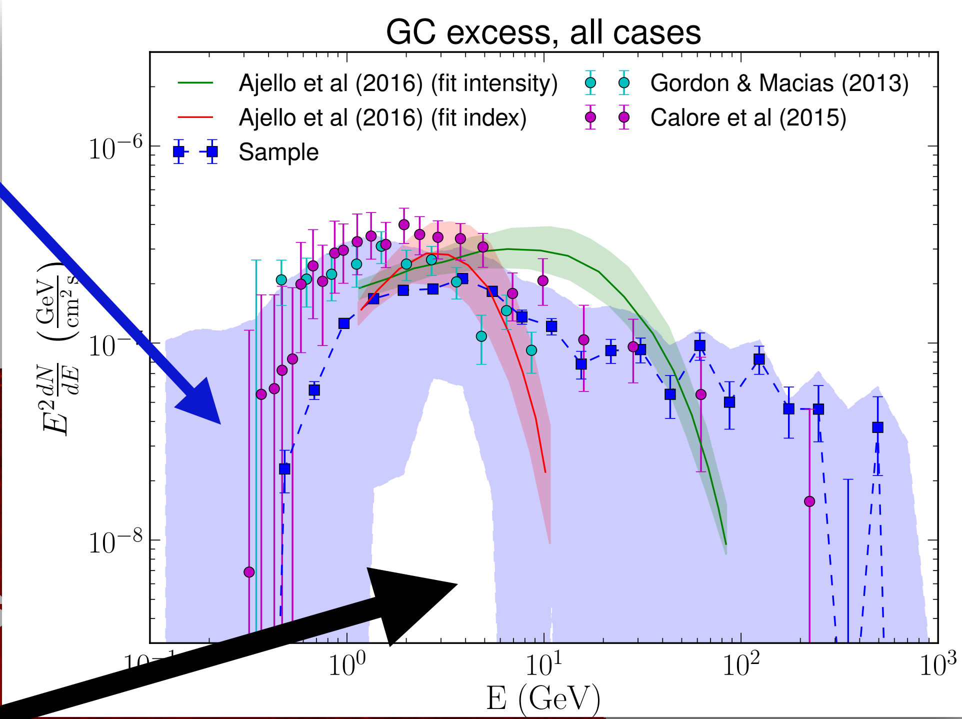
29th August 2023,
TAUP 2023
Based on: JCAP 06 (2023) 013

On behalf of: Sascha Caron, Luc Hendriks, Guðlaugur Jóhannesson, Roberto Ruiz de Austri, Gabrijela Zaharijas

What is the Fermi GeV excess ?

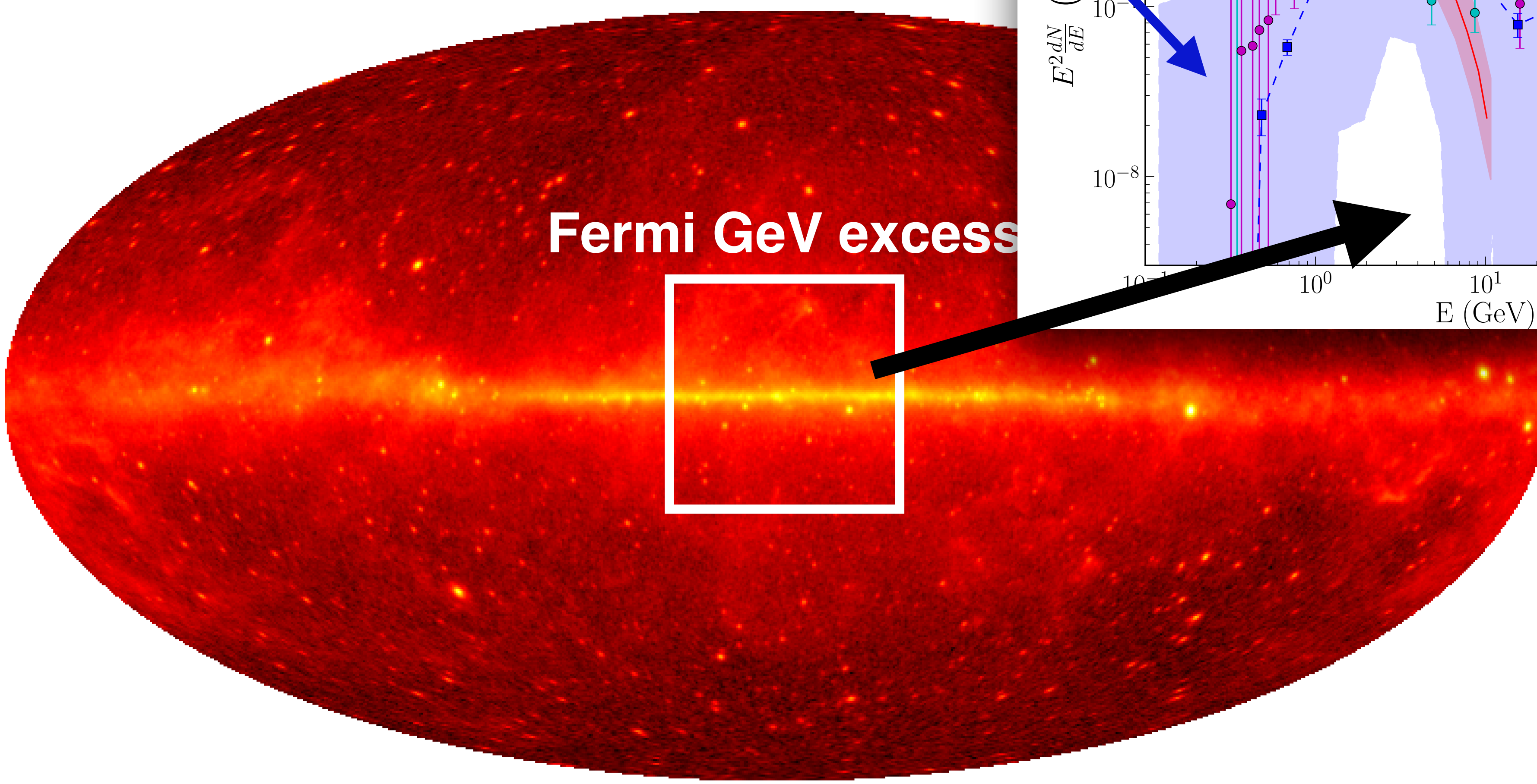
We all agree: There is an excess of GeV gamma rays (GCE) toward the Galactic centre measured by Fermi-LAT **above known astrophysical backgrounds**.

[Fermi collab. ApJ 840 (2017) 1]



An incomplete list of works:

Goodenough & Hooper (2009)
Vitale & Morselli (2009)
Hooper & Goodenough (2011)
Hooper & Linden (2011)
Boyarsky et al (2011)
Abazajian & Kaplinghat (2012)
Gordon & Macias (2013)
Macias & Gordon (2014)
Abazajian et al (2014, 2015)
Calore et al (2014)
Daylan et al (2014)
Selig et al (2015)
Huang et al (2015)
Gaggero et al (2015)
Carlson et al (2015, 2016)
de Boer et al (2016)
Yang & Aharonian (2016)
Fermi Coll. (2016)
Horiuchi et al (2016)
Linden et al (2016)
Ackermann et al (2017)
Macias et al (2018)
Bartels et al (2018)
Balaji et al (2018)
Zhong et al (2019)
Macias et al (2019)
Chang et al (2020)
Buschmann et al (2020)
Leane & Slatyer (2020)
Abazajian et al (2020)
List et L (2020)
Di Mauro (2020)
Burns et al (2020)
Cholis et al (2022)
Pohl, Macias+ (2022)
...



