UNIVERSITÀ DEGLI STUDI
DI GENOVA

INFN GENOVA
Istituto Nazionale di Fisica Nucleare
Sezione di Genova

Parametrising profiled likelihoods with neural networks. (work in progress)

Humberto Reyes-González
University of Genoa

LHC reinterpretation workshop
Durham, 1/9/2023

Introduction

- **Likelihood functions (full statistical models)** parametrise the full information of an LHC analysis; wether it is New Physics (NP) search or an SM measurement.
- Their **preservation** is a key part of the **LHC legacy.**
- **Usage:** Resampling, Reinterpretation in the context of different NP models and/or with with different statistical approaches,….

A brief story on full statistical model publication and usage:

- ATLAS started publishing full likelihoods of NP searches (2019) ATL-PHYS-PUB-2019-029.
- Release of the pyhf package to construct statistical models (2020) 10.21105/joss.02823, L Heinrich, M Feickert, G Stark
- Interface with reinterpretation tools: SmodelS (2020) arXiv:2009.01809, MadAnalysis (2022) arXiv:2206.14870,
- Spey: Generalised framework for likelihood handling (2023) arXiv:2307.06996. ⟵ see Jack's talk

Why (Machine) Learning Profiled Likelihoods?

- In LHC-reinterpretation, to exclude a BSM model, we are mostly interested in the profiled likelihood given a signal strength.
- Optimally, we can compute the profiled likelihood from pyhf's full statistical models.
- However, this computation can take the order of seconds by parameter point.
- A pheno study may require to survey thousands of points.
- This considerably scales-up the time consumption. Specially for fast reinterpretation approaches.
- **Using Neural Networks provides a fast and compact way using profiled likelihoods in our day-to-day pheno studies.**
- **We will super useful when launching a new protomodel-based anomaly search. (arXiv:2105.09020)**

LHC likelihoods in a nutshell

Bayes theorem:

$$P(\Theta, x) = P_x(x\,|\,\Theta)\pi_\Theta(\Theta) = P_\Theta(\Theta\,|\,x)\pi_x(x)$$

LHC Statistical model:

$$P(\boxed{\mu}, \boxed{\theta}; \text{data}) = \Pi_{k=1}^{n_c} P[\boxed{n_i}; \boxed{\mu}\epsilon_{i,k}\boxed{(\overrightarrow{\theta})}N_{S,i,k}\boxed{(\overrightarrow{\theta})} + B_{i.k}\boxed{(\overrightarrow{\theta})}]\Pi_{j=1}^{n_{syst}} G(\boxed{\theta_j^{obs}}; \boxed{\theta_j; 1})$$

(Observed) data

Nuisance parameters (uncertainties)

(Auxiliary) data

Parameters of Interest (signal strength, observables, etc.)

With this we perform global fits, exclude BSM models, find upper limits, search for SM deviations, etc.

The importance of the Profile Likelihood

The Profile Likelihood (PL) is a function where the nuisance parameters are fixed such that the likelihood is maximised given a signal strength μ. So the PL is a function of x, here on n_s.

$$L(x \mid \mu; \hat{\theta}(\mu))$$

With the PL we construct Log Likelihood Ratio (LLR) tests. Depending on how the PL is fitted and defined, we can derive upper limits, exclusion confidence levels and discoveries.

$$t(\mu) = -2 \log \frac{L(\mu; \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta}(\hat{\mu}))}$$

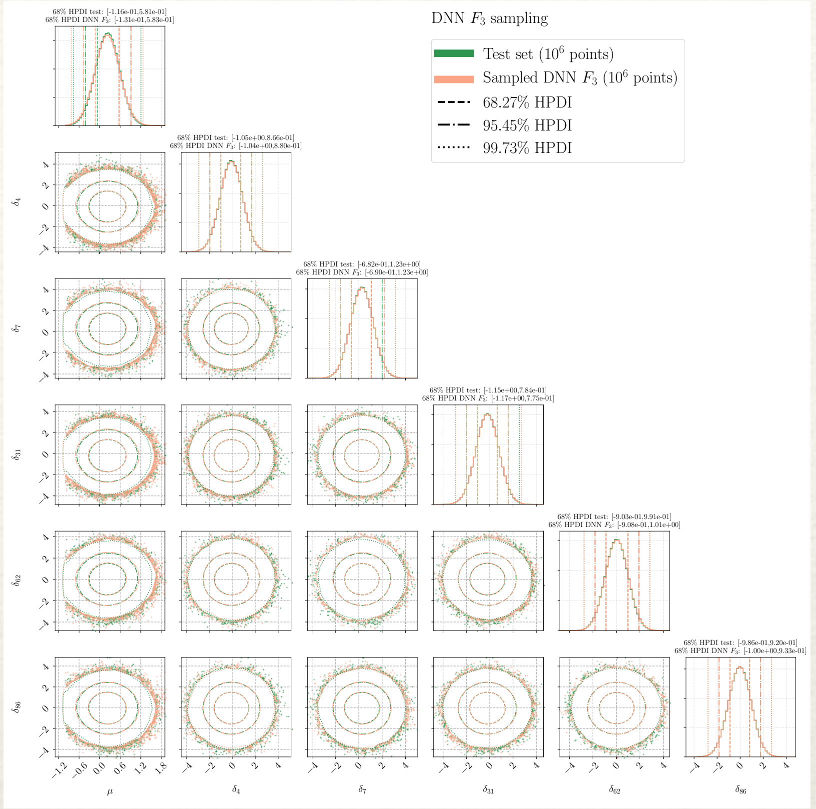
Note that to derive p-values of the test statistics we need to know their distributions. For this you may need the **Asimov Likelihood**. This is derived from the Asimov data set, defined in such a way that "when one uses it to evaluate the estimators for all parameters, one obtains the true parameter values".

(see arXiv:1007.1727)

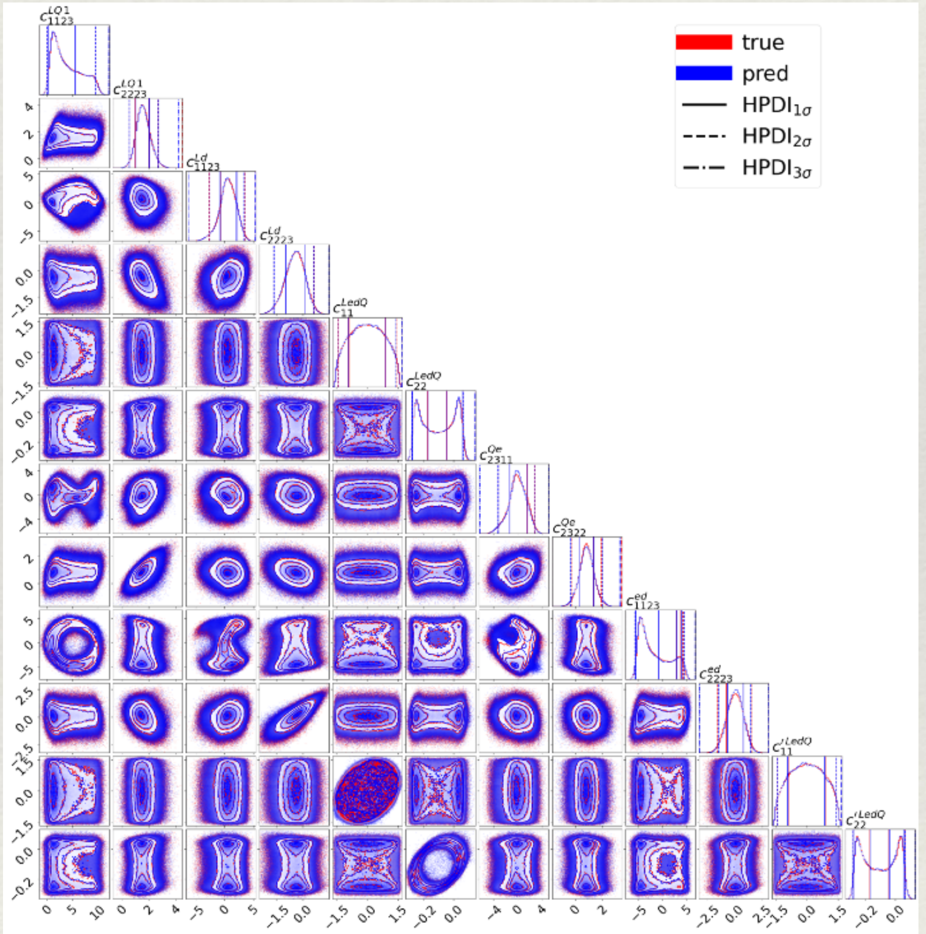
We plan to learn Likelihood Observed,Excluded (fit to $\mu = 1{,}0$) and the corresponding Asimov Likelihoods.

Previous work on learning LHC Likelihoods

$$P_\Theta(\Theta \,|\, x = \text{obs})$$



- **DNNLikelihood**
- Supervised Learning with Deep Neural Networks.
- arXiv:1911.03305 (A. Coccaro, M. Pierini, L. Silvestrini, R. Torre)



- **NFLikelihood.**
- Unsupervised Learning with Normalising Flows.
- See my talk from last RiF: (https://indico.cern.ch/event/1197680/timetable/#13-machine-learning-lhc-likeli)
- Paper to appear soon. (H.R-G., R.Torre)

Also, remember Nathan's talk yesterday on learning profiled EFT analyses!

Example Likelihoods

ATLAS-SUSY-2018-04

- Search for direct stau production in events with two τ-leptons
- Number of SRs: 2
- DOI: 10.1103/PhysRevD.101.032009

ATLAS-SUSY-2018-31 (A,B,C)

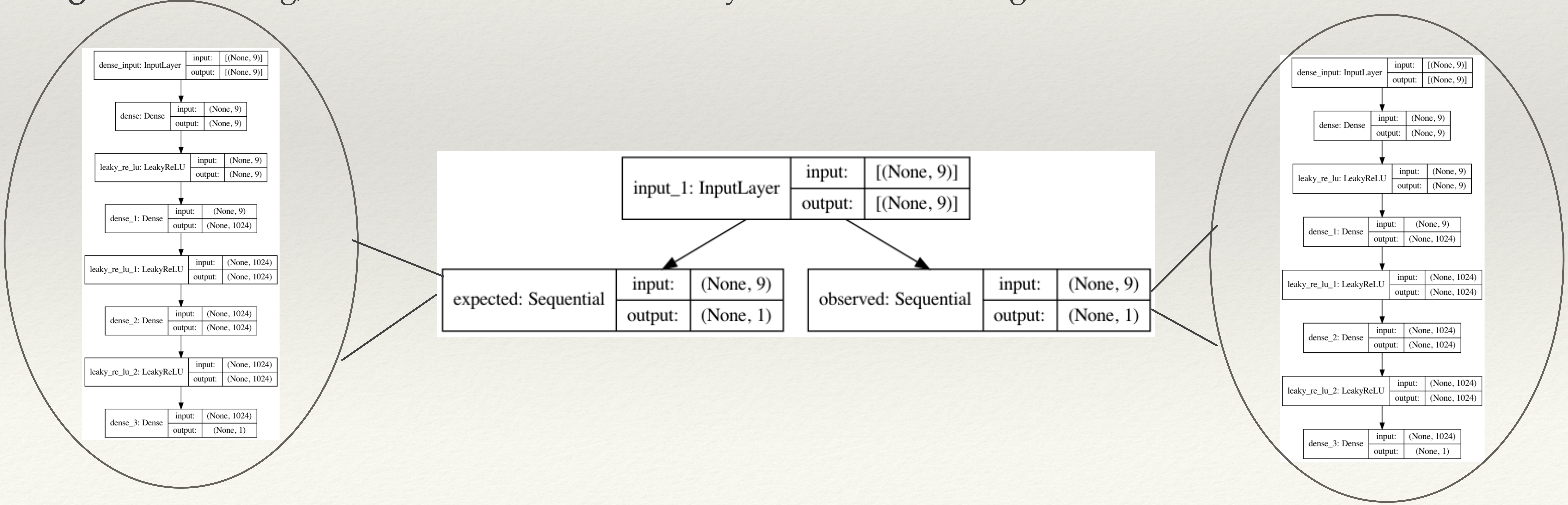
- Search for bottom-squark pair production in final states containing Higgs , b-jets and MET
- Divided into 3 subregions (A,B,C).
- Number of SRs: A: 3,B: 1,C: 4.

