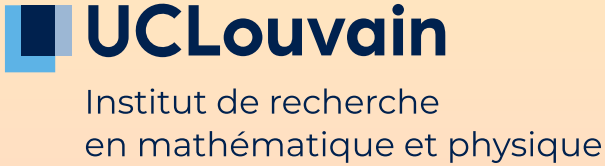
New approaches to the problem of unfolding

Pietro Vischia[1]

[1] CP3 — IRMP, Université catholique de Louvain



ICNFP 2020, Kolumbari

Contents

UCLouvain
Institut de recherche en mathématique et physique

The unfolding problem
Unfolding methods
- Popular methods
- Some early machine learning attempts

Two new proposals
- Simulation settings
- Unfolding with normalizing flows
- Unfolding by folding

Summary

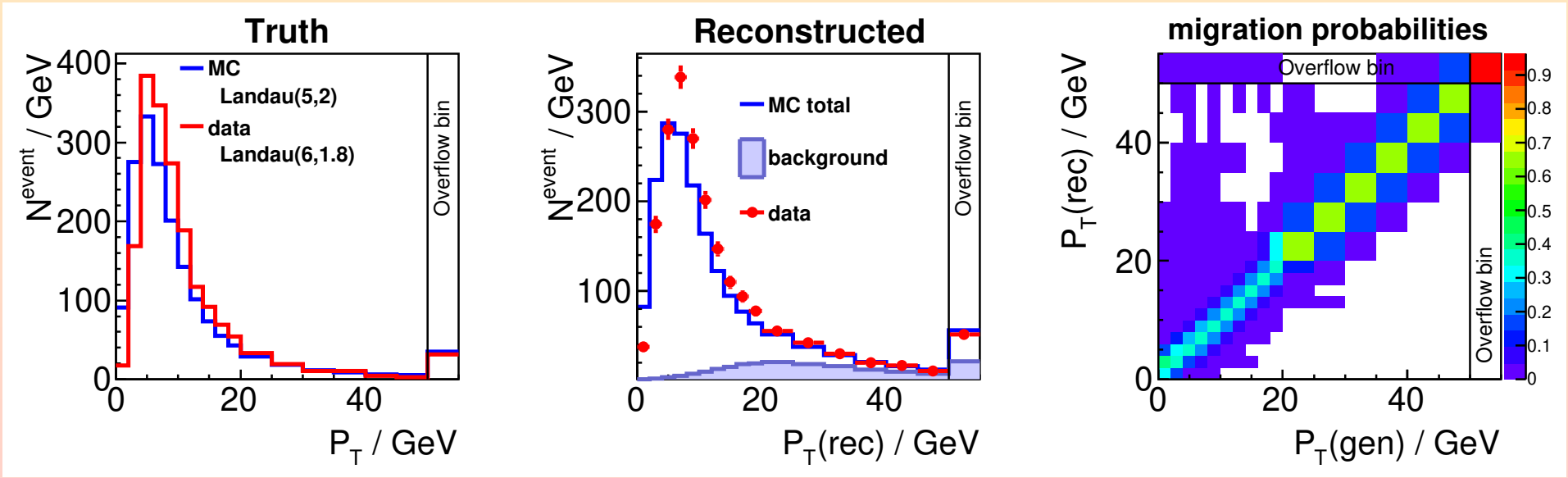
UCLouvain
Institut de recherche
en mathématique et physique

Unfolding: the problem

- Unfolding it's about how to invert a matrix that should not be inverted

$$\mathcal{L} = (\mathbf{y} - \mathbf{Ax})^T \mathbf{V_{yy}} (\mathbf{y} - \mathbf{Ax}),$$

- Observations y, to be transformed in the theory space into x
  - Model the detector as a response matrix
  - Invert the response to convert experimental data to theory space distributions
  - Usually to compare with models in the theory space
- The best solution is to fold any new theory and make comparisons in the experimental data space



Plot from ArXiv:1611.01927

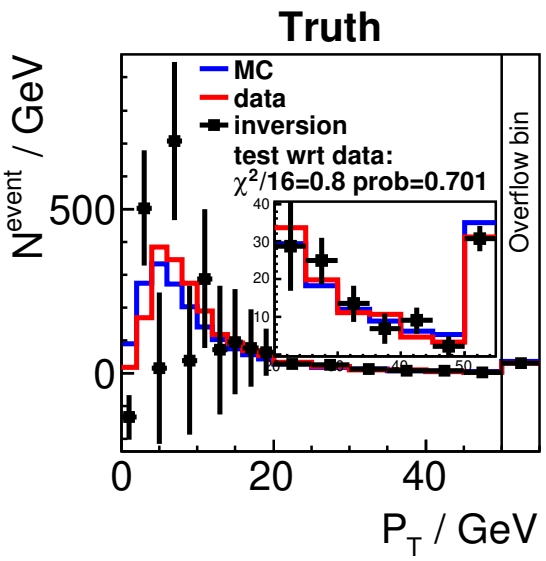
UCLouvain
Institut de recherche
en mathématique et physique