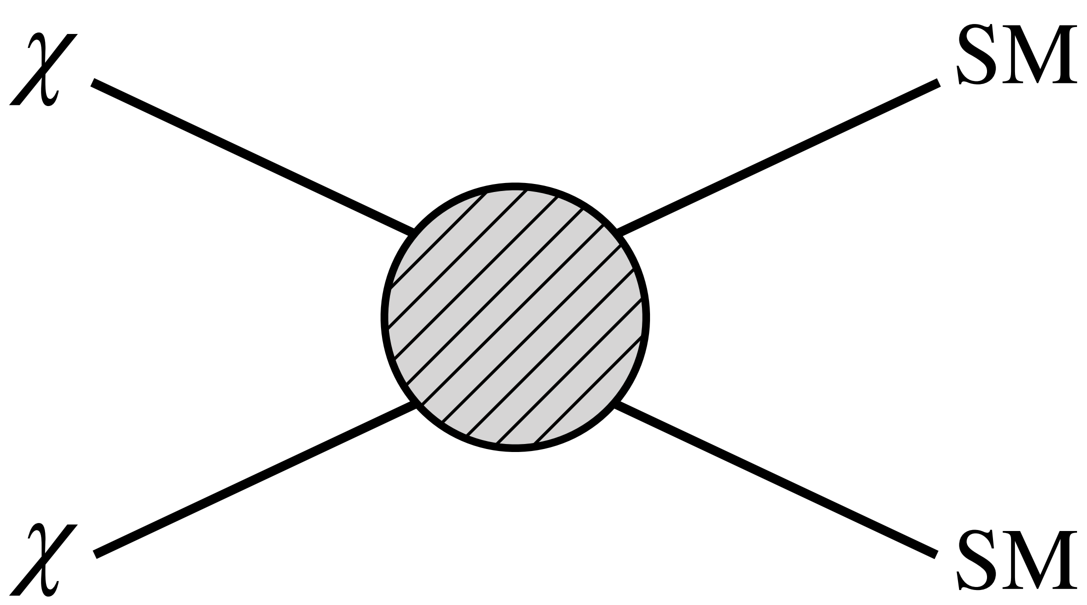
Where we do not agree:

1. What is the preferred spatial morphology of the excess?
2. **What is causing the Fermi GeV excess?**

What produces the excess?

The excess is tantalising since it coincides well with the expectations for the sought-after signal of **thermal dark matter pair-annihilating** in the Galactic centre. However, **unresolved populations of gamma-ray sources** are a strong contender! [See D. Hooper, arXiv:2209.14370 for a different view.]

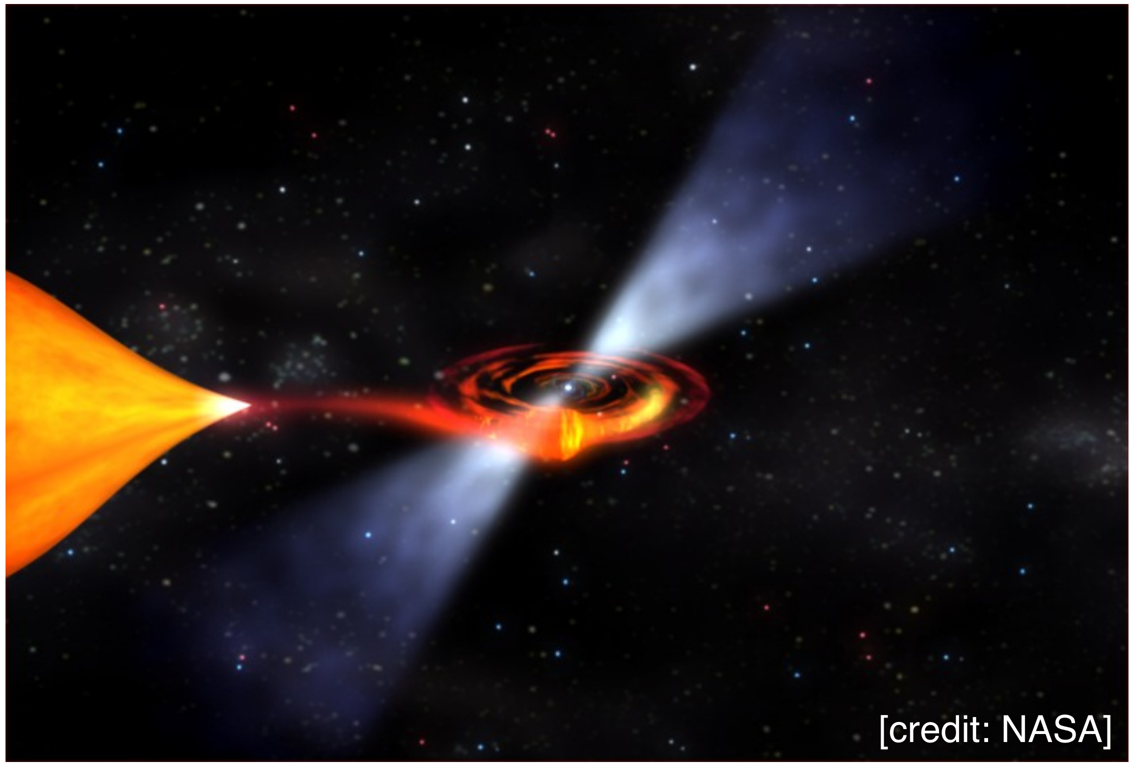
Thermal dark matter



supported by (incomplete collection):
[Fermi collab. ApJ 840 (2017) 1];
[R. K. Leane and T. R. Slatyer, PRL 123 (2019) 24];
[M. di Mauro, PRD 103 (2021) 6]; [I. Cholis et al., PRD 105 (2022) 10];
[S. D. McDermott et al., arXiv:2209.00006]

VS.

Unresolved Galactic source population (here: millisecond pulsars [MSPs])



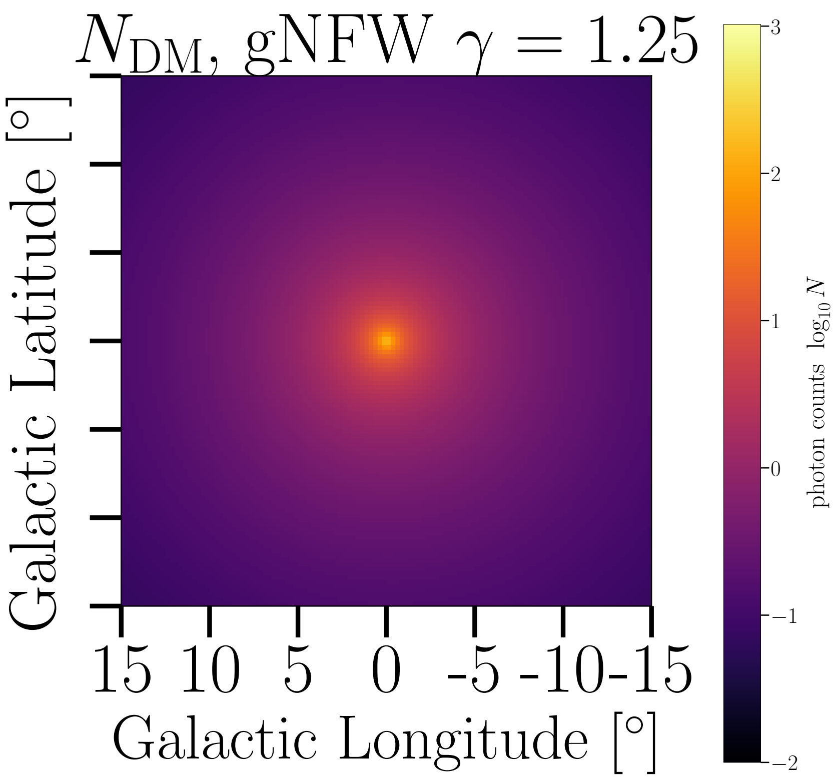
supported by (incomplete collection):
[R. Bartels et al., PRL 116 (2016) 5];
[R. Bartels et al., Nature Astron. 2 (2018) 10];
[O. Macias et al., JCAP 09 (2019) 042];
[F. Calore et al., PRL 127 (2021) 16];
[M. Pohl et al., ApJ 929 (2022) 2]

Our question: How robust are those interpretations given the caveat that the excess is above **known** astrophysical background? **Does the background model uncertainty impact the final results?**