ATLAS-SUSY-2019-08

- Search for direct production of e-winos in final states with 1 lepton, MET and a Higgs boson decaying into 2 -jets
- Number of SRs: 9.
- DOI: https://doi.org/10.17182/hepdata.90607.v4

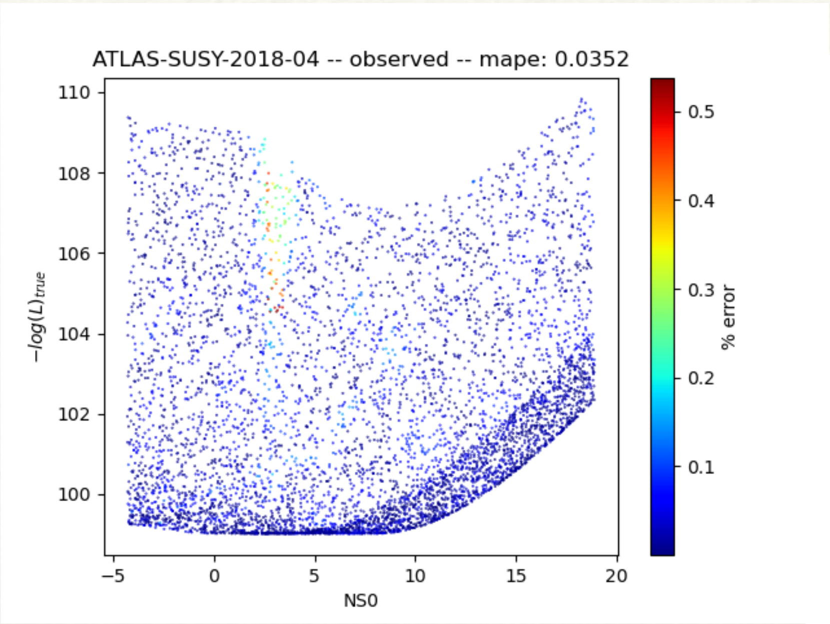
Training strategy.

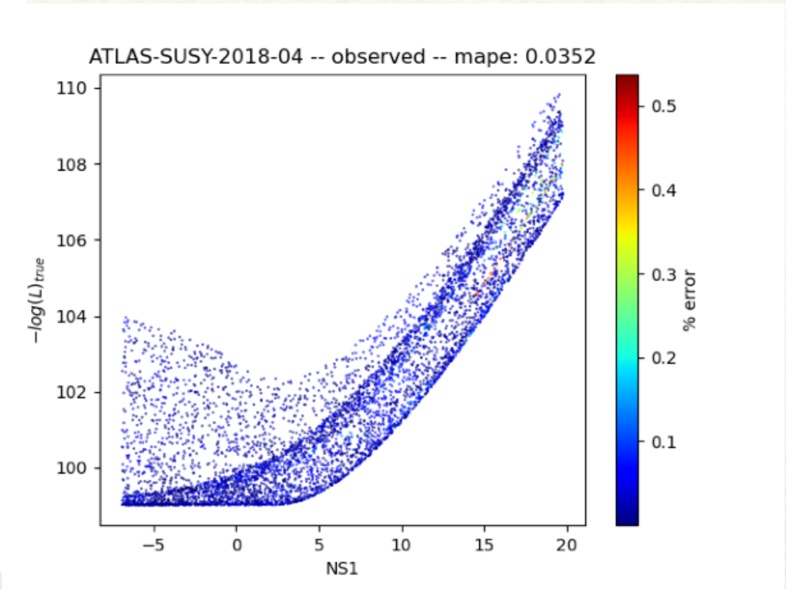
- **Sampling**: MCMC Metropolis-Hasting towards the min and max, to cover the full parameter space. The data was generated using an pyhf-SModelS interface. The Asimov Likelihood data was obtained with spey.
- The **Input** is n_s and the **Output** $-\ln(L)$.
- **Training:** All models were Multi-Layer Perceptrons (MPE) trained using Mean Squared Error loss function, ADAM optimiser and LeakyReLU activation functions. Data was divided as training-validation-test on a 60-20-20 scheme.
- **Testing:** The accuracy of the NN models was measured with the Mean (MAPE) and Max (MaxAPE) Absolute Percentage Error.
- **Saving:** After training, the best models for each analysis are ensemble together and saved as ONNX files.

