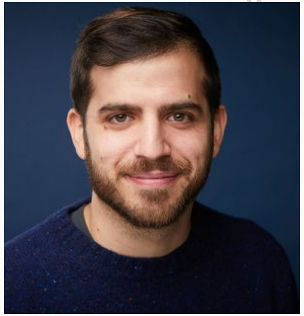
Reliable Uncertainty Quantification

Advances in Calibration and Likelihood-Free Inference

Rafael Izbicki (UFSCar)

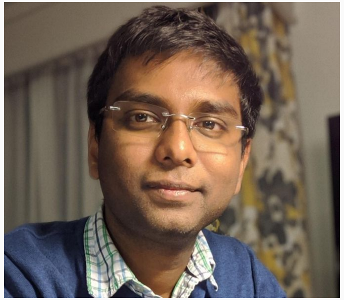
Agradecimentos: FAPESP, CAPES, CNPq

L. Masserano
(Stats&DS/CMU)



Nic Dalmasso
(JPMorgan)

Ann B. Lee
(Stats&DS/CMU)



Biprateep Dey
(U Toronto/CITA)

Jeffrey A. Newman
(Astro/Pitt)

Brett H. Andrews
(Astro/Pitt)

T. Dorigo (INFN)

Part 1

arxiv 2205.14568

Towards Instance-Wise Calibration: Local Amortized Diagnostics and Reshaping of Conditional Densities (LADaR)

Biprateep Dey[1,2,3,4,5,6], David Zhao[7], Brett H. Andrews[1,2], Jeffrey A. Newman[1,2], Rafael Izbicki[8] and Ann B. Lee[7]

[1]Department of Physics and Astronomy, University of Pittsburgh
[2]Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh
[3]Department of Statistical Sciences, University of Toronto
[4]Canadian Institute for Theoretical Astrophysics (CITA), University of Toronto
[5]Dunlap Institute for Astronomy & Astrophysics, University of Toronto
[6]Vector Institute
[7]Department of Statistics and Data Science, Carnegie Mellon University
[8]Department of Statistics, Federal University of São Carlos (UFSCar)

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

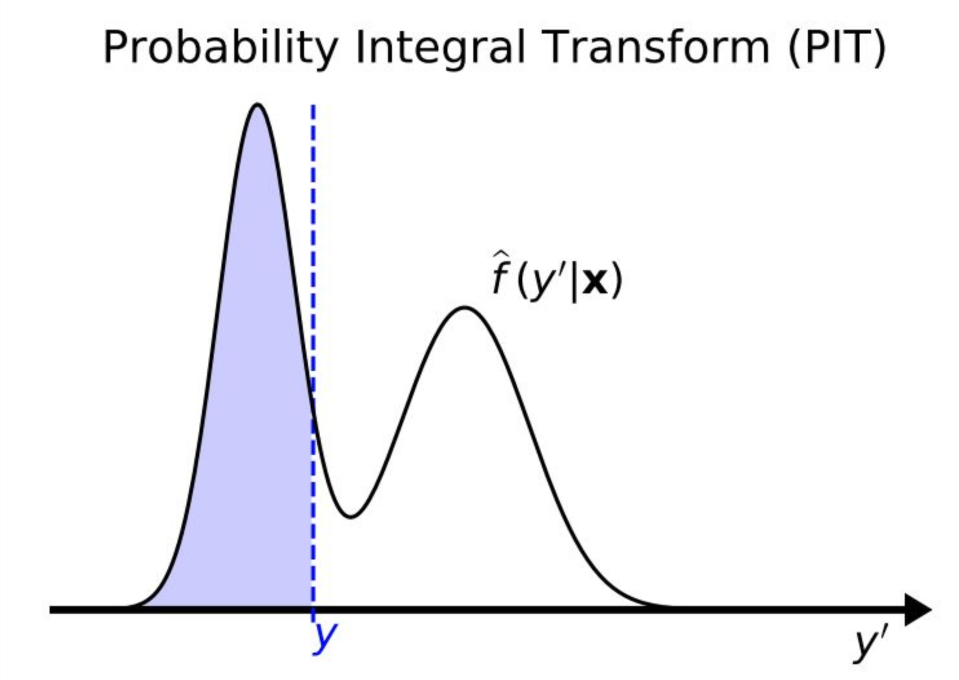
MNRAS **499,** 1587–1606 (2020)
Advance Access publication 2020 September 14