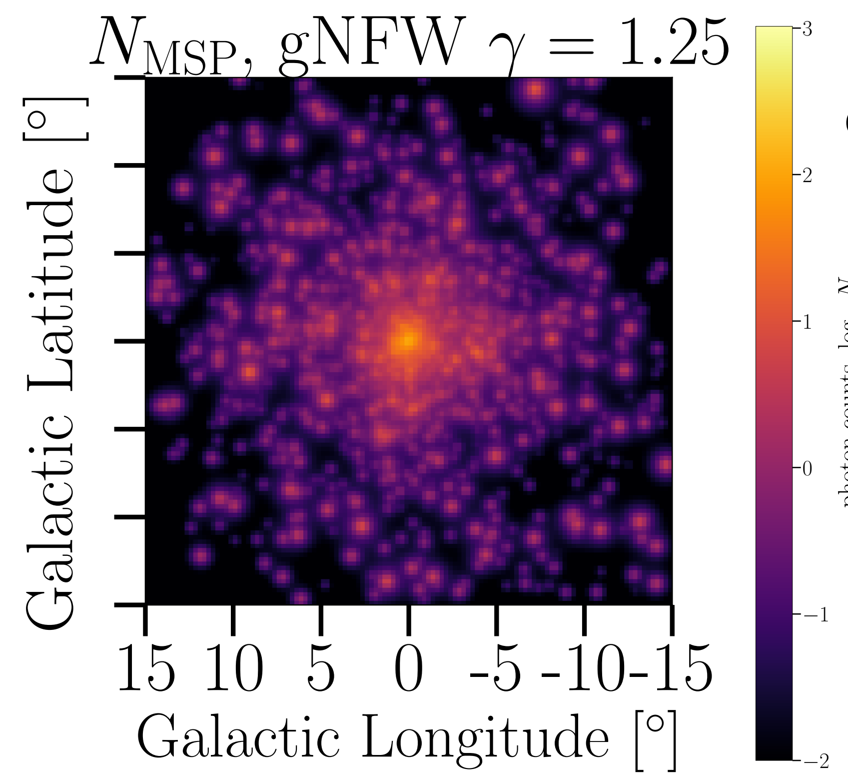
How machine learning can help

A decisive feature of the GeV excess is its **photon clustering behaviour**, spectrally they can be almost identical.

DM annihilation
(smooth morphology,
Poisson-distributed photon events)



Faint millisecond pulsar population
(photon clustering on small scales,
non-Poissonian noise component)



1 - 2 GeV energy band

- Traditional likelihood methods cannot explore this difference in any practical way (probabilistic nature of point source locations and fluxes!)
- Effective methods have been proposed: **non-Poissonian template fitting, 1pPDF, wavelet analysis.** These approaches seem to prefer an excess due to MSPs (e.g. [F. Calore et al., PRL 127 (2021) 16]; [M. Buschmann et al. PRD 102 (2020) 2]; [R. Bartels et al., PRL 116 (2016) 5]).
- **Machine learning with convolutional networks could generalise over point source distribution as a generic feature and include uncertainties in astrophysical background modelling!**

How machine learning can help

Machine learning techniques have already been applied to the GeV excess.

— pioneering: [S. Caron et al., JCAP 05 (2018) 058] —> The work presented here extends this pioneering study.

— more recent ones: [F. List et al. PRD 104 (2021) 12]; [S. Mishra-Sharma and K. Cranmer, PRD 105 (2022) 6]

—> Results point rather to an admixture of both components.

Our approach:

Convolutional neural networks trained on images of the Galactic centre composed of background and signal templates to reconstruct the parameters of the GeV excess.

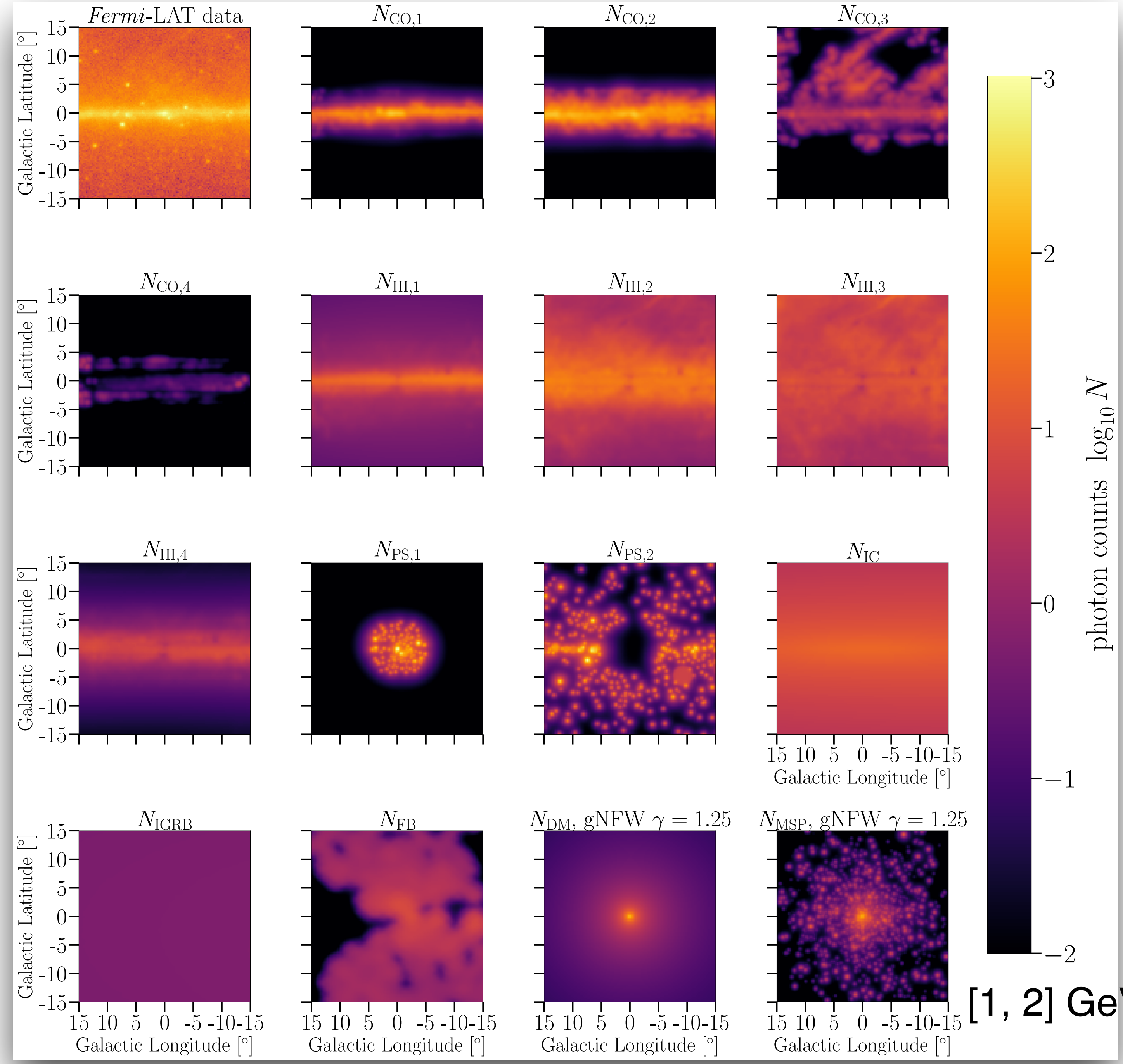
— 4 rings for diffuse background model (split into HI and CO contribution) + inverse-Compton (single) [Fermi collab., ApJ. Suppl. 224 (2016) 1];

— all 4FGL-DR2 point sources within 20°x20° region of interest (matches period selected for real data) [Fermi collab., ApJ.Suppl. 247 (2020) 1];

— Fermi Bubbles [Fermi collab. ApJ 840 (2017) 1];

— isotropic component

— GCE: smooth DM component + individually drawn MSPs both following a gNFW profile



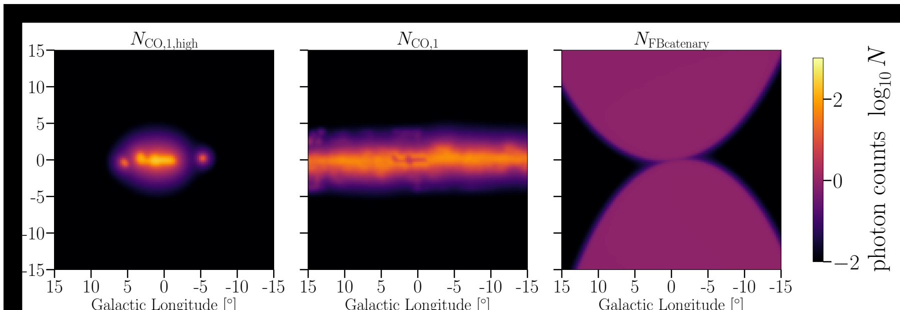
Neural network architecture and scope

Model setup to explore the impact of the **background model complexity** on the interpretation of the GCE with **Bayesian convolutional neural networks** used in a **DeepEnsembles** setup. We probe the '**reality gap**' — the discrepancy between modelled and real data.