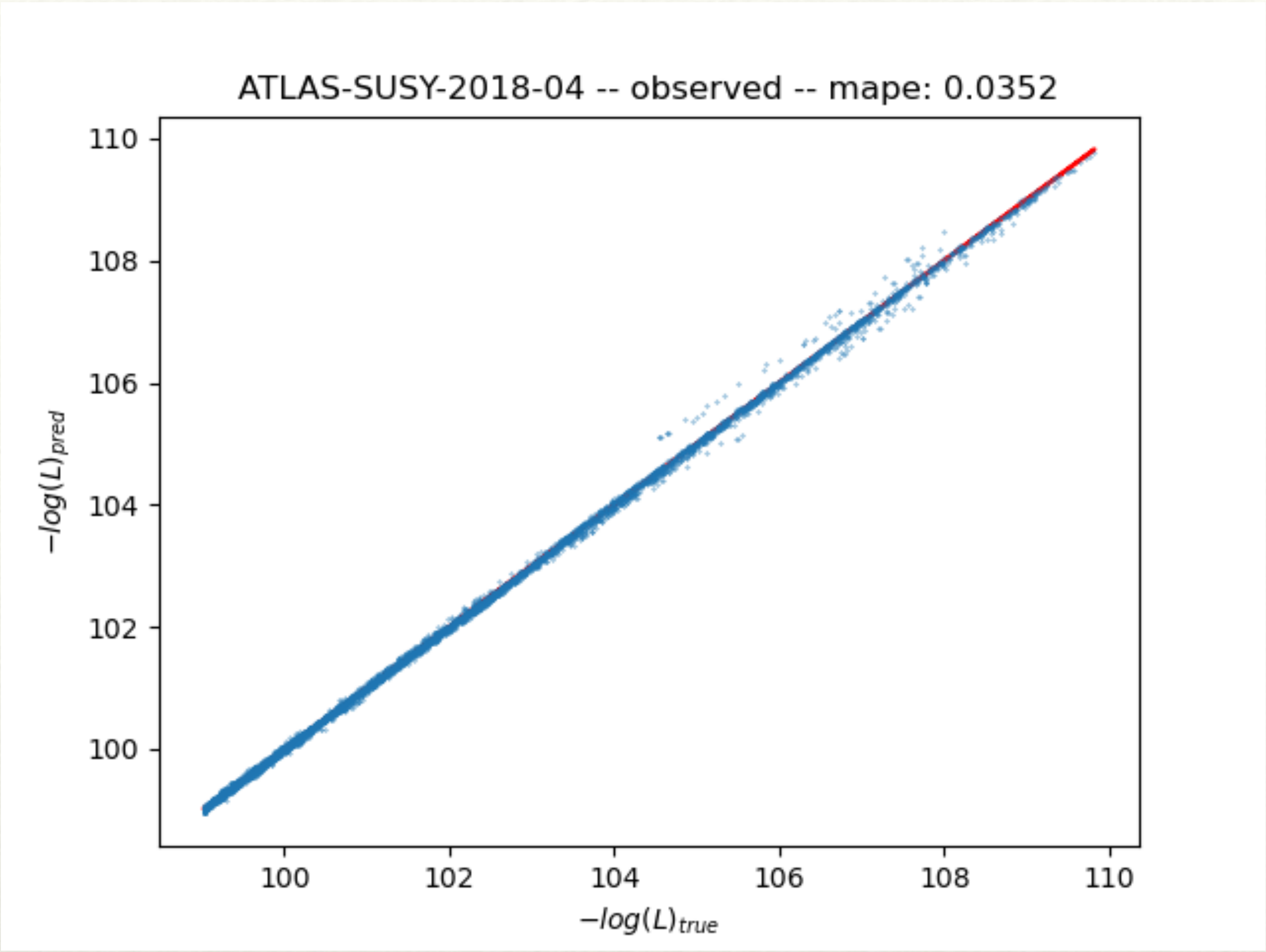


ATLAS-SUSY-2018-04, 2 SRS

OBSERVED



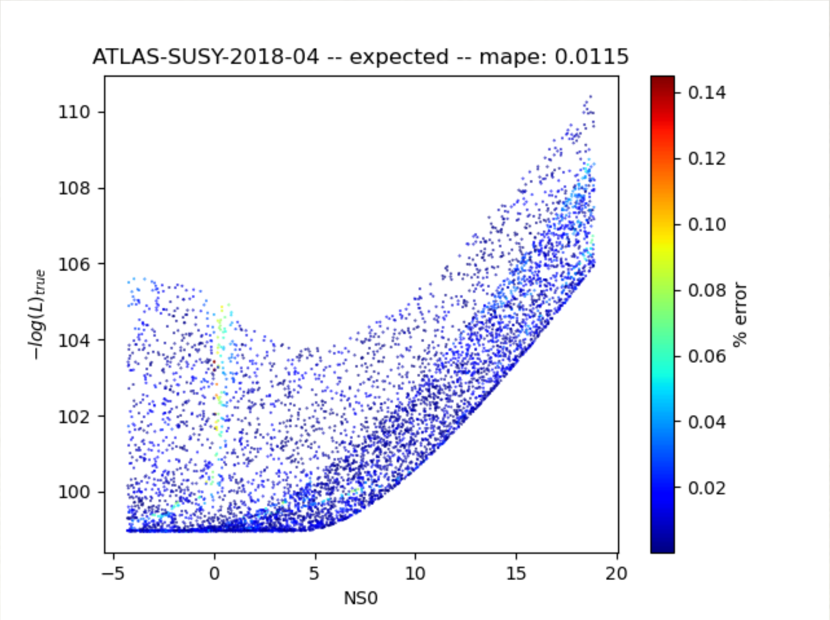


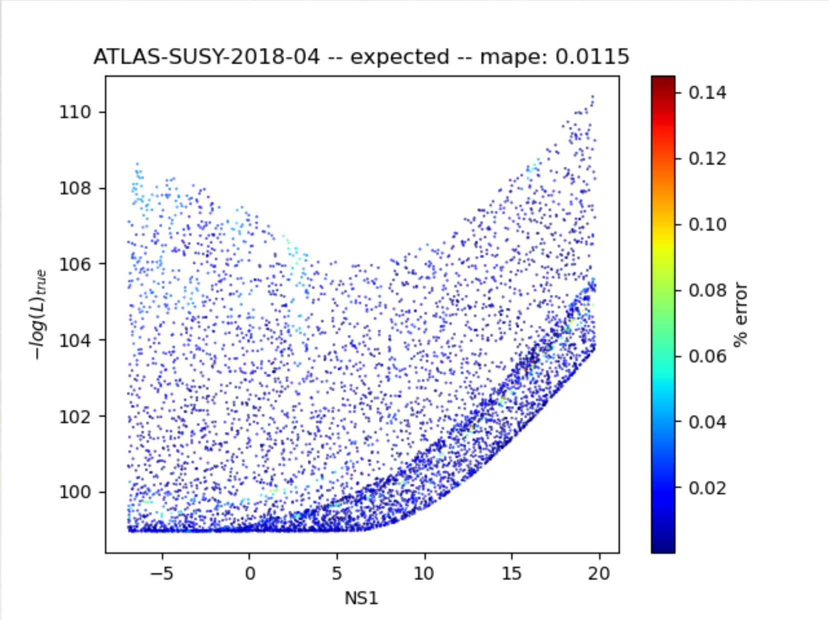


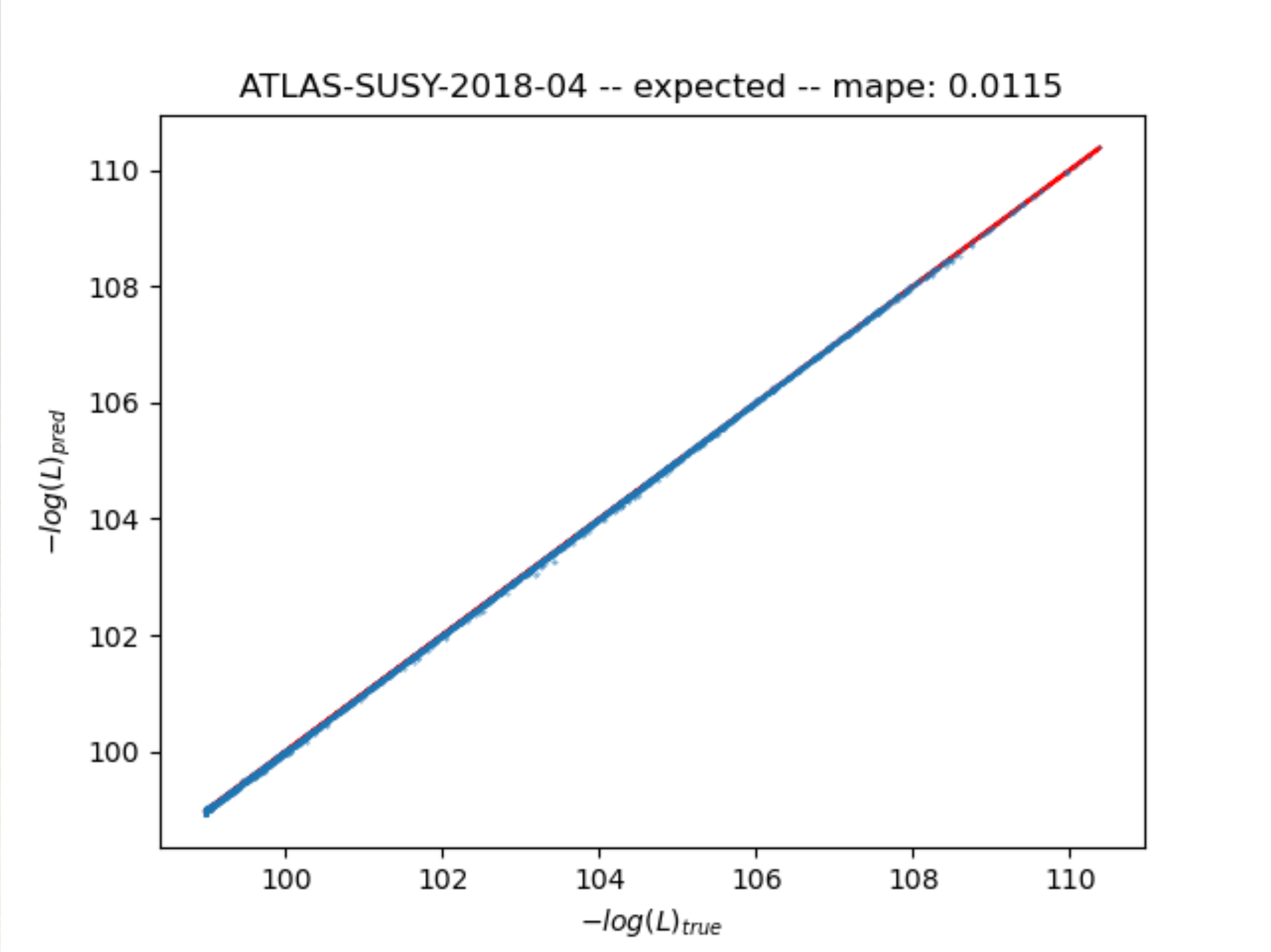
HYPERPARAMS	
Hidden layers	3X128
L2 regularisation	10^-5
Max epochs	300 X 4
Batch size	32
N samples	30k

METRICS
MAPE
0.0352
MaxAPE
0.5379

EXPECTED





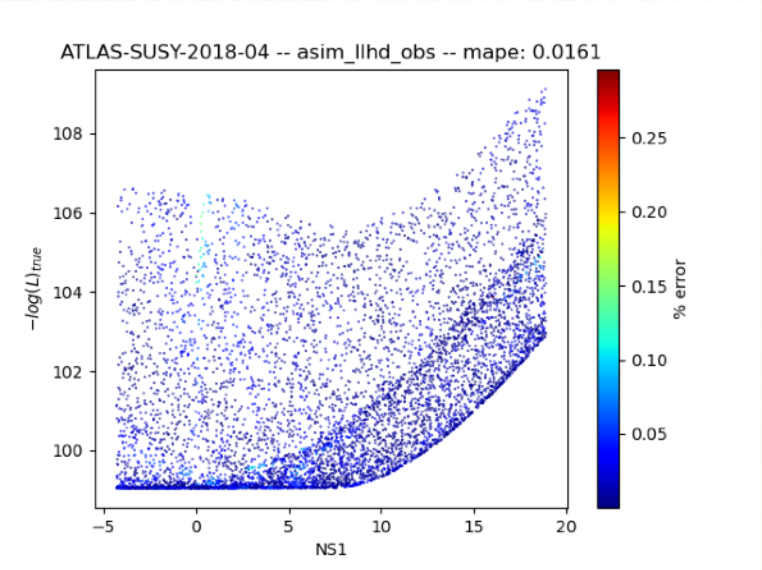


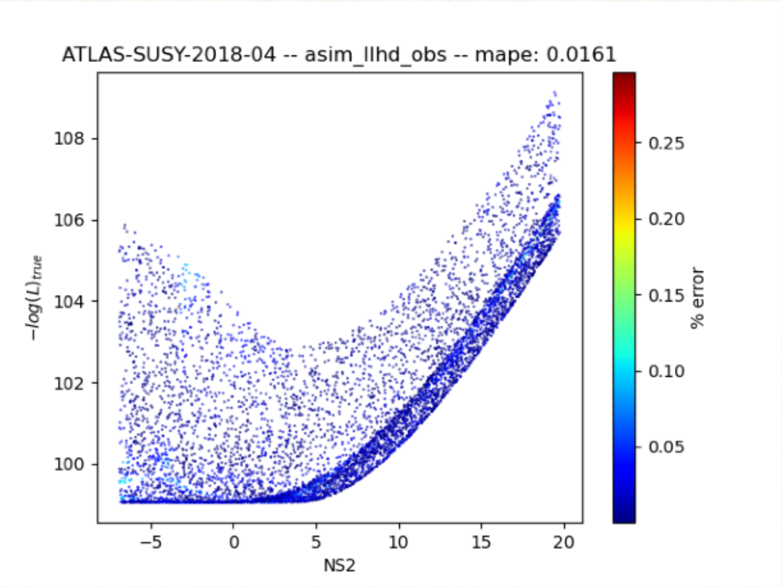
HYPERPARAMS	
Hidden layers	3X256
L2 regularisation	10^-5
Max epochs	300 X 4
Batch size	32
N samples	30k