doi:10.1093/mnras/staa2799

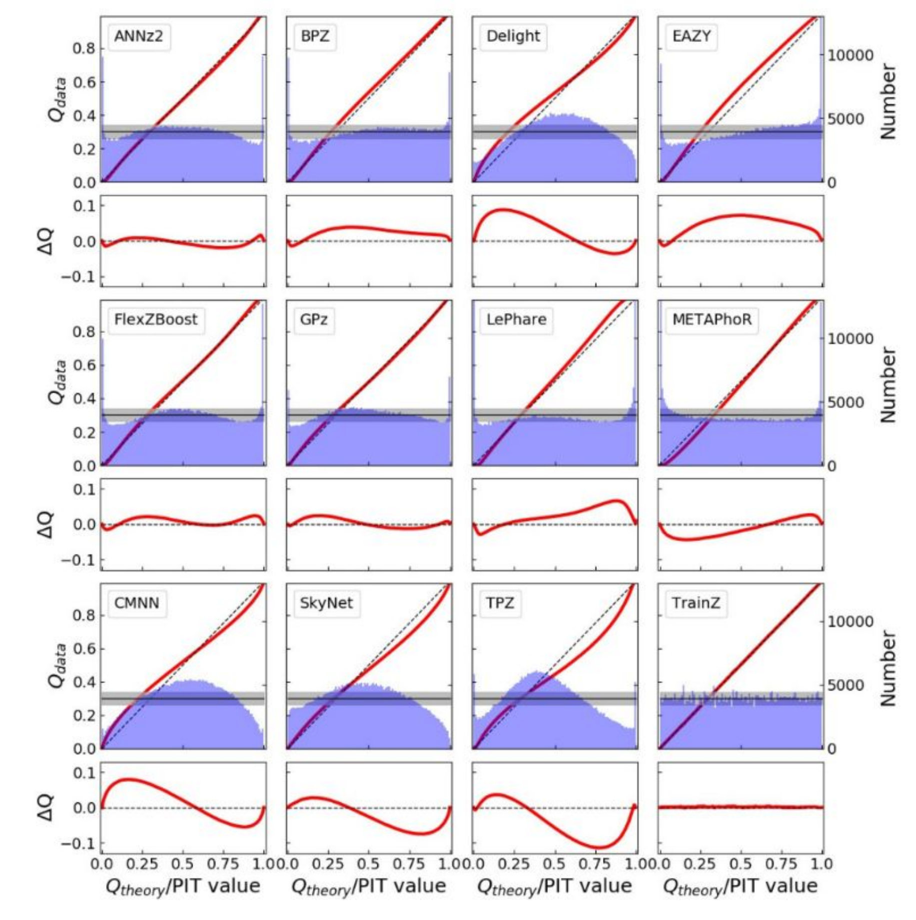
Evaluation of probabilisti … mation approaches for The Rubin Observatory L … Time (LSST)

S. J. Schmidt,[1★] A. I. Malz,[2,3,4★] J. … M. Brescia,[9] S. Cavuoti,[9,10]
J. Cohen-Tanugi,[11] A. J. Connolly,[12] … M. L. Graham,[12]
K. G. Iyer,[19,20] M. J. Jarvis,[21,22] J. … e,[18] G. Longo,[10]
C. B. Morrison,[12] J. A. Newman,[24] E … [18] H. Tranin,[11]
R. H. Wechsler,[16,25,26] R. Zhou,[15,24 … nergy Science Collaboration)

Photo-z Code	CDE Loss
ANNz2	−6.88
BPZ	−7.82
Delight	−8.33
EAZY	−7.07
FlexZBoost	−10.60
GPz	−9.93
LePhare	−1.66
METAPhoR	−6.28
CMNN	−10.43
SkyNet	−7.89
TPZ	−9.55
trainZ	−0.83



$$\mathrm{PIT}(Y_1;\mathbf{X}_1),\ldots,\mathrm{PIT}(Y_n;\mathbf{X}_n) \overset{i.i.d.}{\sim} \mathrm{Unif}(0,1)$$



ANNz2
BPZ
Delight
EAZY
FlexZBoost
GPz
LePhare
METAPhoR
CMNN
SkyNet
TPZ
TrainZ
Q_{data}
ΔQ
Number
Q_{theory}/PIT value