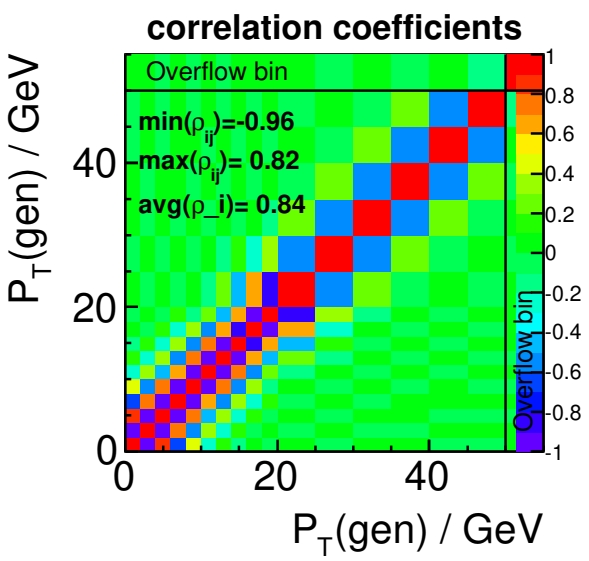
Unfolding: naïve solutions

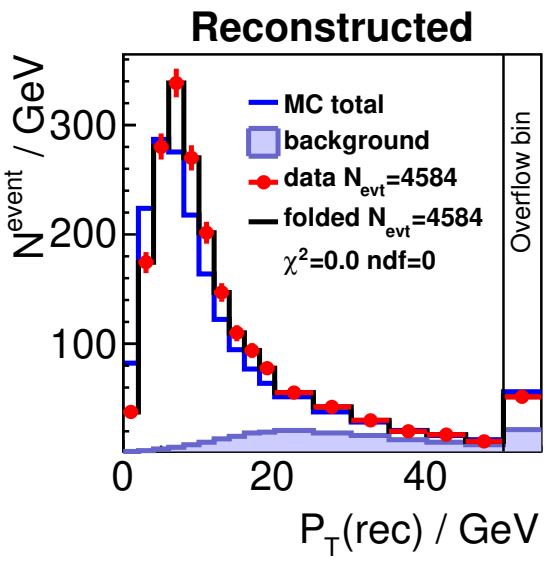
- Bin-by-bin correction factors $\hat{x}_i = (y_i - b_i)\frac{N_i^{\text{gen}}}{N_i^{\text{rec}}}$; disfavoured
  - Heavy biases due to the underlying MC truth
  - Yields the wrong normalization for the unfolded distribution
- Invert the response matrix $\hat{\boldsymbol{x}} = \boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})$
  - Only for square matrices, but always unbiased
  - Oscillation patterns (small determinants in matrix inversion)
  - Patterns also seen as large negative $\rho_{ij} \sim -1$ near diagonal
  - Result is correct within uncertainty envelope given by $\boldsymbol{V}_{xx}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

determinant





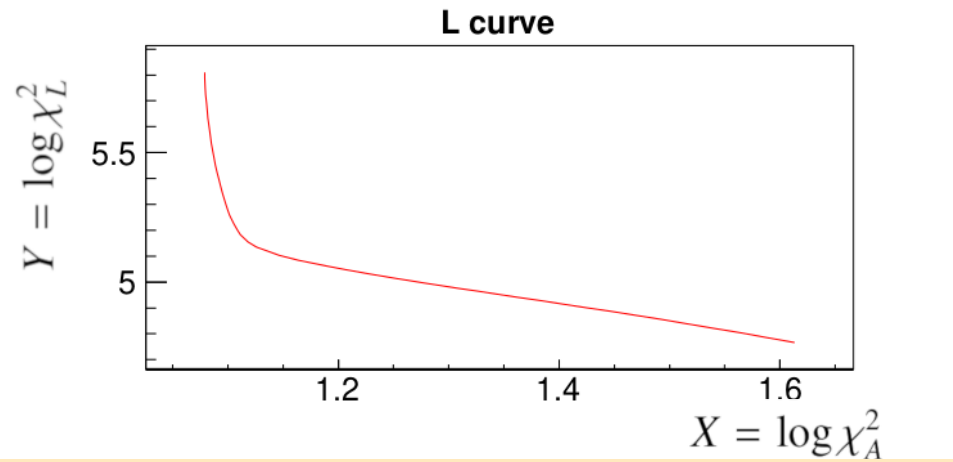


Cartoon from https://www.mathsisfun.com/algebra/matrix-inverse.html, plots from ArXiv:1611.01927