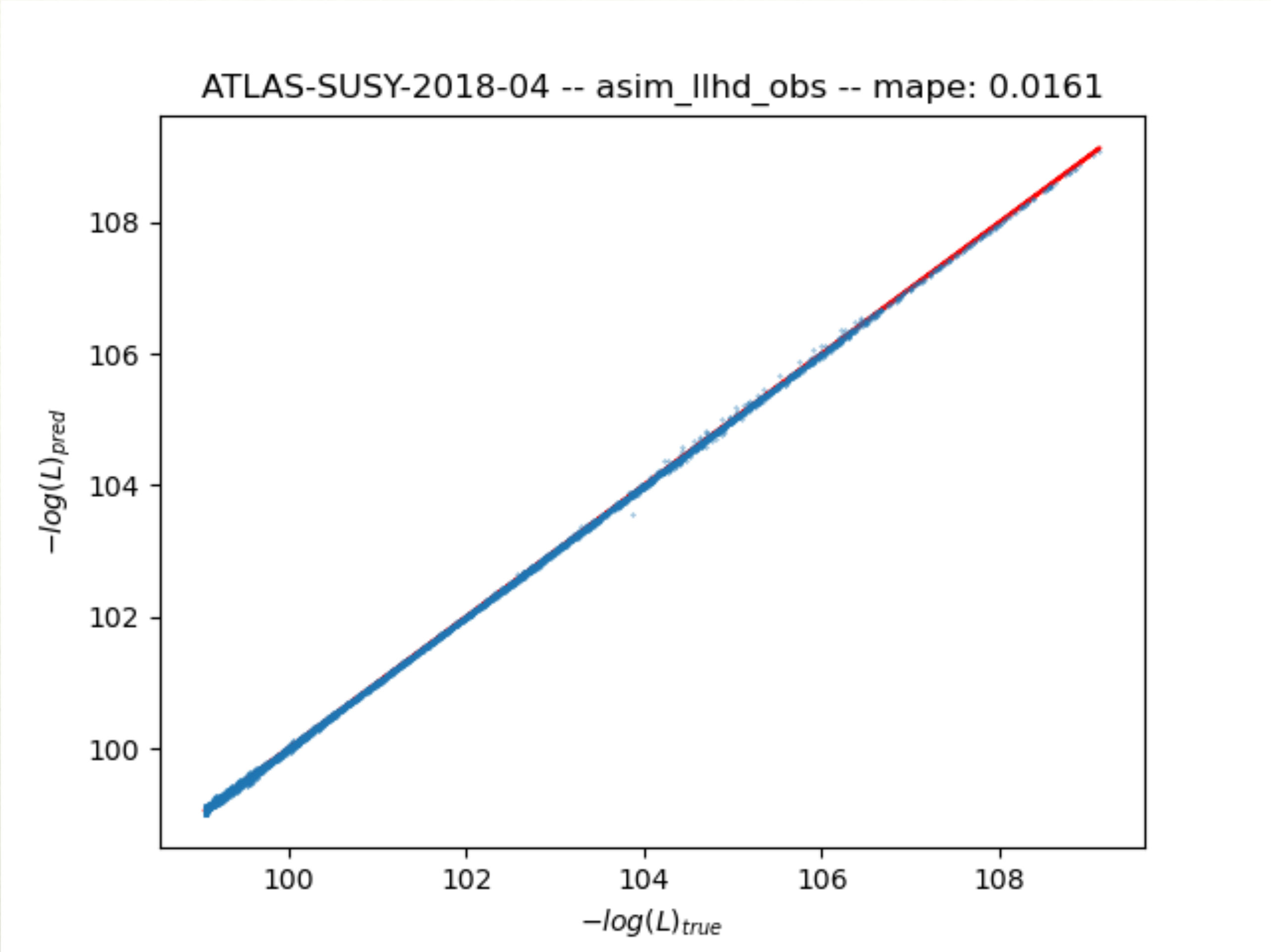
METRICS
MAPE
0.01154
MaxAPE
.14525

ATLAS-SUSY-2018-04, 2 SRS

ASIMOV OBSERVED



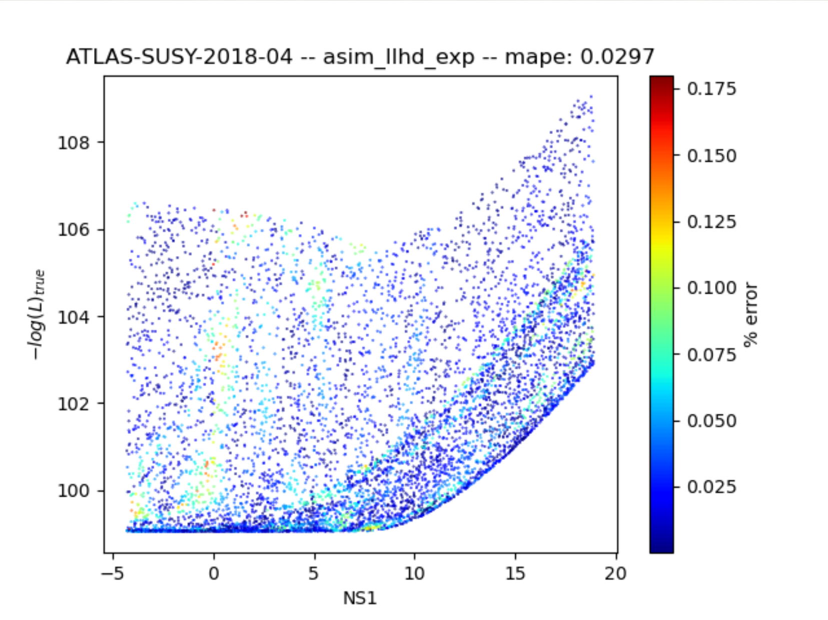


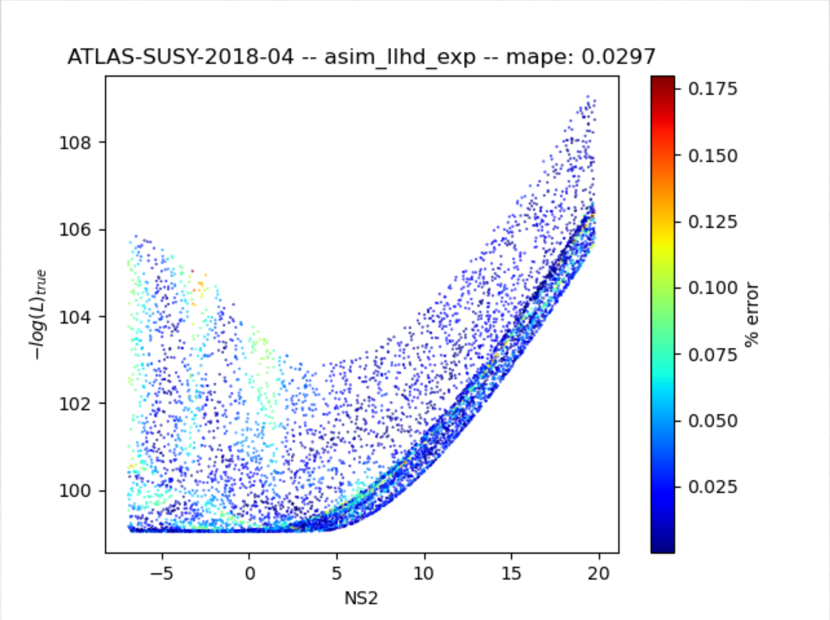


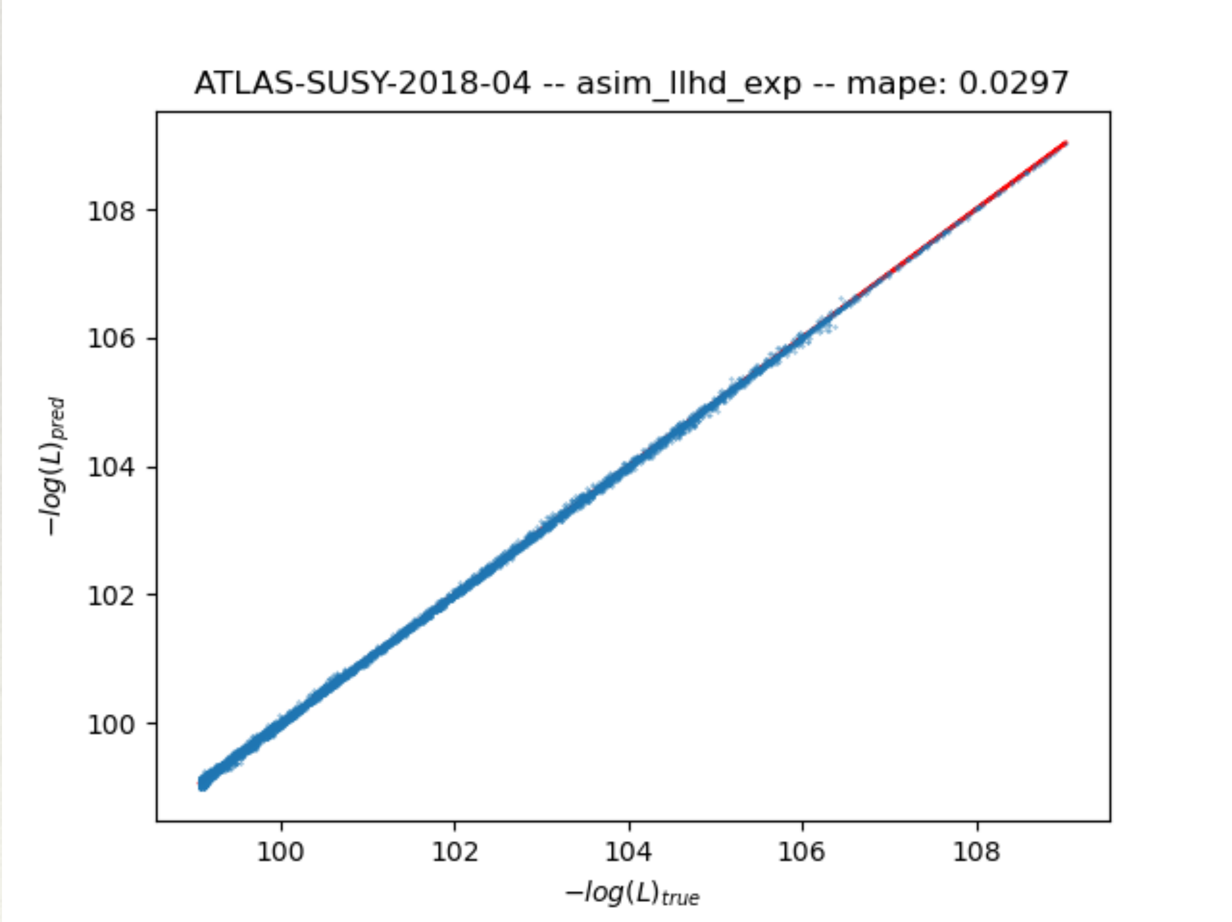
HYPERPARAMS	
Hidden layers	3X128
L2 regularisation	0
Max epochs	200 X 4
Batch size	64
N samples	12k

METRICS
MAPE
0.0161
MaxAPE
0.2964

ASIMOV EXPECTED





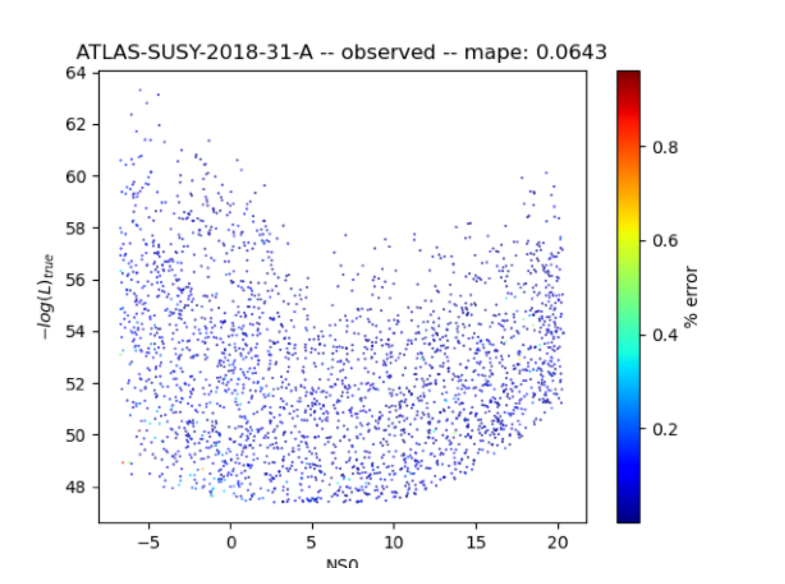


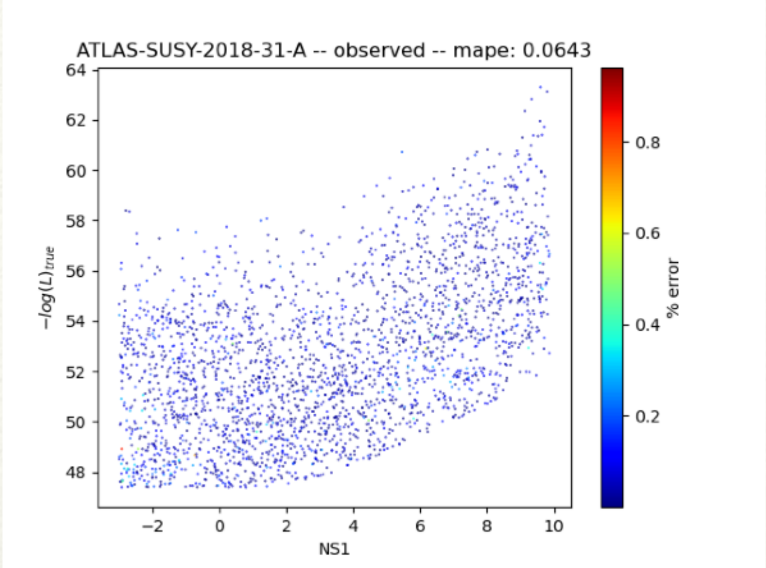
HYPERPARAMS	
Hidden layers	3X128
L2 regularisation	0
Max epochs	200 X 4
Batch size	64
N samples	20k