Algorithm 1 `Cal-PIT`

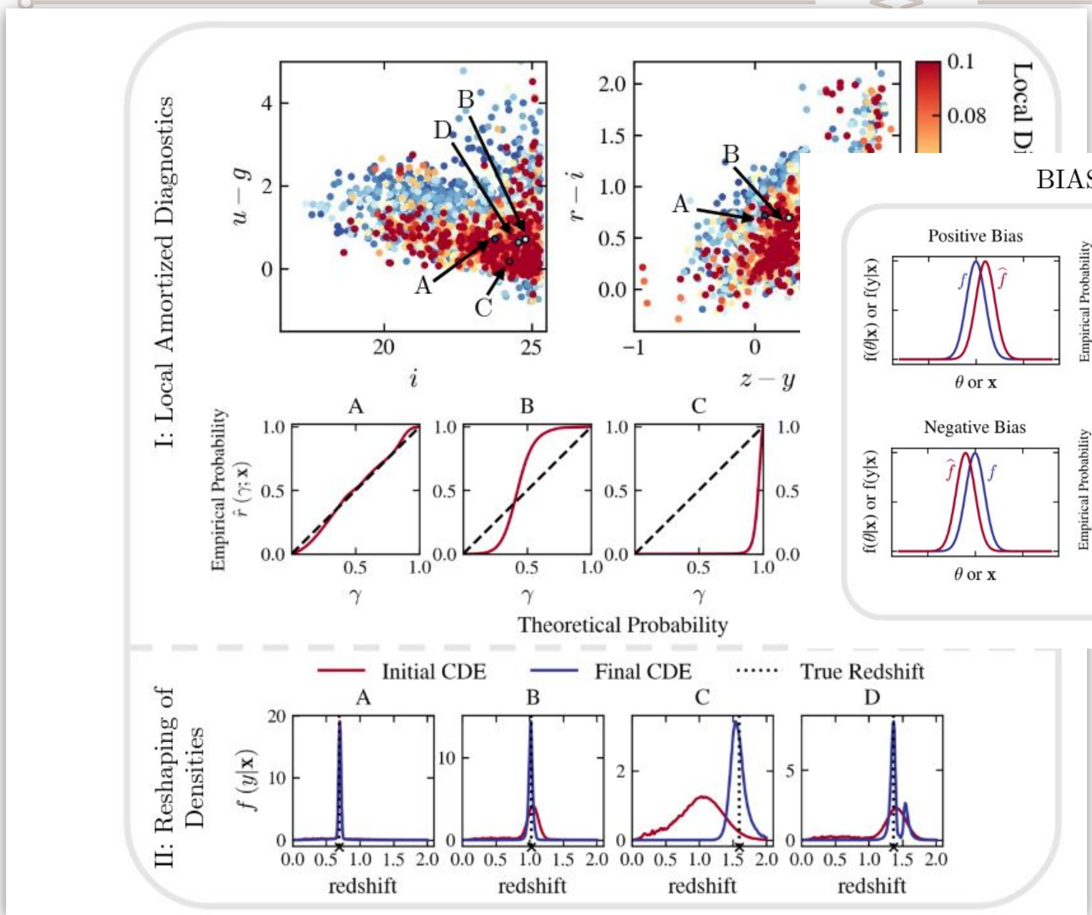
Require: initial CDE $\widehat{f}(y|\mathbf{x})$ evaluated at $y \in G$; calibration set $\mathcal{D} = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$; oversampling factor K; evaluation points $\mathcal{V} \subset \mathcal{X}$; nominal miscoverage level α, flag `HPD` (true if computing HPD sets)

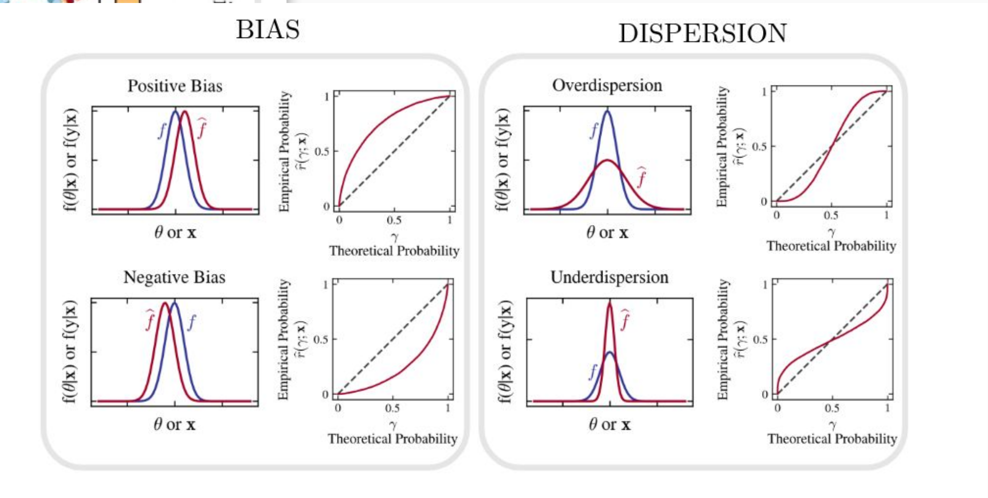
Ensure: new distribution $\widetilde{F}(y|\mathbf{x})$, `Cal-PIT` interval $C(\mathbf{x})$, new density estimate $\widetilde{f}(y|\mathbf{x})$, for all $\mathbf{x} \in \mathcal{V}$

// **Learn PIT-CDF from augmented and upsampled calibration data** $\mathcal{D}'$
Set $\mathcal{D}' \leftarrow \emptyset$
for i in $\{1, ..., n\}$ **do**
  for j in $\{1, ..., K\}$ **do**
    Draw $\gamma_{i,j} \sim U(0, 1)$
    Compute $W_{i,j} \leftarrow \mathbb{I}\left(\mathrm{PIT}(Y_i; \mathbf{X}_i) \leq \gamma_{i,j}\right)$
    Let $\mathcal{D}' \leftarrow \mathcal{D}' \cup \{(\mathbf{X}_i, \gamma_{i,j}, W_{i,j})\}$
  end for
end for
Use $\mathcal{D}'$ to learn $\widehat{r}^{\widehat{f}}(\gamma; \mathbf{x}) := \widehat{\mathbb{P}}\left(\mathrm{PIT}(Y; \mathbf{x}) \leq \gamma \mid \mathbf{x}\right)$ via a regression of W on $\mathbf{X}$ and γ, which is monotonic w.r.t. γ.