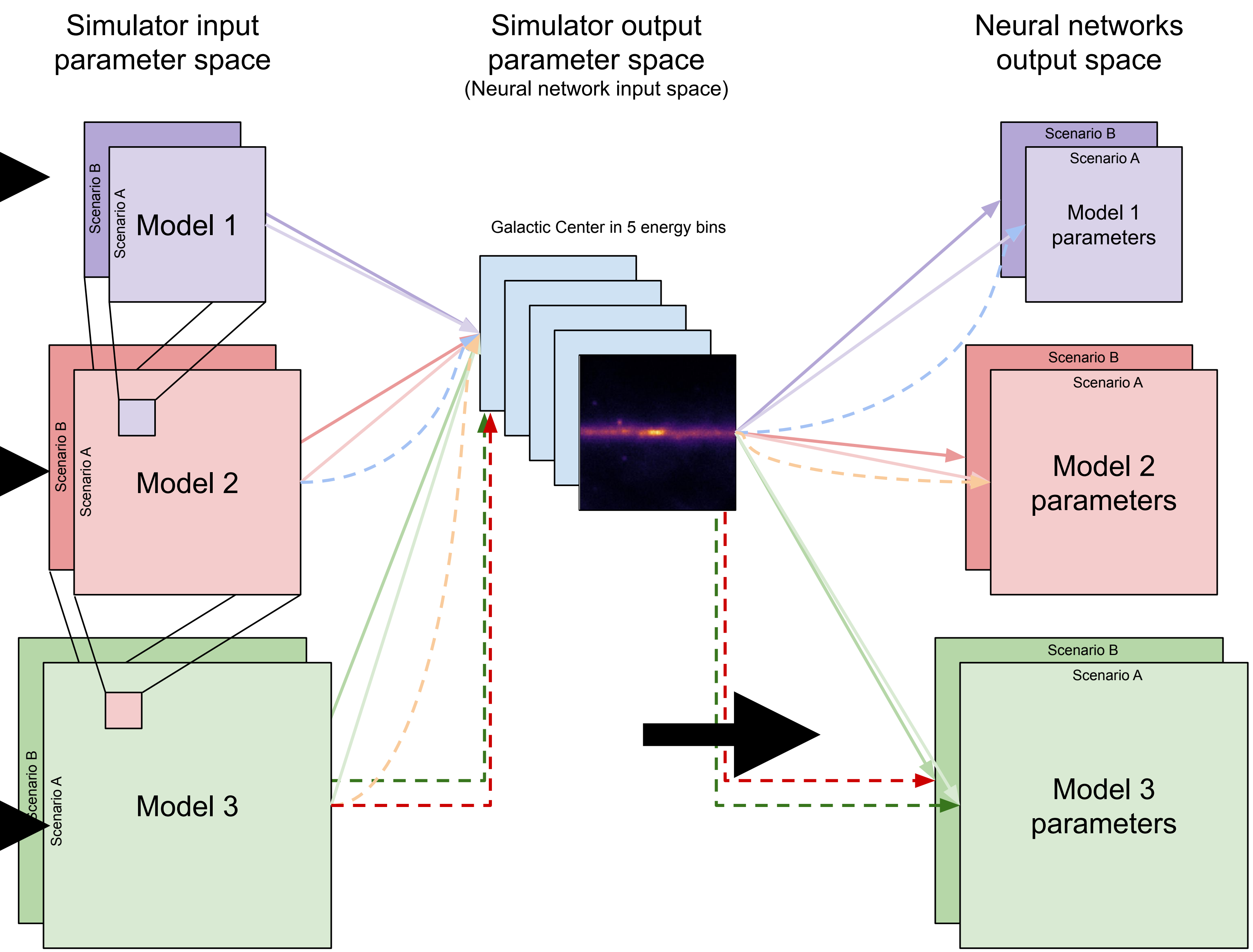
Reduce to 7 parameters, only one for astro diffuse emission (relative norm. of astro components fixed by likelihood fit)

compares to [S. Caron et al., JCAP 05 (2018) 058]

Initial 18 parameters



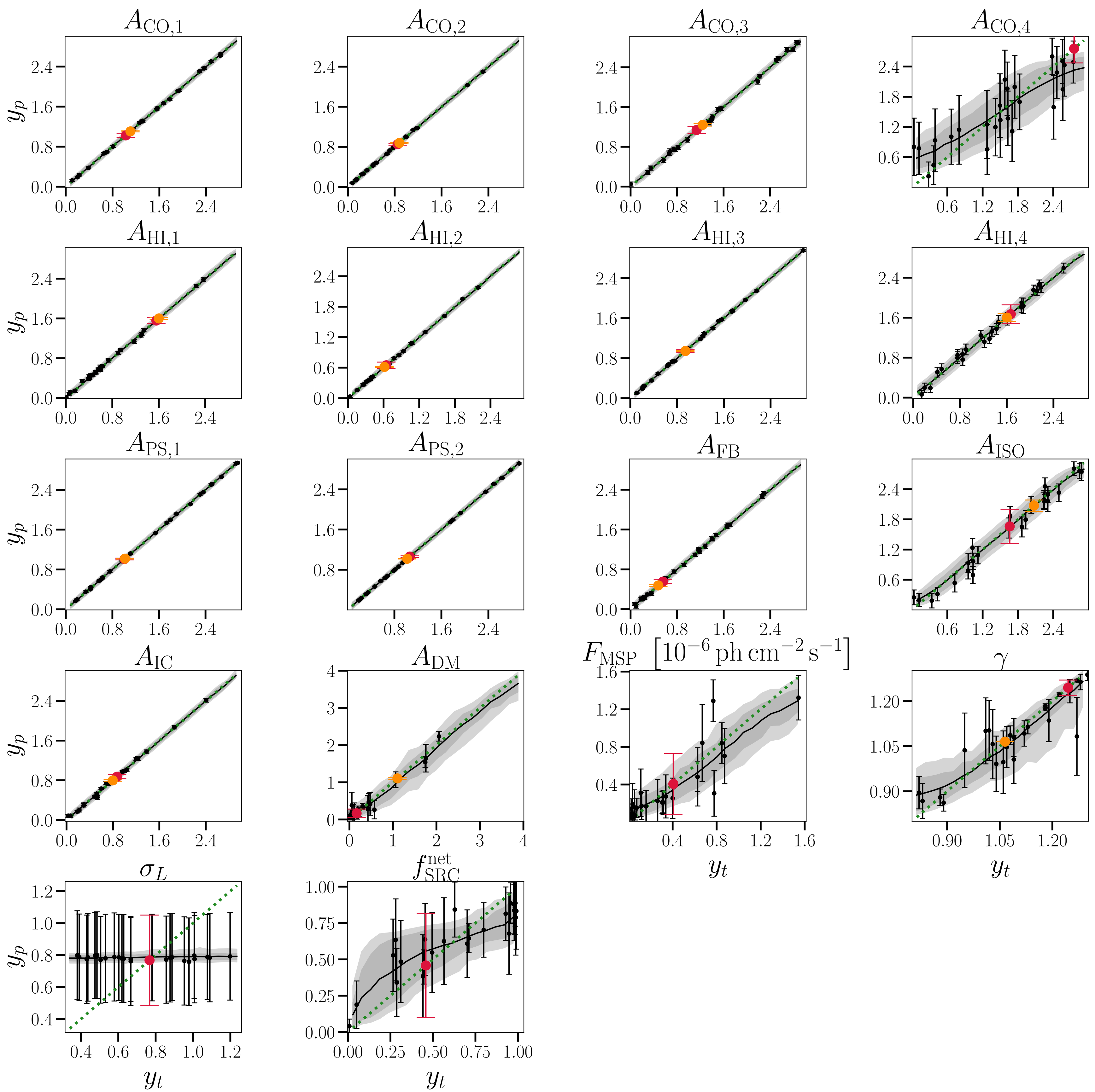
+ spectral freedom —> 25 parameters



scenario A: discard all MSPs above the Fermi LAT detection threshold

scenario B: keep all MSPs

Results on fiducial model (2)