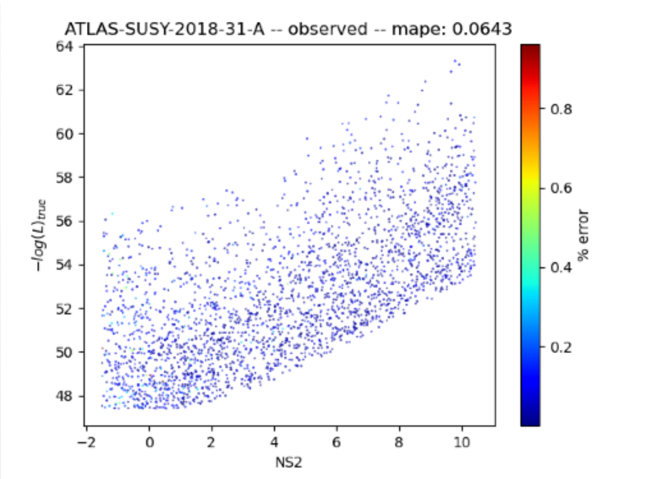
METRICS
MAPE
0.0297
MaxAPE
.180

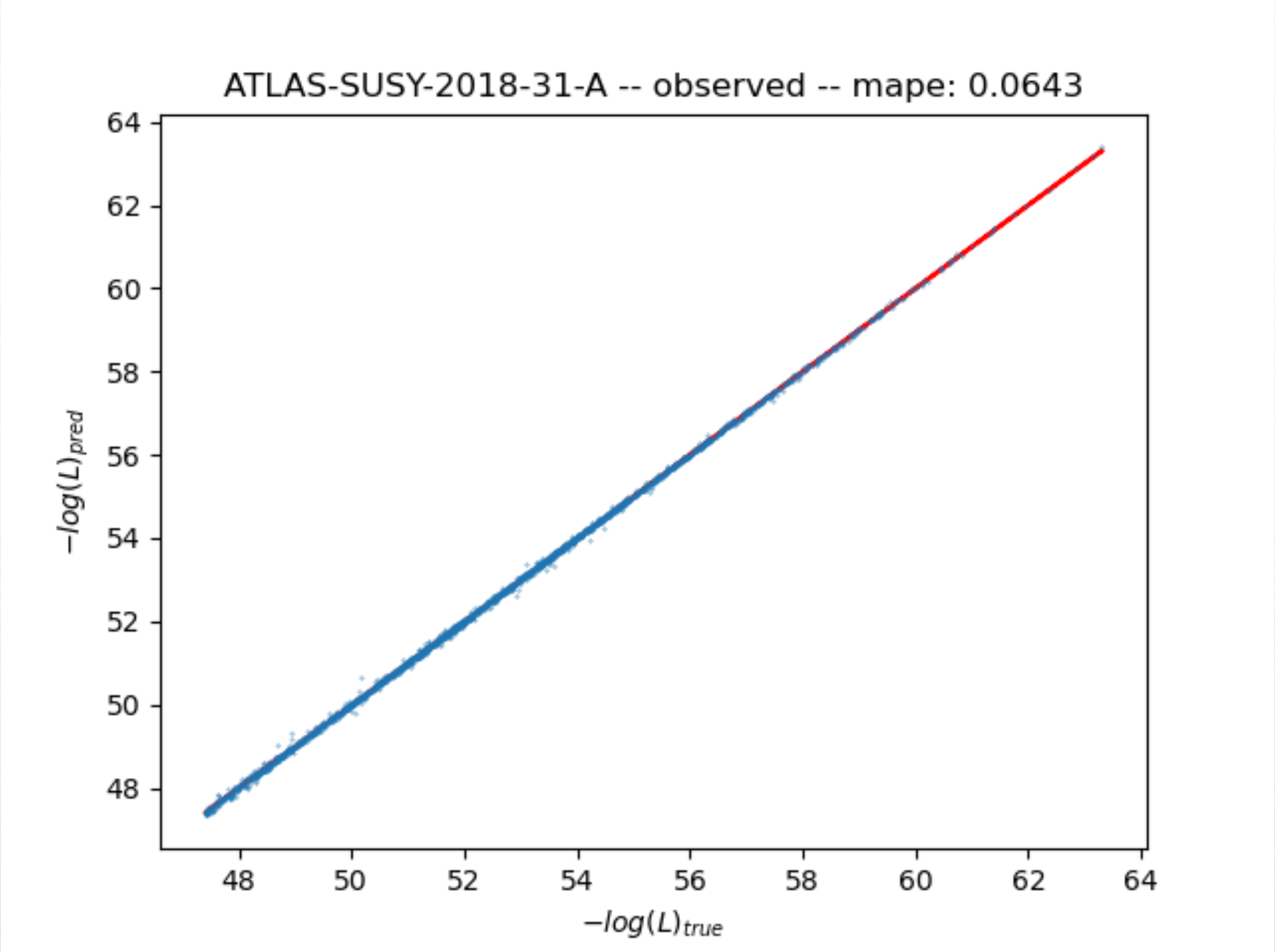
ATLAS-SUSY-2018-31-A, 3 SRS

OBSERVED





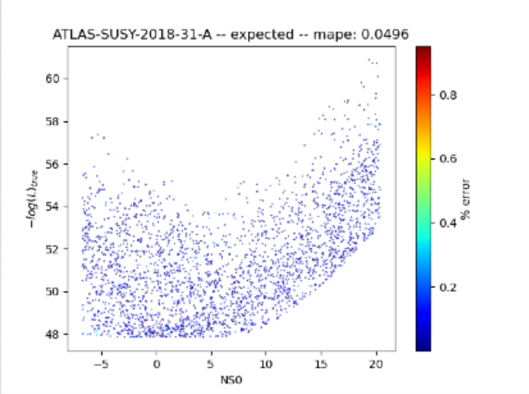


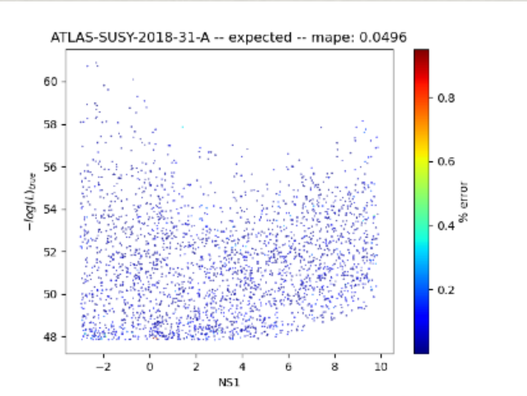


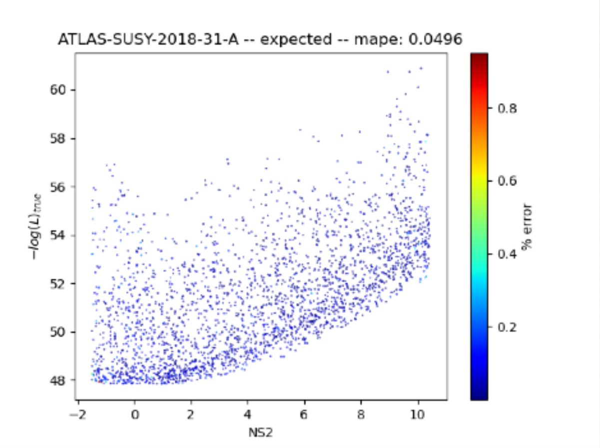
HYPERPARAMS	
Hidden layers	3X1024
L2 regularisation	10^-5
Max epochs	300 X 4
Batch size	32
N samples	20k