// **Map initial CDE into a new CDE by applying learnt PIT-CDF**
for $\mathbf{x} \in \mathcal{V}$ **do**
  // **Construct recalibrated CDE**
  Compute $\widehat{F}(y|\mathbf{x}) \leftarrow \mathrm{cumsum}(\widehat{f}(y|\mathbf{x}))$ for $y \in G$
  Let $\widetilde{F}(y|\mathbf{x}) \leftarrow \widehat{r}^{\widehat{f}}\left(\widehat{F}(y|\mathbf{x}); \mathbf{x}\right)$ for $y \in G$
  Apply interpolating (or smoothing) splines to obtain $\widetilde{F}(\cdot|\mathbf{x})$ and $\widetilde{F}^{-1}(\cdot|\mathbf{x})$
  Differentiate $\widetilde{F}(y|\mathbf{x})$ to obtain new distribution $\widetilde{f}(y|\mathbf{x})$ for $y \in G$
  Renormalize $\widetilde{f}(y|\mathbf{x})$ according to Izbicki & Lee (2016, Section 2.2)

  // **Construct Cal-`PIT` interval with conditional coverage** $1 - \alpha$
  Compute $C(\mathbf{x}) \leftarrow [\widetilde{F}^{-1}(0.5\alpha|\mathbf{x}); \widetilde{F}^{-1}(1 - 0.5\alpha|\mathbf{x})]$.
  if `HPD` **then**
    Obtain HPD sets $C(\mathbf{x}) = \{y : \widetilde{f}(y|\mathbf{x}) \geq \widetilde{t}_{\mathbf{x},\alpha}\}$, where $\widetilde{t}_{\mathbf{x},\alpha}$ is such that $\int_{y \in C_\alpha(\mathbf{x})} \widetilde{f}(y|\mathbf{x}) dy = 1 - \alpha$
  end if
end for
return $\widetilde{F}(y|\mathbf{x})$, $C(\mathbf{x})$, $\widetilde{f}(y|\mathbf{x})$, for all $\mathbf{x} \in \mathcal{V}$





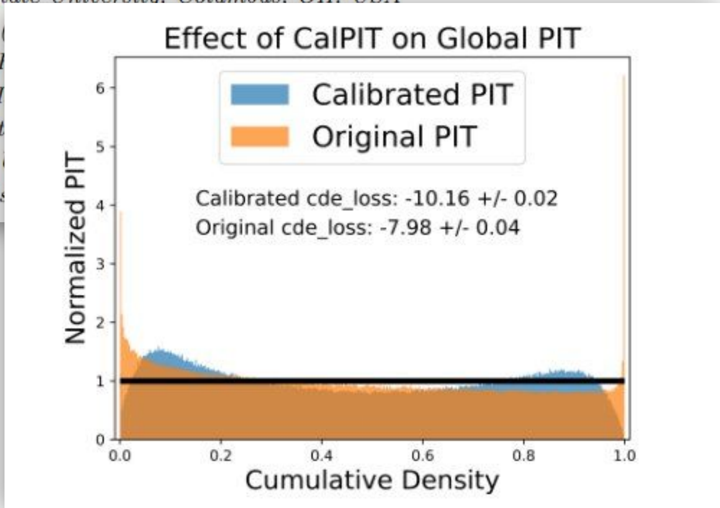
https://lee-group-cmu.github.io/cal-pit-paper/fig_1_interactive/

Photo-z Algorithm	CDE Loss
ANNz2 (Sadeh et al., 2016)	–6.88
BPZ (Benítez, 2000)	–7.82
Delight (Leistedt & Hogg, 2017)	–8.33
EAZY (Brammer et al., 2008)	–7.07
FlexZBoost (Izbicki & Lee, 2017)	**–10.60**
GPz (Almosallam et al., 2016)	–9.93
LePhare (Arnouts et al., 1999)	–1.66
METAPhoR (Cavuoti et al., 2017)	–6.28
CMNN (Graham et al., 2018)	–10.43
SkyNet (Graff et al., 2014)	–7.89
TPZ (Carrasco Kind & Brunner, 2013)	–9.55
trainZ (Schmidt et al., 2020)	–0.83
Cal-PIT	**–10.80**

Mantis Shrimp: Exploring Photometric Band Utilization in Computer Vision Networks for Photometric Redshift Estimation

Andrew W. Engel,[1,2,3] Nell Byler,[1] Adam Tsou,[4] Gautham Narayan,[5,6,7] Emmanuel Bonilla,[8] and Ian Smith[8]

[1] National Security Directorate, Pacific Northwest National Laboratory, Richland, WA, USA
[2] Department of Physics, Ohio State University, Columbus, OH, USA
[3] Center for Cosmology and AstroParticle Physics
[4] Department of Mathematics, John
[5] Department of Astronomy, University of
[6] Center for AstroPhysical Surveys, National Cent
[7] Illinois Center for Advanced Studies of the
[8] Research Computing, Pacific Northwes