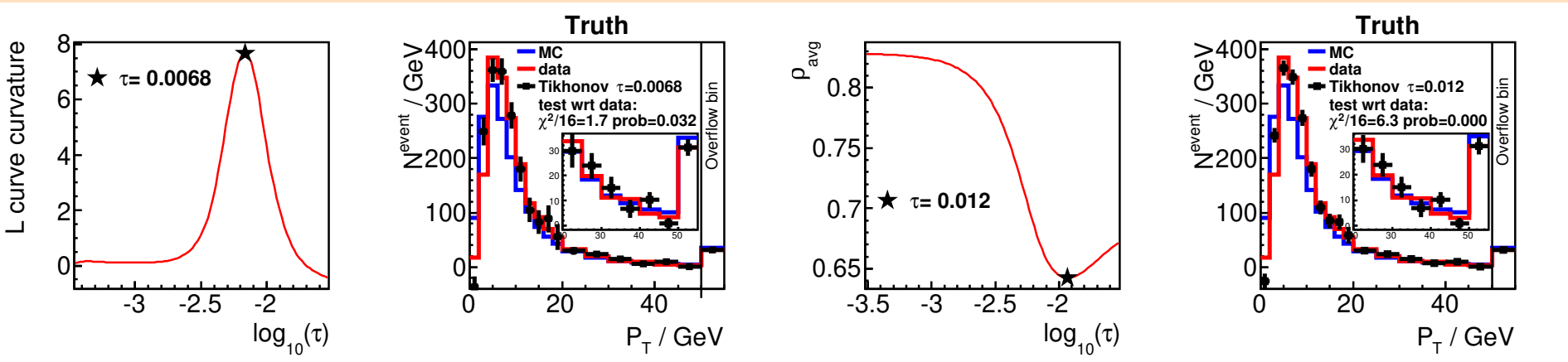
UCLouvain
Institut de recherche
en mathématique et physique

Unfolding: regularization 1/

$$\chi^2_{\text{TUnfold}} = \chi^2_A + \tau^2 \chi^2_L$$
$$\chi^2_A = (\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{b} - \boldsymbol{y})^{\mathsf{T}} (\boldsymbol{V_{yy}})^{-1} (\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{b} - \boldsymbol{y})$$
$$\chi^2_L = (\hat{\boldsymbol{x}} - \boldsymbol{x_B})^{\mathsf{T}} \boldsymbol{L}^{\mathsf{T}} \boldsymbol{L} (\hat{\boldsymbol{x}} - \boldsymbol{x_B})$$



- Choose τ corresponding to maximum curvature of L-curve
- Or minimize the global $\rho_{\text{avg}} = \frac{1}{M_x} \sum_{j=1}^{M_x} \rho_j$
  - Often results in stronger regularization than L-curve



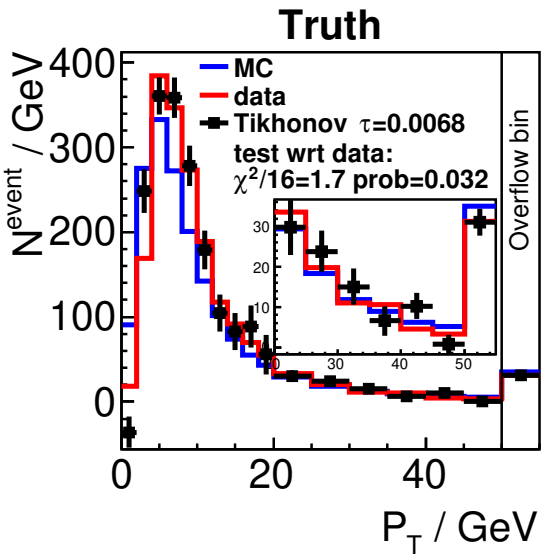
Plots from ArXiv:1611.01927

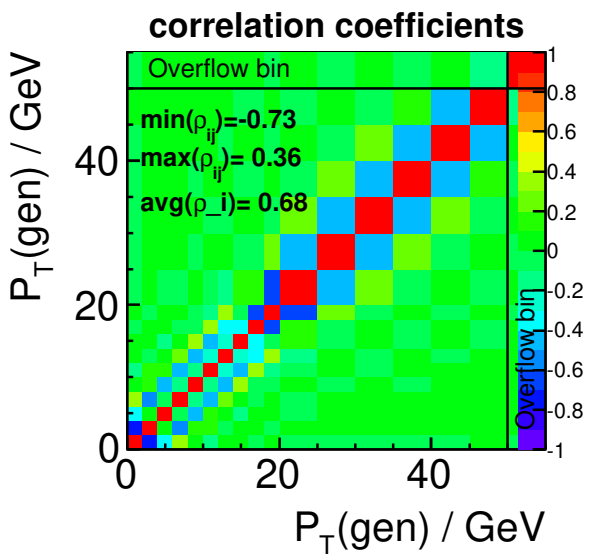
UCLouvain
Institut de recherche en mathématique et physique