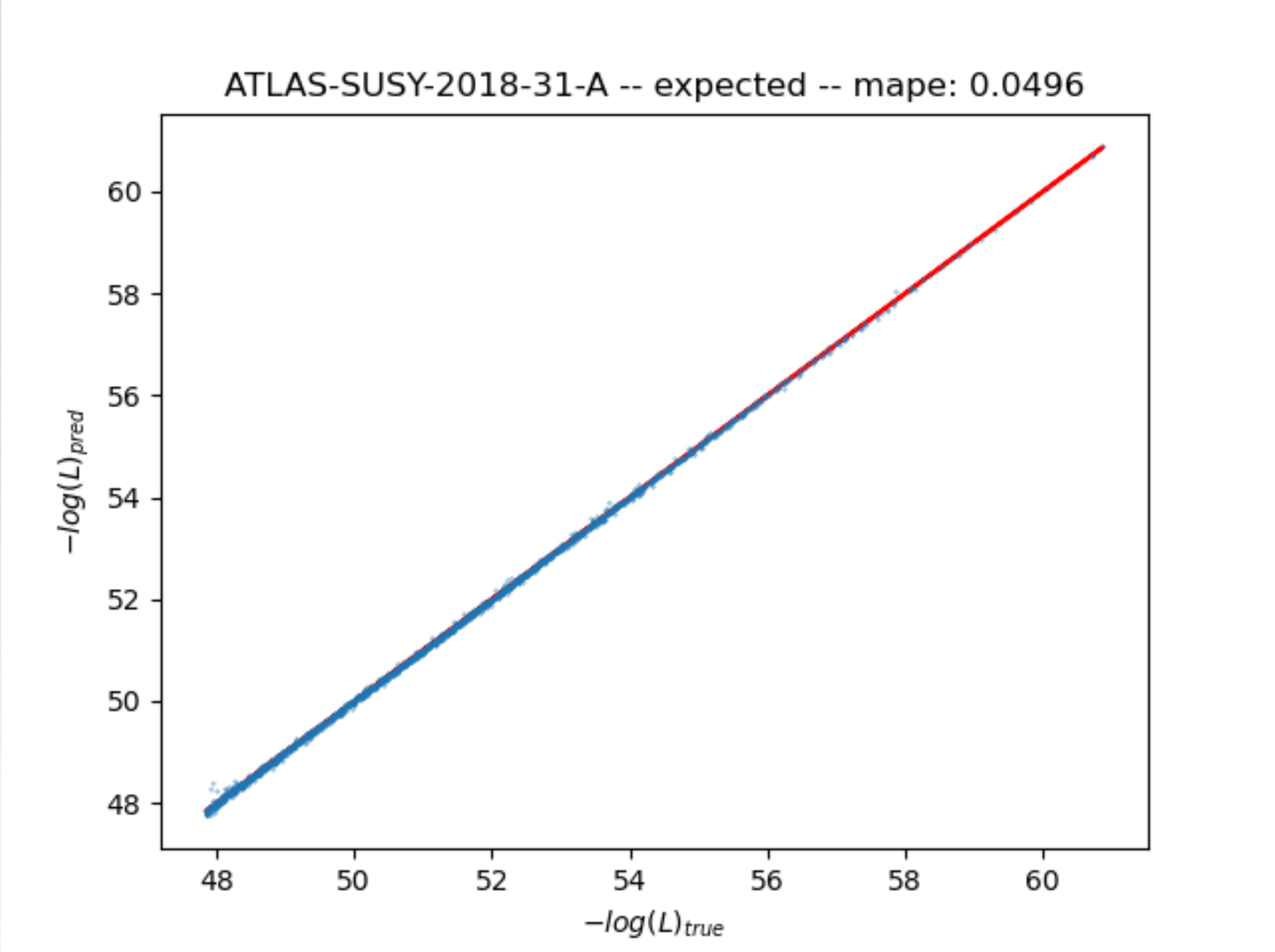
METRICS
MAPE
0.0643
MaxAPE
0.96305

EXPECTED







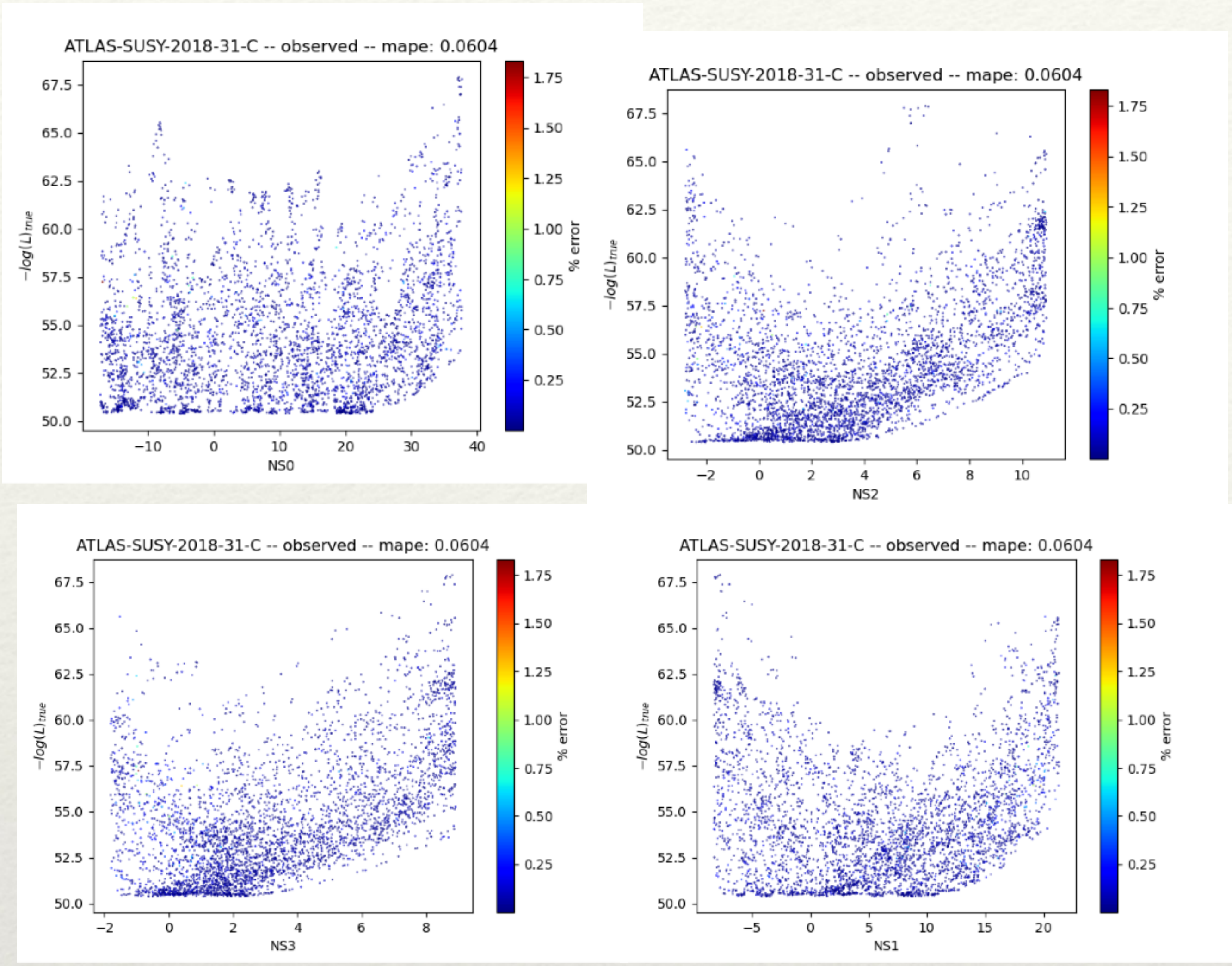


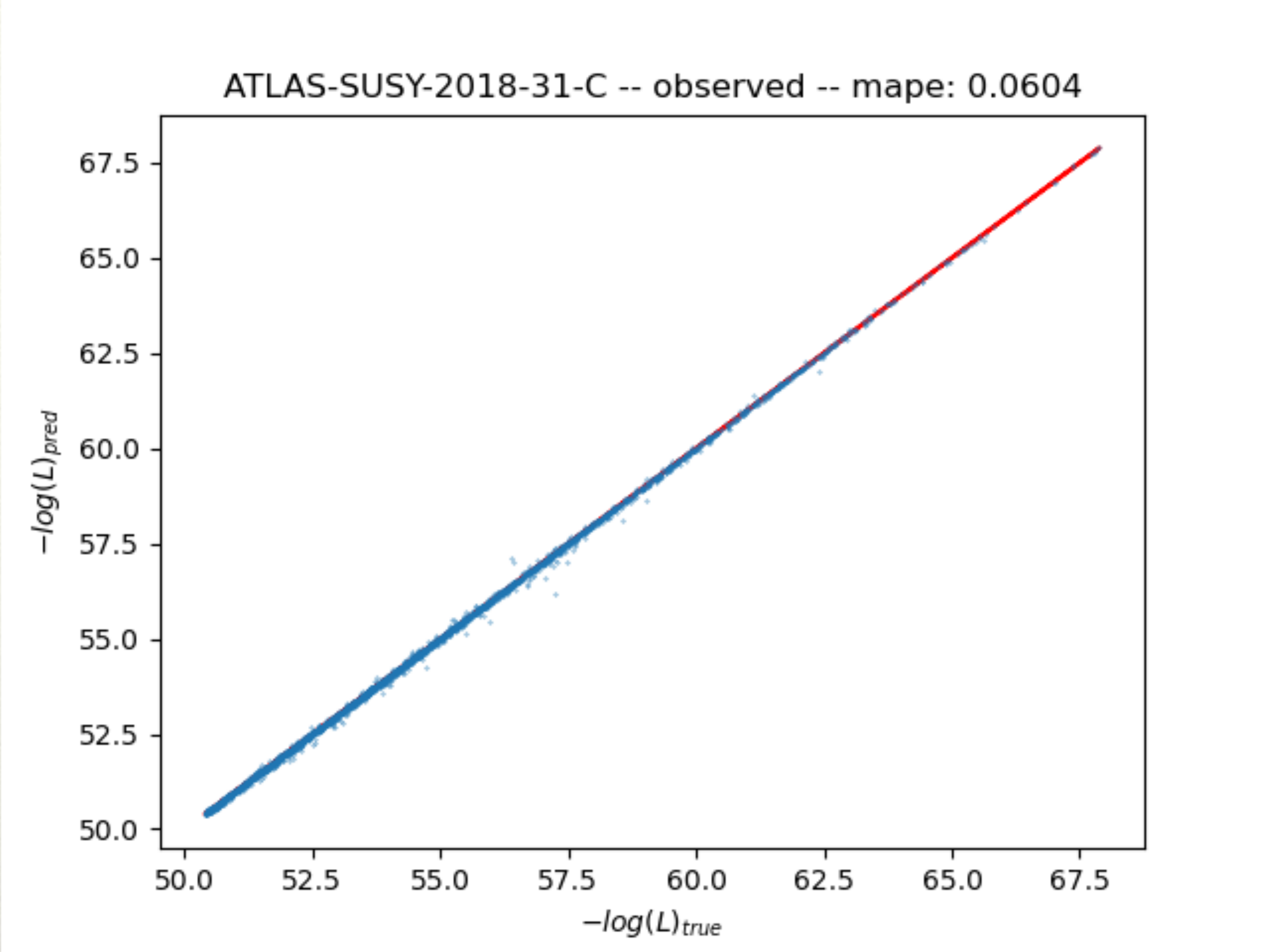
HYPERPARAMS	
Hidden layers	2X1024
L2 regularisation	10^-5
Max epochs	300 X 4
Batch size	32
N samples	20k