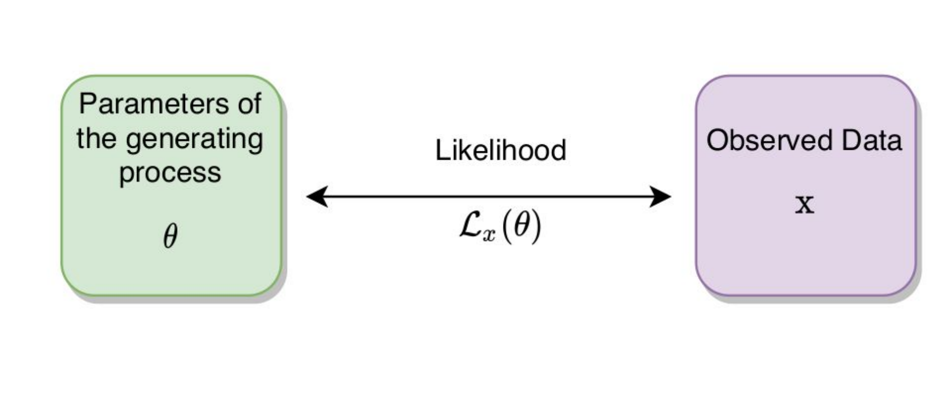
Part 2

Electronic Journal of Statistics
Vol. 18 (2024) 5045–5090
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2307

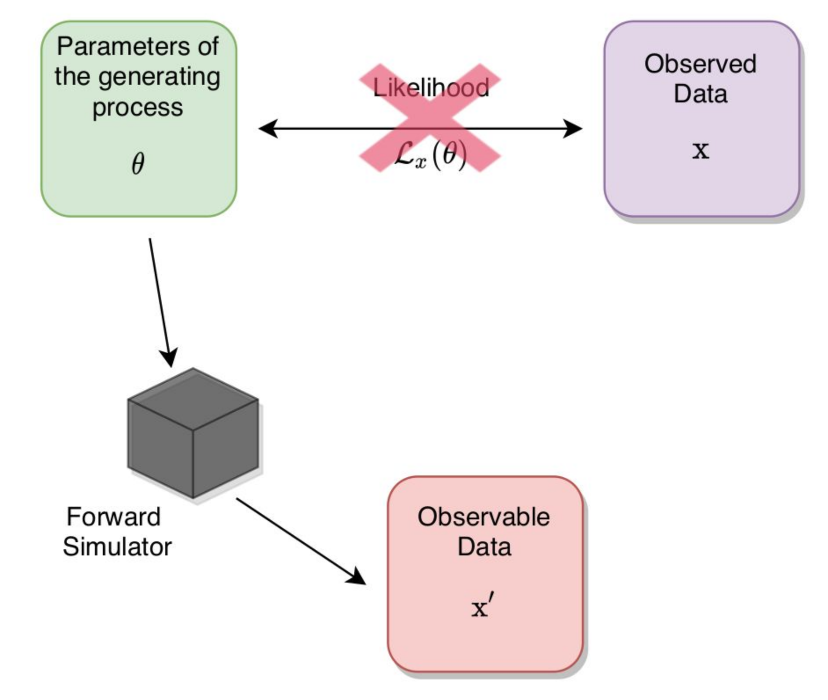
Likelihood-free frequentist inference: bridging classical statistics and machine learning for reliable simulator-based inference*

Niccolò Dalmasso[†,1], Luca Masserano[†,2],
David Zhao[2], Rafael Izbicki[3], Ann B. Lee[2]