Gray band: 1(2) σ scatter of the mean prediction in the validation dataset

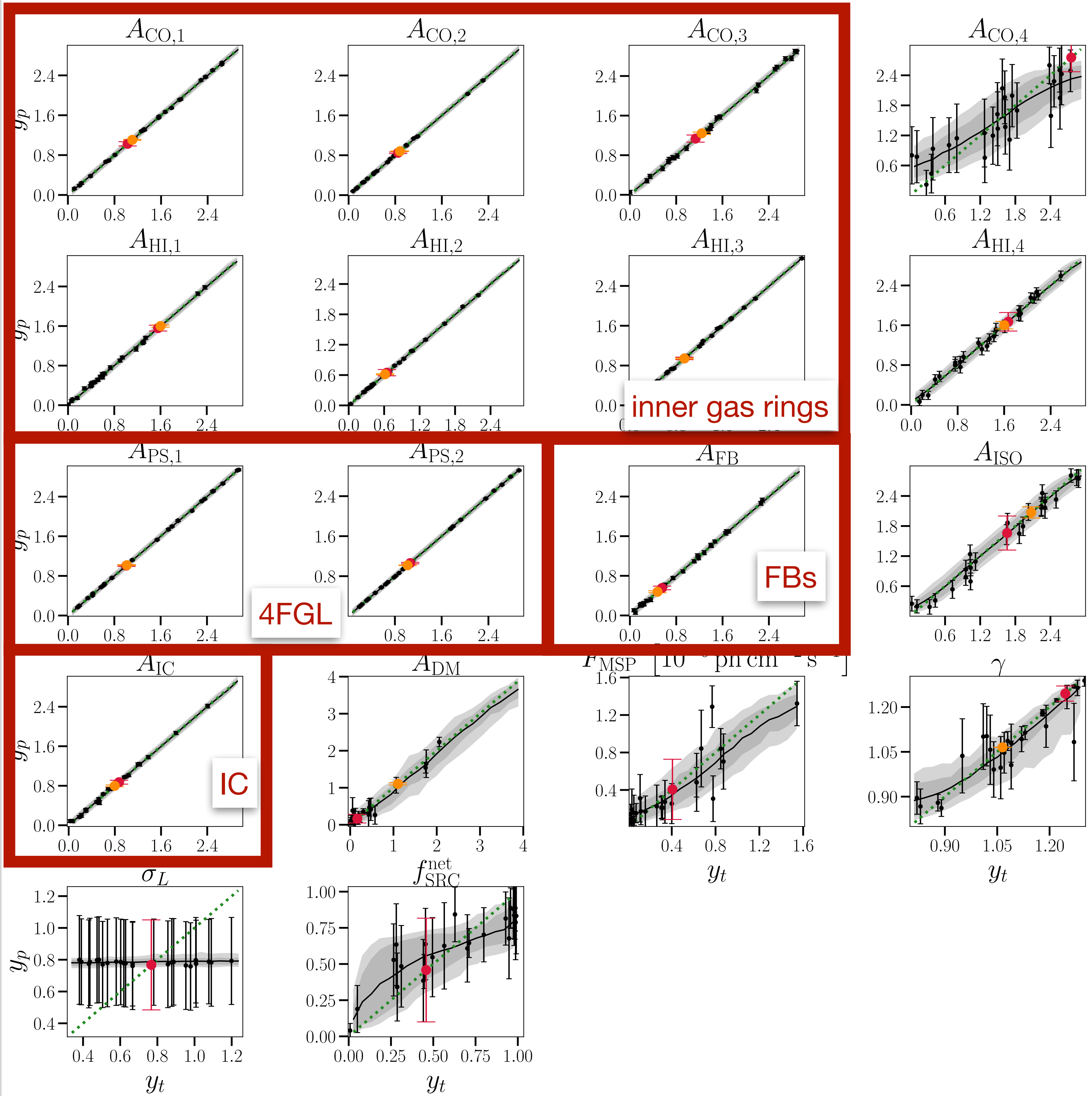
Error bars: network predicted uncertainty for selected points

Red point: results on real data
Orange point: results of maximum likelihood fit

X/Y-axis: True/predicted value (means should fall on the diagonal

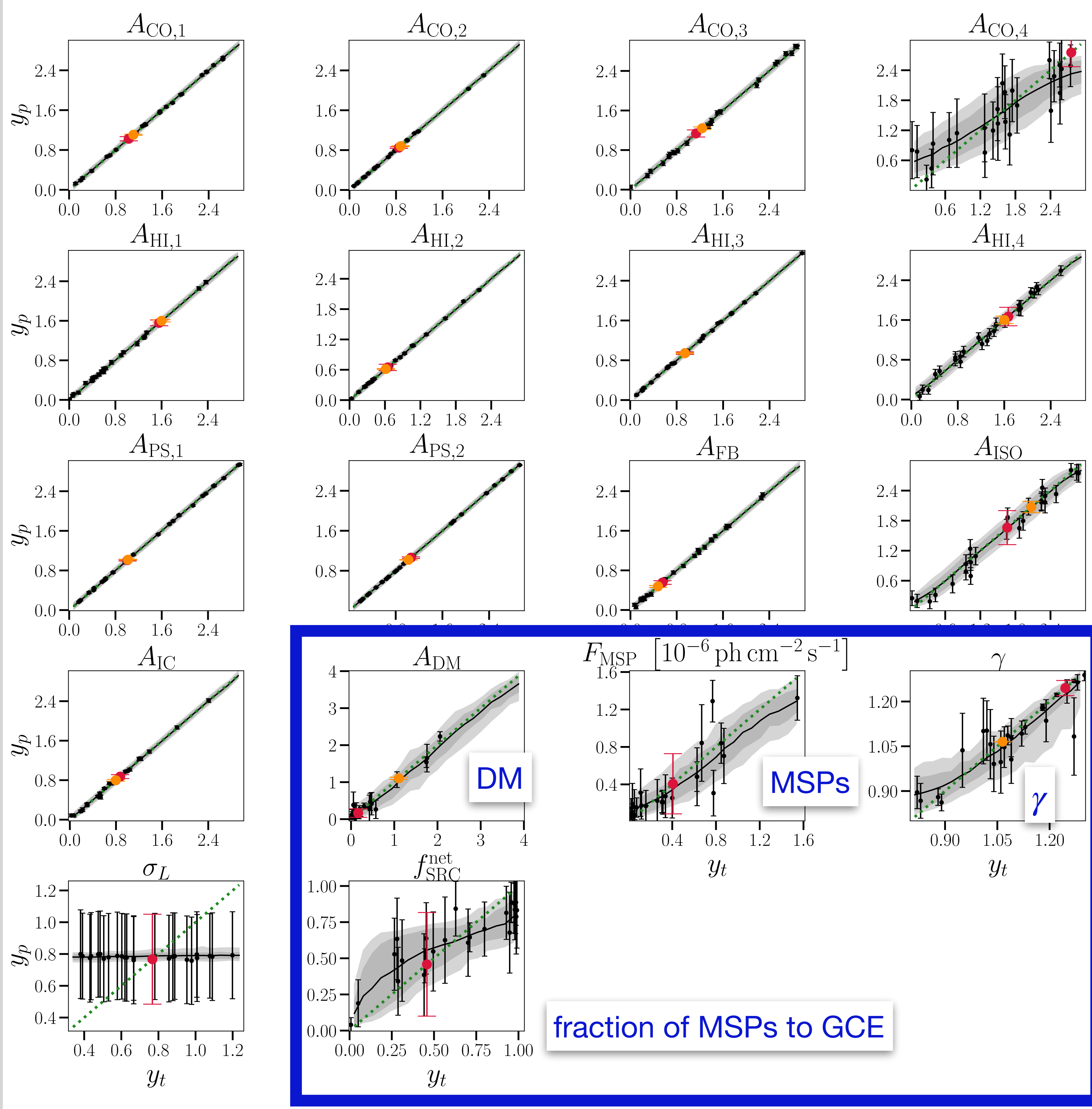
Results on fiducial model (2)

Bright components predicted with high precision and consistent with the traditional likelihood analysis.