METRICS
MAPE
.04969
MaxAPE
0.9501

ATLAS-SUSY-2018-31-C, 3 SRS

OBSERVED

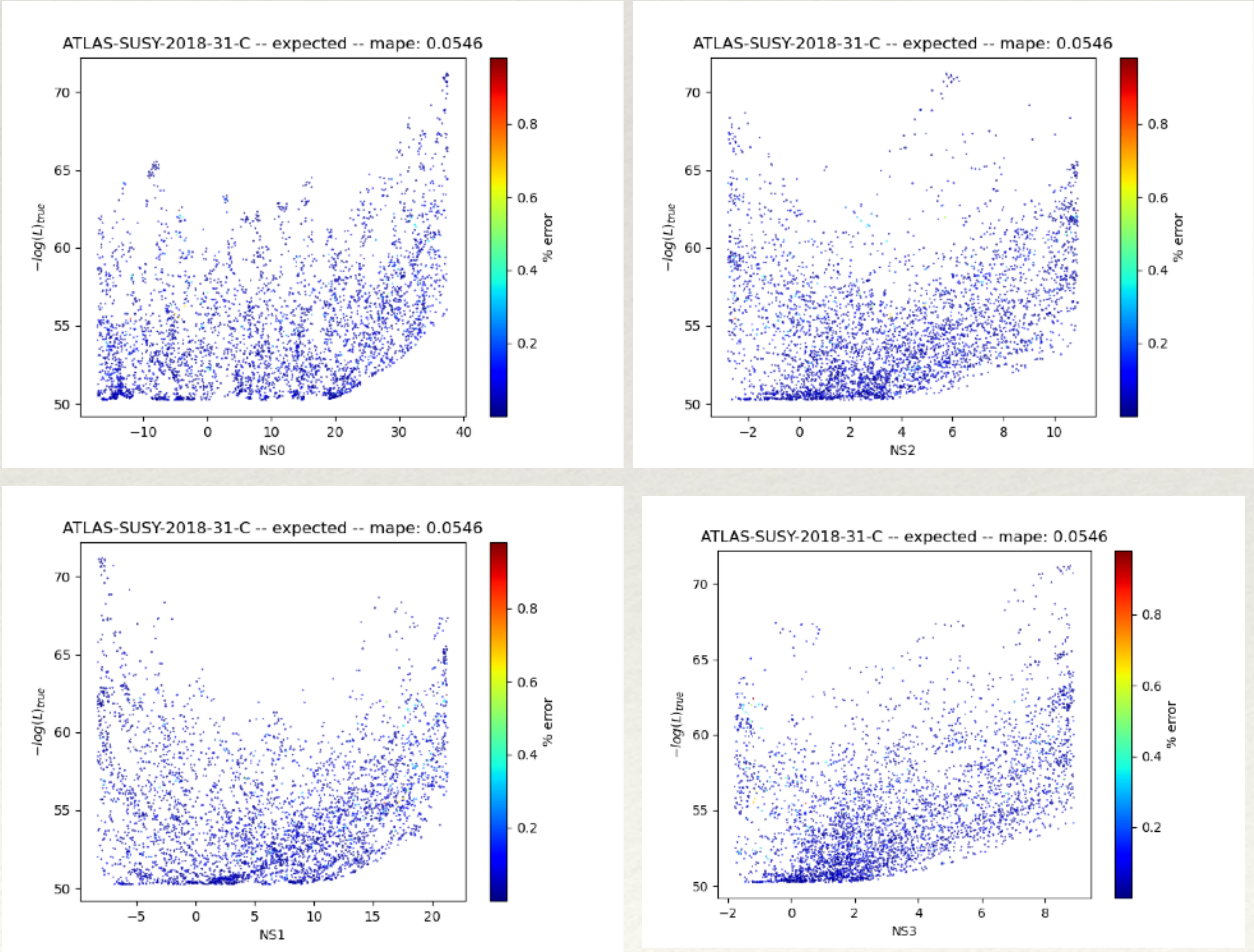


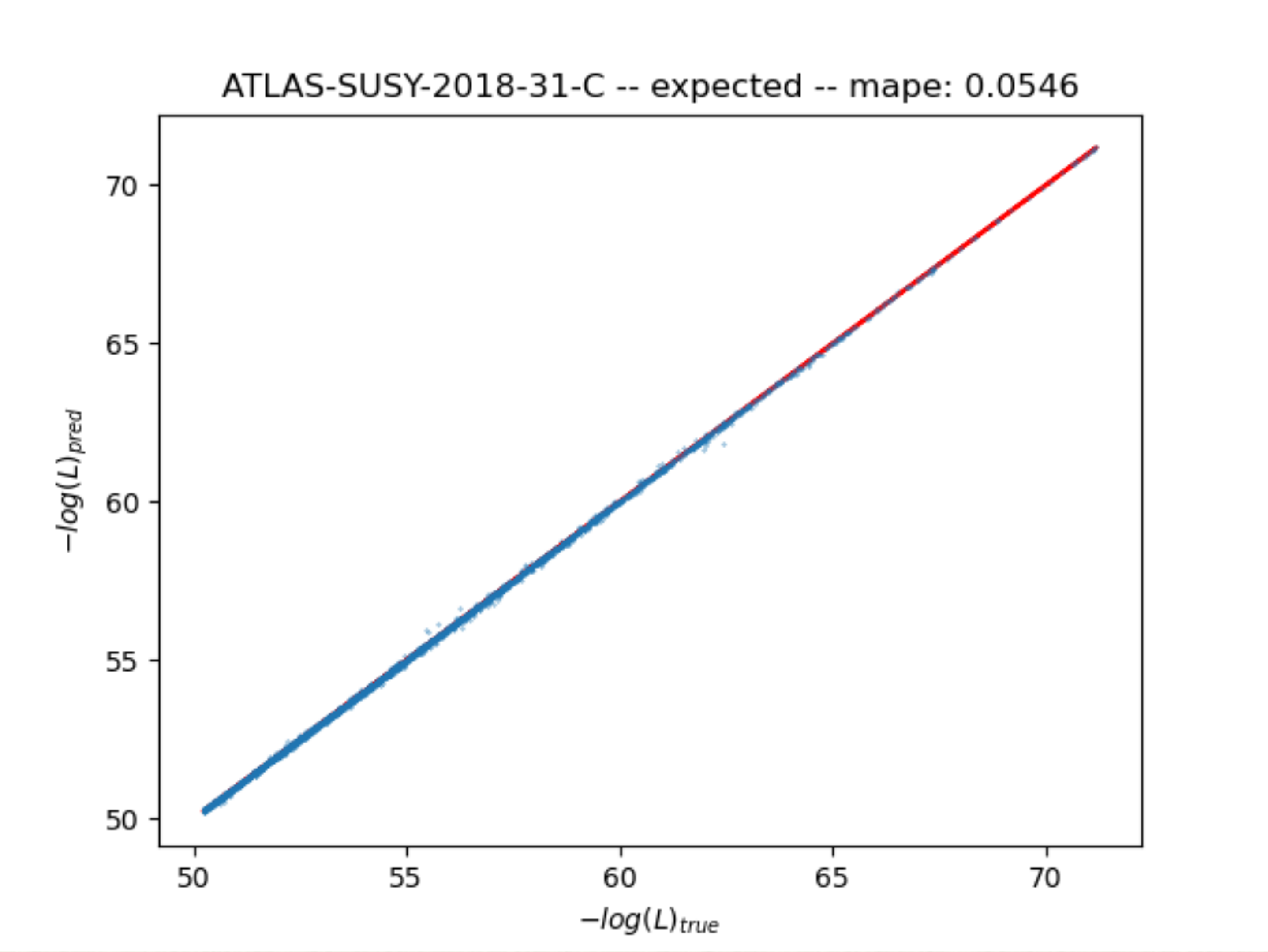


HYPERPARAMS	
Hidden layers	3X1024
L2 regularisation	0
Max epochs	300 X 4
Batch size	32
N samples	20k

METRICS
MAPE
0.0604
MaxAPE
1.8334

EXPECTED



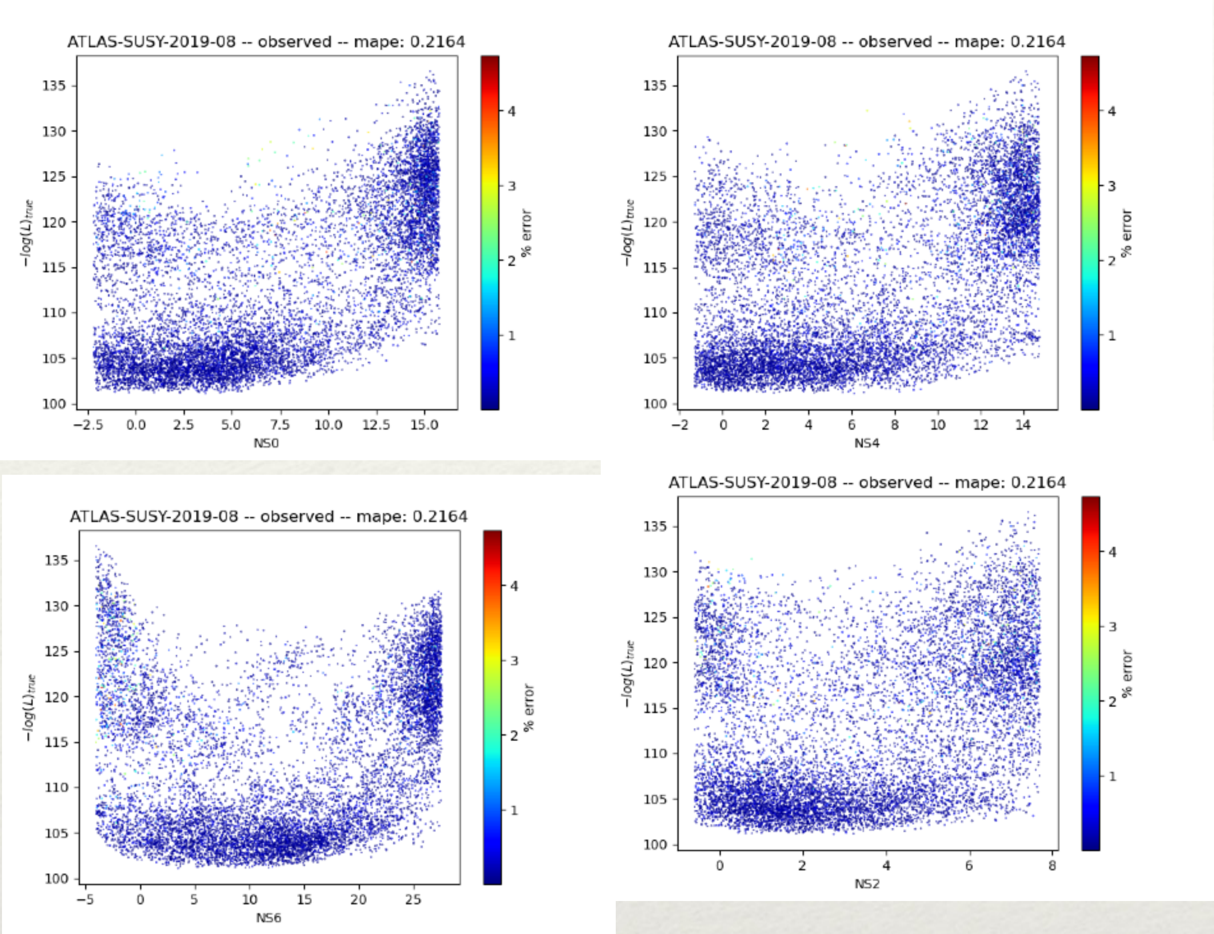


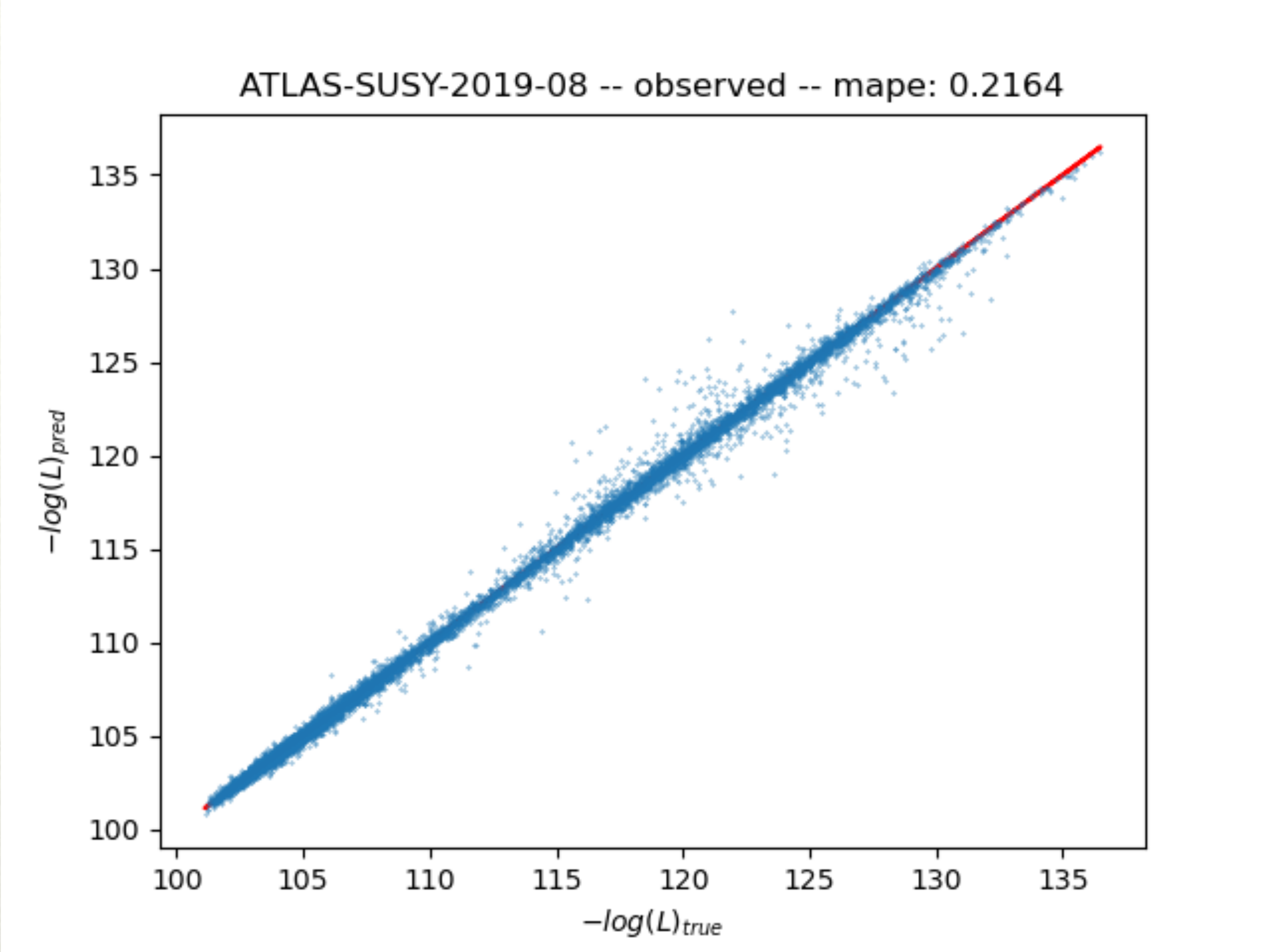
HYPERPARAMS	
Hidden layers	3X1024
L2 regularisation	0
Max epochs	300 X 4
Batch size	32
N samples	20k