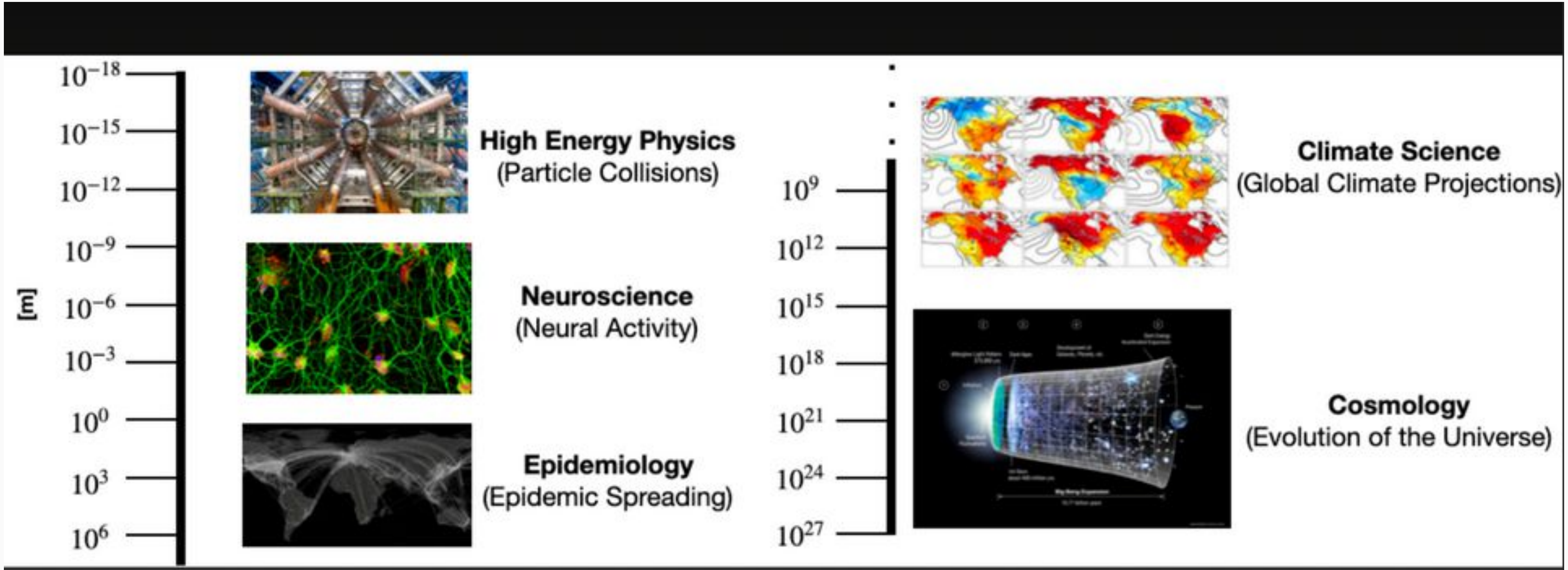
Statistical Inference



Statistical Inference



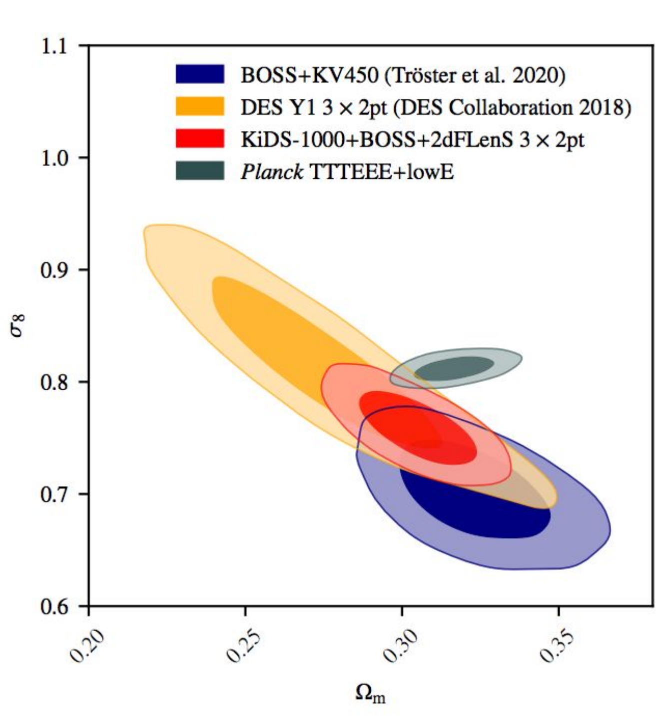
Model/theory comes in the form of simulations



Credit: Dalmasso (adapted from Cranmer et al, 2020)

How about valid frequentist inference?

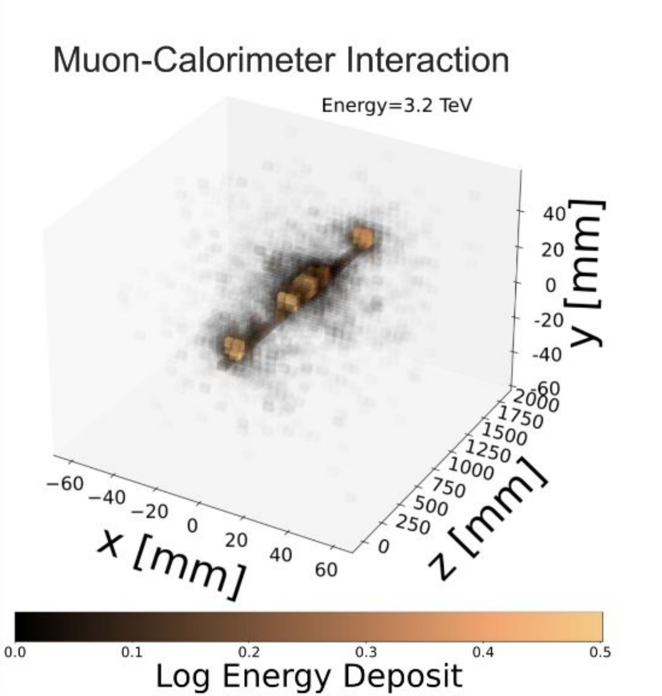
$$\mathbb{P}_{\mathcal{D}|\theta}(\theta \in R(\mathcal{D})) = 1 - \alpha, \ \forall \theta \in \Theta$$



Source: KiDS collaboration 2020

$$R(\mathbf{X}) := \{\theta \in \Theta \,|\, \tau(\mathbf{X};\theta) \geq C_\theta\}$$