Results on fiducial model (2)

GCE related components predicted robustly, though fSRC has some bias and significant uncertainty



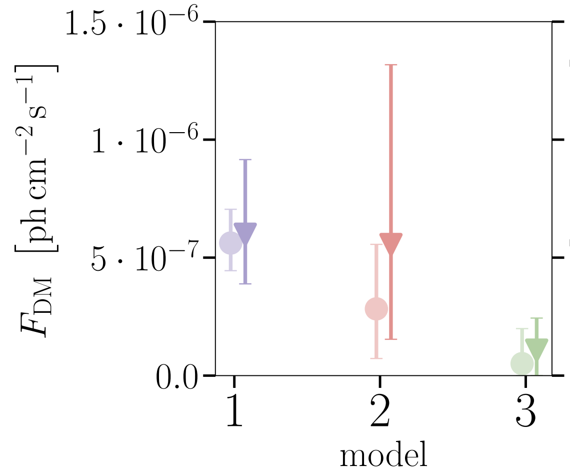
Evolution of results with model iteration

Predictions broadly consistent among the models (also conventional astro components),

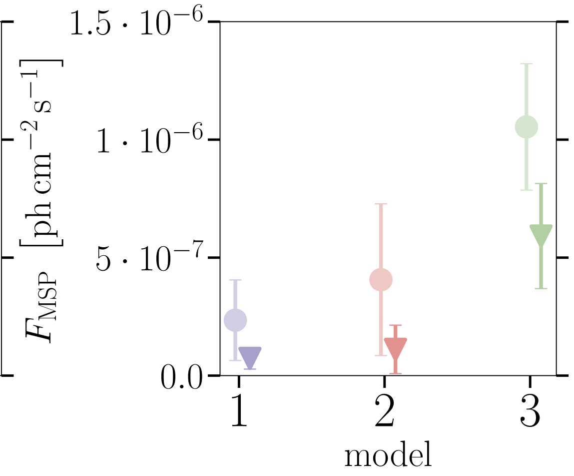
BUT there seems to be a trend towards lower values of DM and high values of MSP normalisation when increasing the complexity of the models.

Reducing the number of reconstructed parameters to one (marginalising over the remaining ones) yields very high precision.

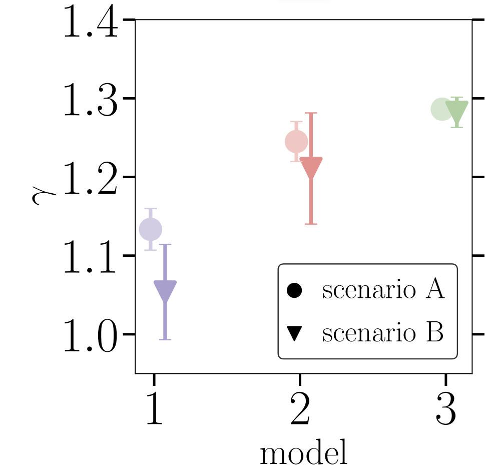
DM flux



MSP flux

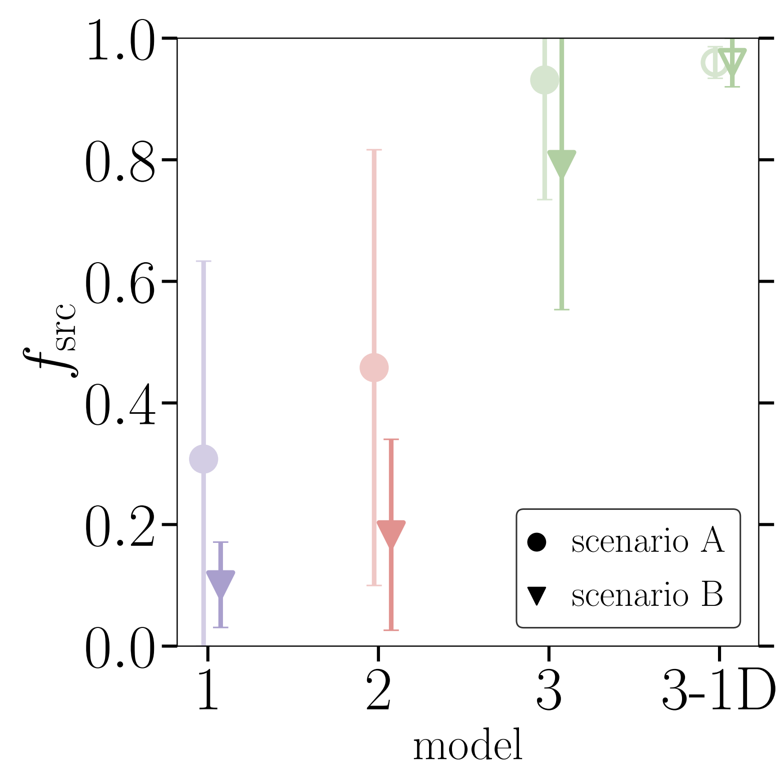


γ



$$f_{\rm src} = \frac{F_{\rm MSP}}{F_{\rm MSP} + F_{\rm DM}}$$

fraction of MSPs contribution to GCE

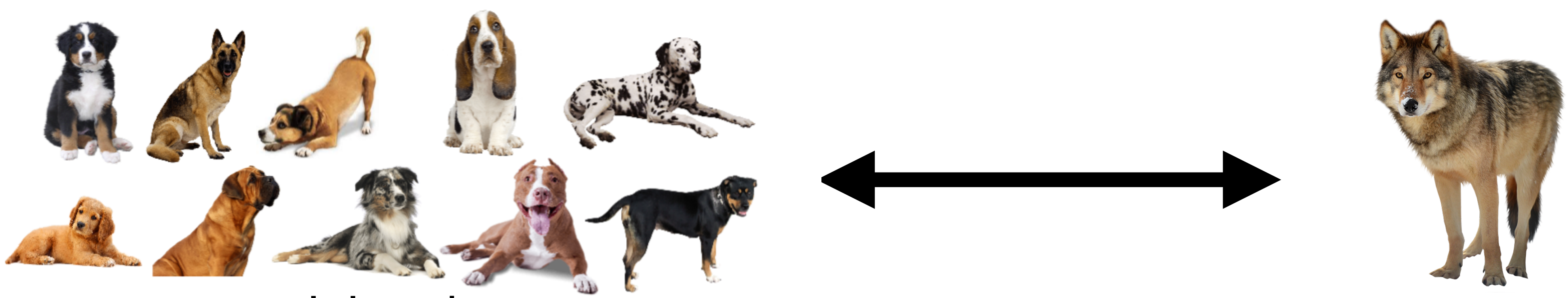


Mind the Gap

There is only one gamma-ray sky…
Machine learning (as any analysis) yields a result, but how to check if this particular result is robust?

+ Such checks especially relevant when using a new (black box?) tool

A likely situation for studies of the Galactic centre:



training data

real data

How far is a wolf from dogs in the world of the neural network?