METRICS
MAPE
0.0546
MaxAPE
0.9813

ATLAS-SUSY-2019-08, 9 SRS

OBSERVED

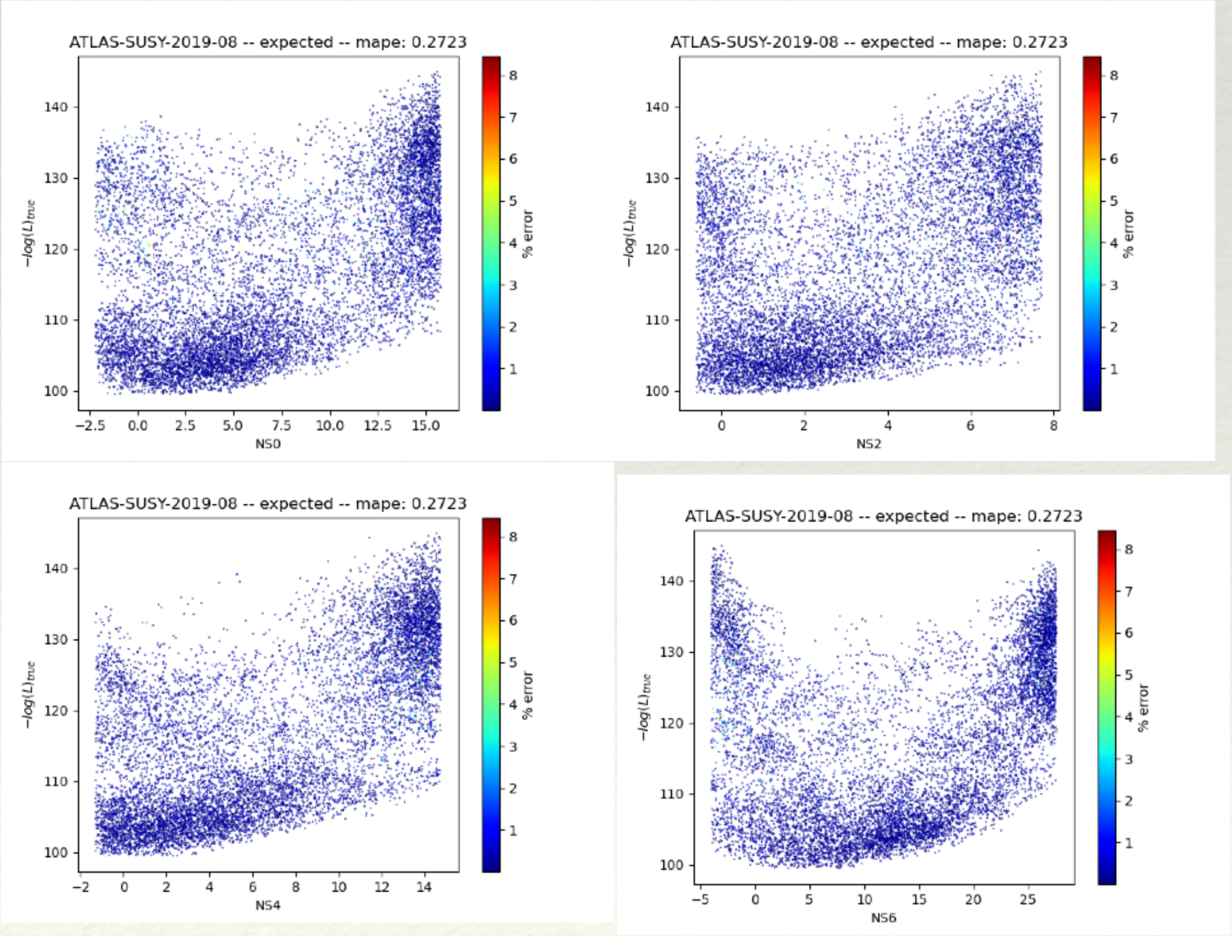


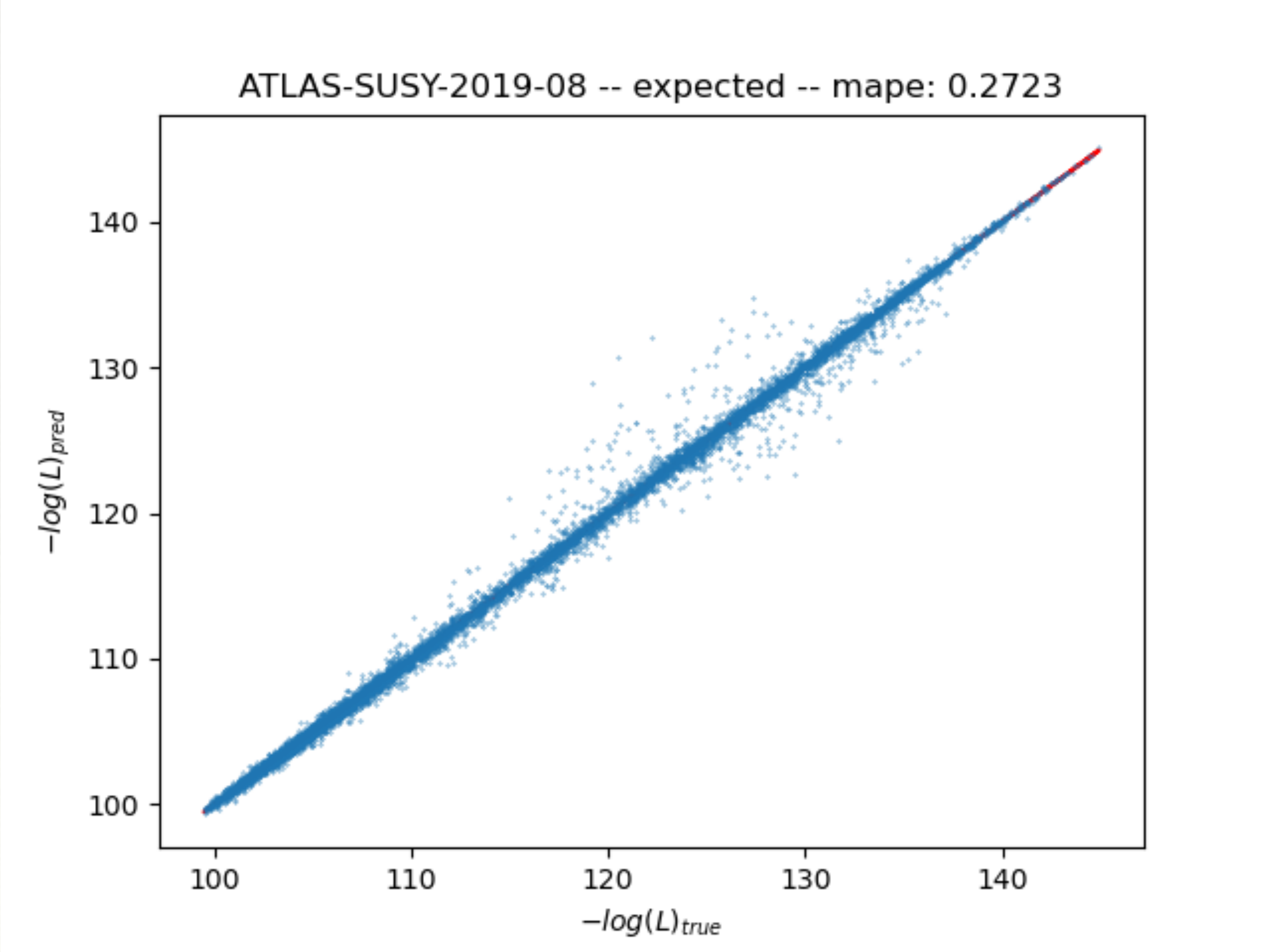


HYPERPARAMS	
Hidden layers	3X1024
L2 regularisation	0
Max epochs	1000 X 4
Batch size	32
N samples	80k

METRICS
MAPE
0.2164
MaxAPE
4.735

EXPECTED





HYPERPARAMS	
Hidden layers	4X1024
L2 regularisation	0
Max epochs	1000 X 4
Batch size	32
N samples	80k

METRICS
MAPE
0.2723
MaxAPE
8.455

SPEED GAIN*

	PYHF	NN 1 sample	NN 1000 samples (single pass)
ATLAS-SUSY-2018-04	.75s	.026s	.037s
ATLAS-SUSY-2018-31-A	.46s	.027s	.063s
ATLAS-SUSY-2018-31-C	.58s	.03s	.063s
ATLAS-SUSY-2019-08	2.81s	.026s	.06s

At least an order of magnitude faster!

*On a Macbook Air

DEPLOYMENT STRATEGY

Under construction

- ONNX models will be available on GitHub/Zenodo.
- Write a Spey backend for smooth statistical interpretations.
- Planned interface to SModelS.
- Ready to use for your reinterpretation studies.

Conclusions

- Full statistical models are a key aspect of the LHC legacy.
- From them, we extract accurate profile likelihoods required for our pheno studies.
- Computing them from pyhf's statistical models takes a significant amount of time in the large scale.
- **NNs provide an orders of magnitude faster solution.**
- **We found that profiled likelihoods are easily learnable by NNs**.
- **They can easily be integrated into modern reinterpretation frameworks.**

THANK YOU!