Confidence Sets for Muon Energies using CNN

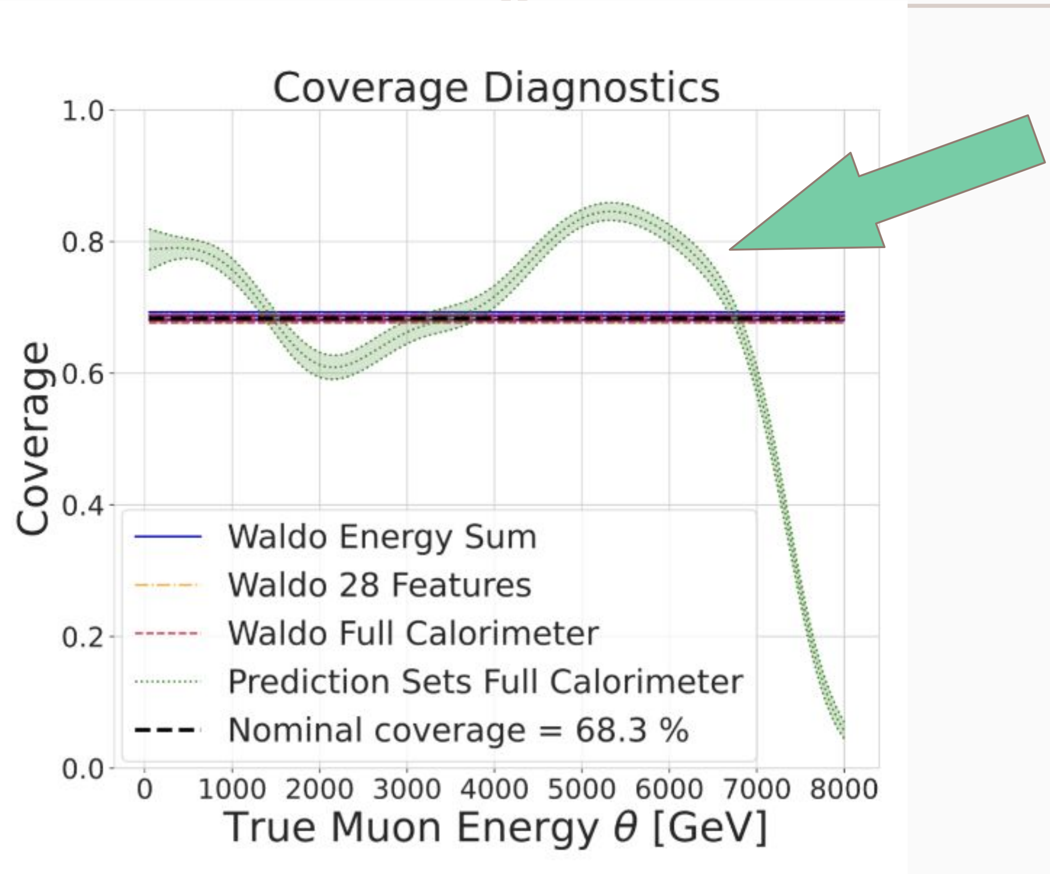


Calorimetric Measurement of Multi-TeV Muons via Deep Regression

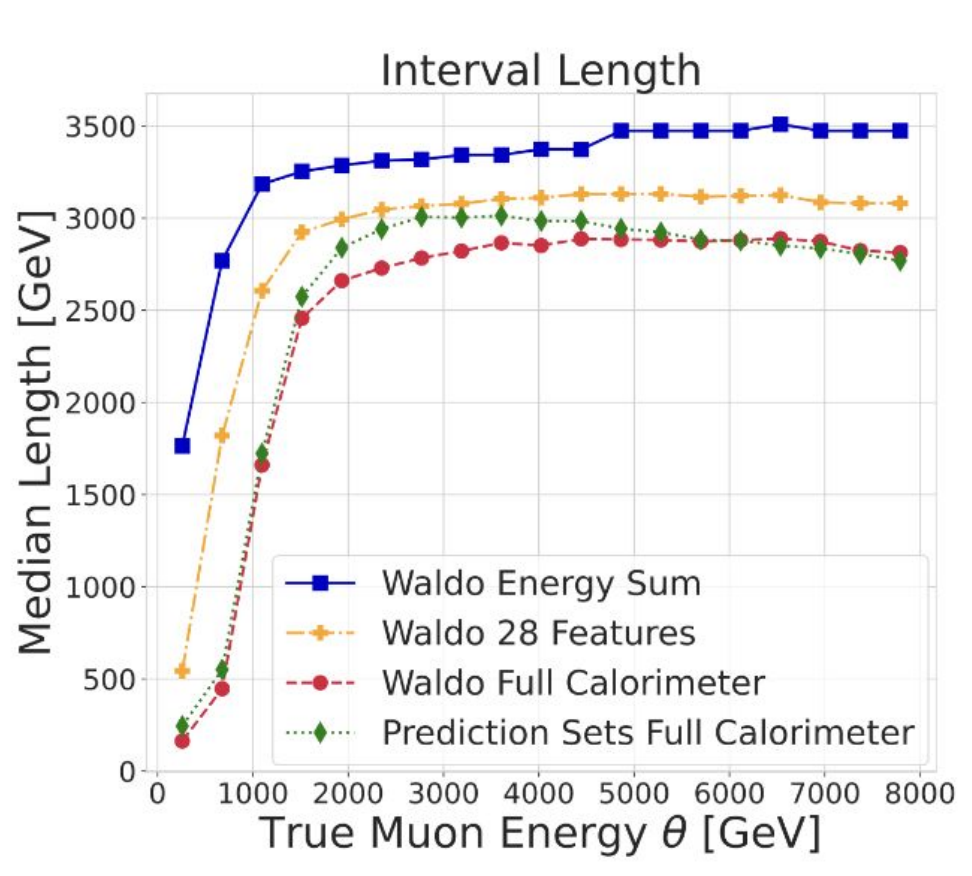
Jan Kieseler[a,1], **Giles C. Strong**[b,2,3], **Filippo Chiandotto**[c,2], **Tommaso Dorigo**[d,3], **Lukas Layer**[e,4,3]