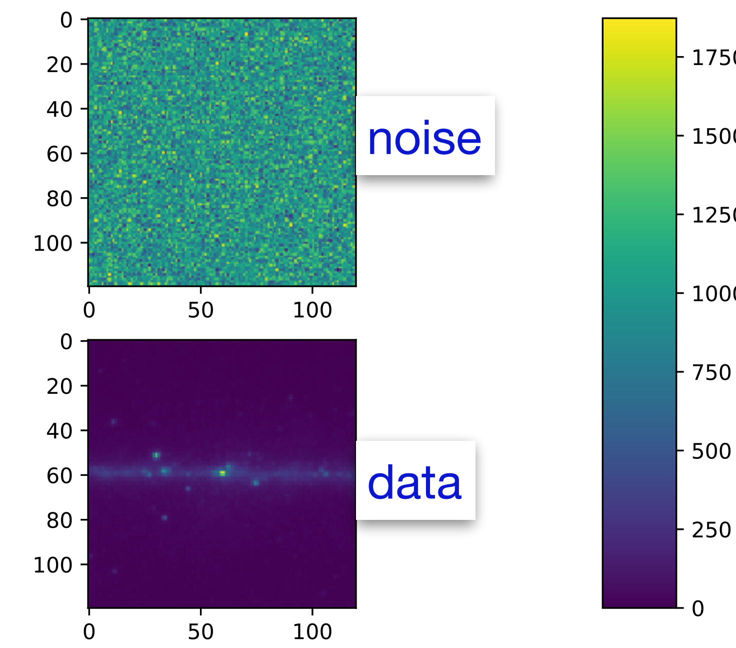
Machine learning technique called: **One-Class Deep Support Vector Data Description** (Deep SVDD) — **Anomaly detector**

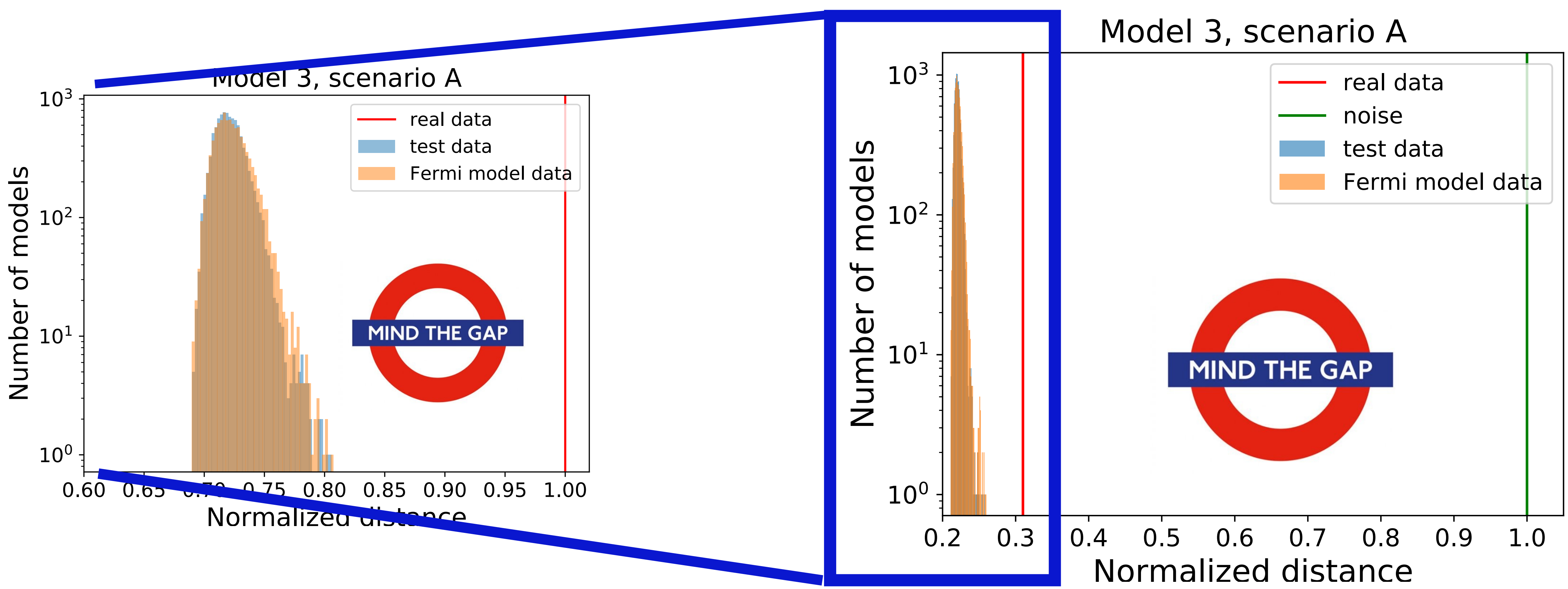
—> Classifies images as simulation-like or not
—> Relies on same network architecture as previous inference network except for the last layer: vector of fixed length with identical scalar value in all components

Mind the Gap — A means to quantify it

For Deep SVDD we stay in the framework of Model 3:

— A network trained on Model 3 is then shown validation data from: Model 3, the Fermi diffuse background model, real data and pure Gaussian noise.
— It quantifies the difference (*in latent space*) between what it 'knows' (its universe derived from the simulated training data) and what is shown to it.
— It answers the question: **Can this universe produce such an image?**





Christopher Eckner, eckner@lapth.cnrs.fr

Conclusions

DeepEnsemble Networks are capable of recovering the background and the presence of the GCE. We found that:

- Bright components are detected robustly and consistently between our models. They are also detected consistently with the prediction from the traditional likelihood method.

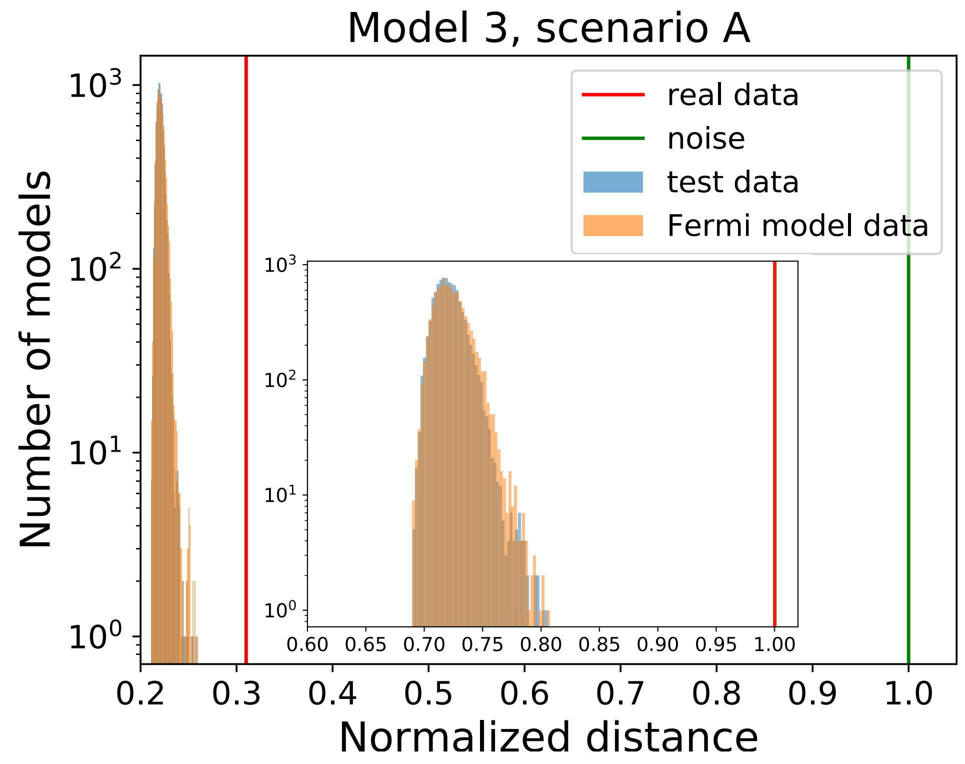
- The networks robustly detect the presence of the GCE in all our models, with the properties (flux and spatial distribution) consistent with other works.

However, the picture is not as clear as we (and everyone else!) wished:

- The nature of the GCE however, while well predicted within each model, **does not appear to be robust when networks are applied outside of their domain. We can predict anything from no DM to no MSPs by selecting a fitting background model.**

- ***Mind the gap: - the fact that reality is not part of the (background) model has been a limiting factor of many (all?) current works. What results can we trust at the moment?***

- **Deep SVDDs** offer a possibility to test severity of the reality gap. We are currently probing state-of-the-art models of the GC in this way. **Stay tuned!**



Backup slides

Neural network architecture and scope

Network architecture:

— **Bayesian convolutional networks** to determine the simulation parameters in an inference task
—> input: (5, 120, 120) images of the GC: 5 energy bins ([0.5, 1], [1, 2], [2, 7],[7, 20], >20 GeV)
—> predict every template parameter with uncertainty

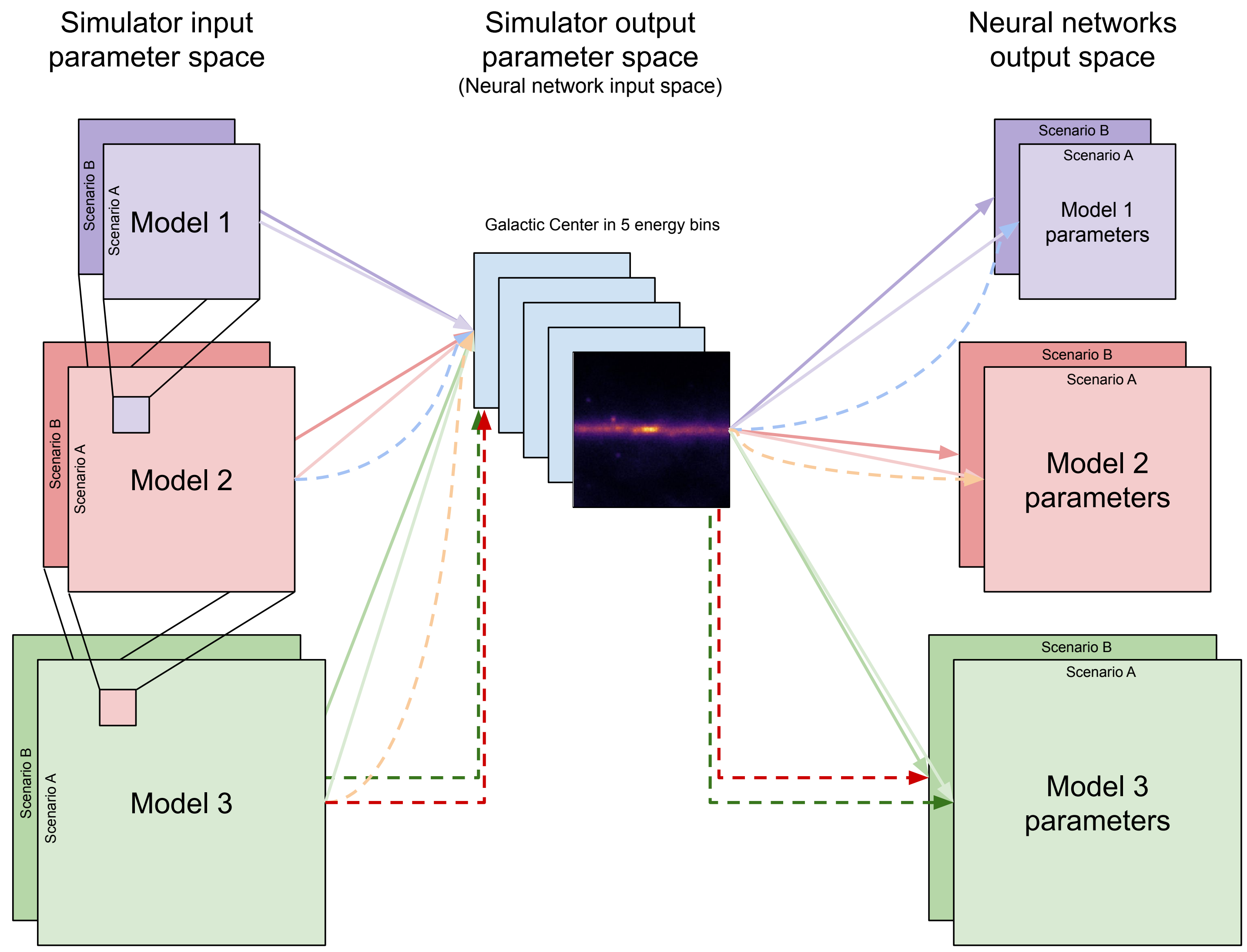
— **Deep ensemble networks** to also predict ***uncertainties due to the network itself***: What would have happened if the network was initialised differently (initial parameters, ordering of images, etc.)?
—> mean and scatter evaluated per network run and later combined

Additional uncertainties:

— DM and MSP templates follow the same spatial morphology. We are only interested in the fractional contribution of both components.

— The ‘**reality gap**’ — the discrepancy between modelled and real data — may dominate all reconstruction uncertainties!

— **Increase/decrease complexity of the model and check if results remain stable.**



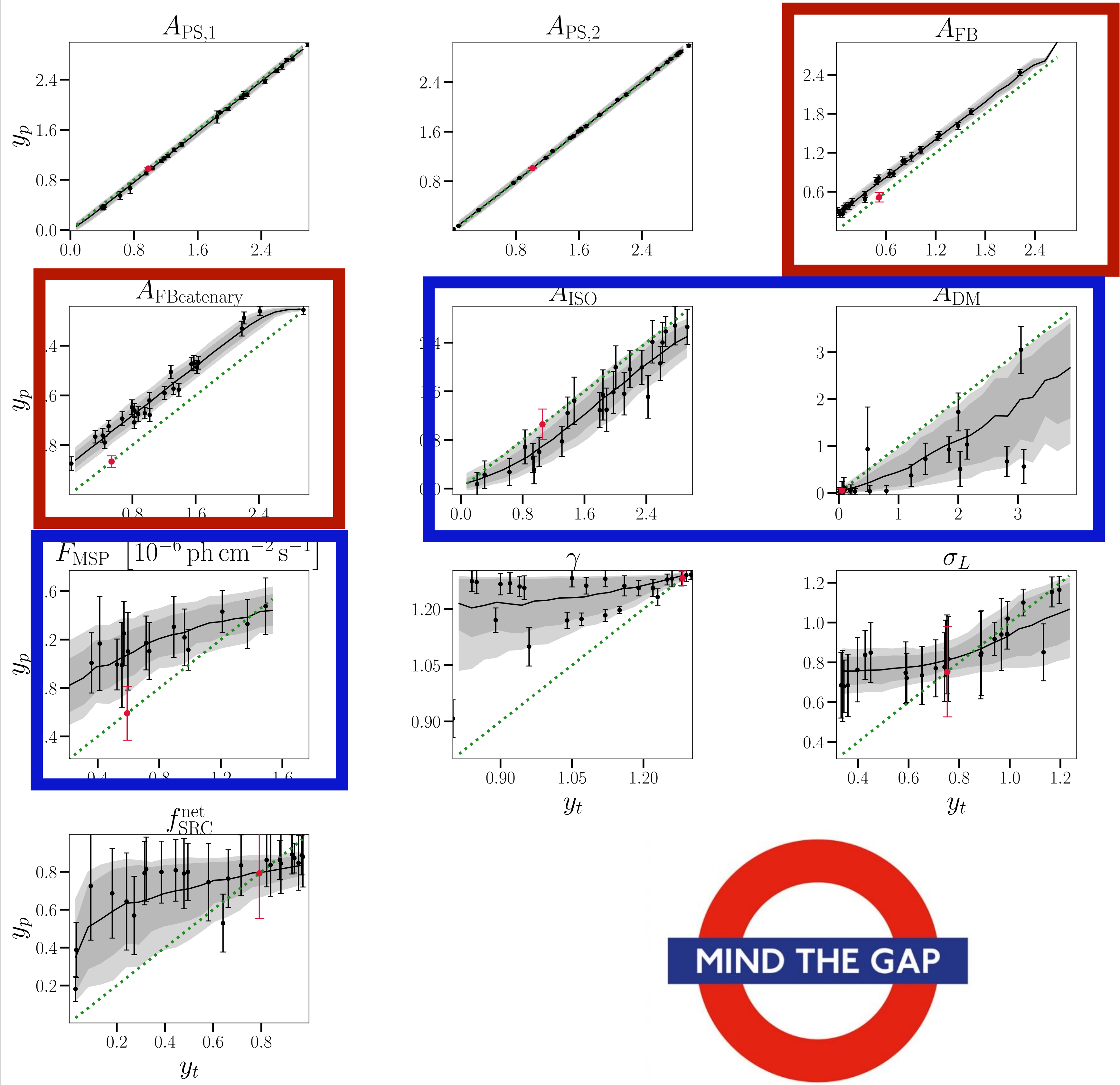
Mind the Gap — Fermi diffuse background model

Network trained on **Model 3A verified on the Fermi diffuse model**

FB templates are over predicted as expected

However, ISO and DM components are under predicted and (consequentially?) MSP template is over predicted

ILLUSTRATES THAT IN THE CASE THERE IS A ‘GAP’ BETWEEN THE TRAINING AND VERIFICATION DATA, RESULTS ARE UNRELIABLE!



MIND THE GAP