[1]CERN
[2]Università di Padova
[3]INFN, Sezione di Padova
[4]Università di Napoli “Federico II”

Received: 6 August 2021 / Accepted: 4 January 2022 / Published online: 27 January 2022
Eur. Phys. J. C (2022) 82:79 https://doi.org/10.1140/epjc/s10052-022-09993-5



Coverage Diagnostics
1.0
0.8
0.6
0.4
0.2
0.0
Coverage
0
1000
2000
3000
4000
5000
6000
7000
8000
True Muon Energy θ [GeV]
Waldo Energy Sum
Waldo 28 Features
Waldo Full Calorimeter
Prediction Sets Full Calorimeter
Nominal coverage = 68.3 %


Interval Length
Median Length [GeV]
3500
3000
2500
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000
7000
8000
True Muon Energy θ [GeV]
Waldo Energy Sum
Waldo 28 Features
Waldo Full Calorimeter
Prediction Sets Full Calorimeter

Statistics > Methodology

[Submitted on 28 Nov 2024]

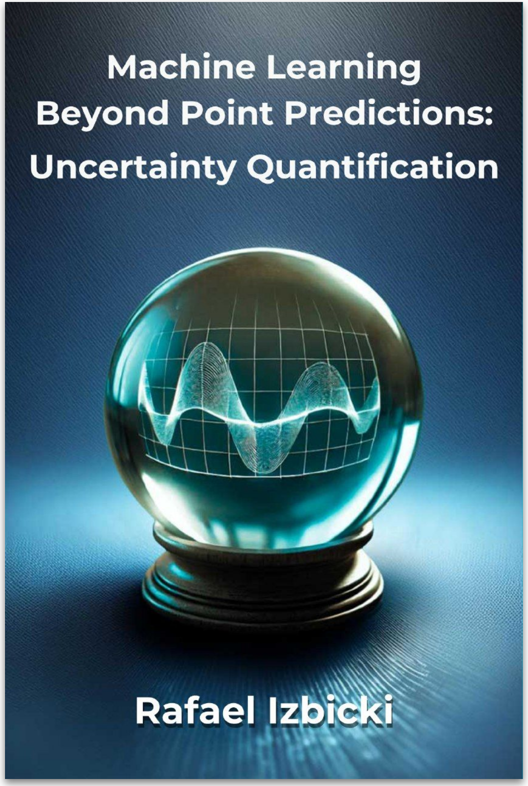
Distribution-Free Calibration of Statistical Confidence Sets

Luben M. C. Cabezas, Guilherme P. Soares, Thiago R. Ramos, Rafael B. Stern, Rafael Izbicki

Constructing valid confidence sets is a crucial task in statistical inference, yet traditional methods often face challenges when dealing with complex models or limited observed sample sizes. These challenges are frequently encountered in modern applications, such as Likelihood-Free Inference (LFI). In these settings, confidence sets may fail to maintain a confidence level close to the nominal value. In this paper, we introduce two novel methods, TRUST and TRUST++, for calibrating confidence sets to achieve distribution-free conditional coverage. These methods rely entirely on simulated data from the statistical model to perform calibration. Leveraging insights from conformal prediction techniques adapted to the statistical inference context, our methods ensure both finite-sample local coverage and asymptotic conditional coverage as the number of simulations increases, even if n is small. They effectively handle nuisance parameters and provide computationally efficient uncertainty quantification for the estimated confidence sets. This allows users to assess whether additional simulations are necessary for robust inference. Through theoretical analysis and experiments on models with both tractable and intractable likelihoods, we demonstrate that our methods outperform existing approaches, particularly in small-sample regimes. This work bridges the gap between conformal prediction and statistical inference, offering practical tools for constructing valid confidence sets in complex models.

Summary

- **LADaR and Cal-PIT:** goodness-of-fit and recalibration of photo-z's to improve local calibration
- **LF2I:** confidence sets with the correct coverage for Simulation-Based Inference Problems



https://rafaelizbicki.com/UQ4MLpt

Thanks!

Questions?

rafaelizbicki@gmail.com

https://rafaelizbicki.com/

SMall

Statistical Machine Learning Lab

Alternative resources

Here's an assortment of alternative resources whose style fits the one of this template:

- Flat sparkling stars collection
- Vector dividers collection in hand drawn style
- Vector calligraphic ornamental divider collection