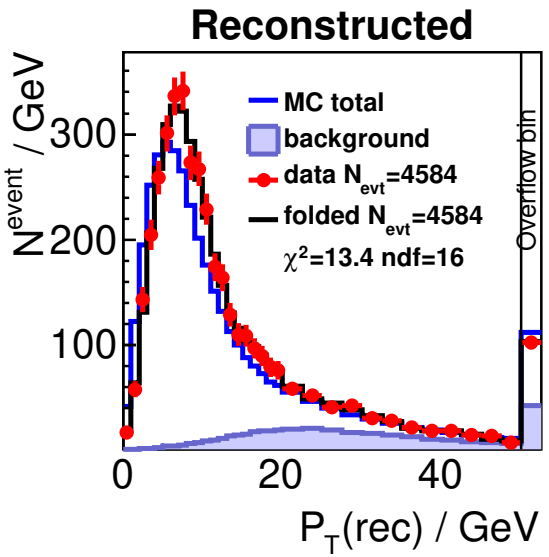
Unfolding: regularization 2/

$$\mathcal{L}(\mathbf{x},\lambda) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3,$$
$$\mathcal{L}_1 = (\mathbf{y}-\mathbf{A}\mathbf{x})^T \mathbf{V}_{\mathbf{yy}} (\mathbf{y}-\mathbf{A}\mathbf{x}),$$
$$\mathcal{L}_2 = \tau^2 (\mathbf{x}-f_b\mathbf{x_0})^T (\mathbf{L}^T\mathbf{L})(\mathbf{x}-f_b\mathbf{x_0}),$$
$$\mathcal{L}_3 = \lambda(Y - \mathbf{e}^T\mathbf{x}),$$
$$Y = \sum_i y_i,$$
$$e_j = \sum_i A_{ij}.$$

- $\mathbf{y}$: observed yields
- $\mathbf{A}$: response matrix
- $\mathbf{x}$: the unfolded result
- $\mathcal{L}_1$: least-squares minimization ($V_{ij} = e_{ij}/e_{ii}e_{jj}$ correlation coefficients)
- $\mathcal{L}_2$: regularization with strength τ
- Bias vector $f_b\mathbf{x_0}$: reference with respect to which large deviations are suppressed
- $\mathcal{L}_3$; area constraint (bind unfolded normalization to the total yields in folded space)







Plots from ArXiv:1611.01927

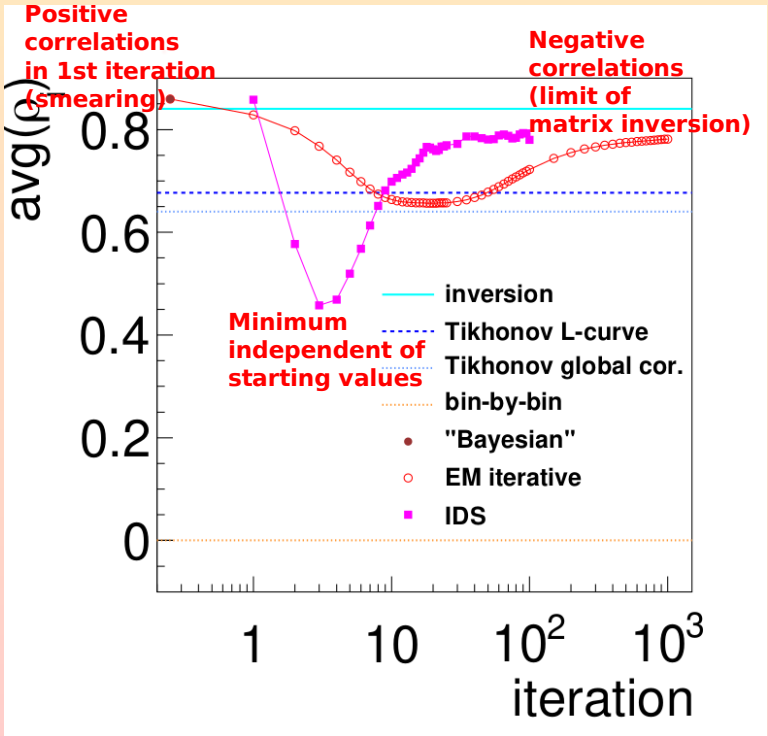
UCLouvain
Institut de recherche en mathématique et physique

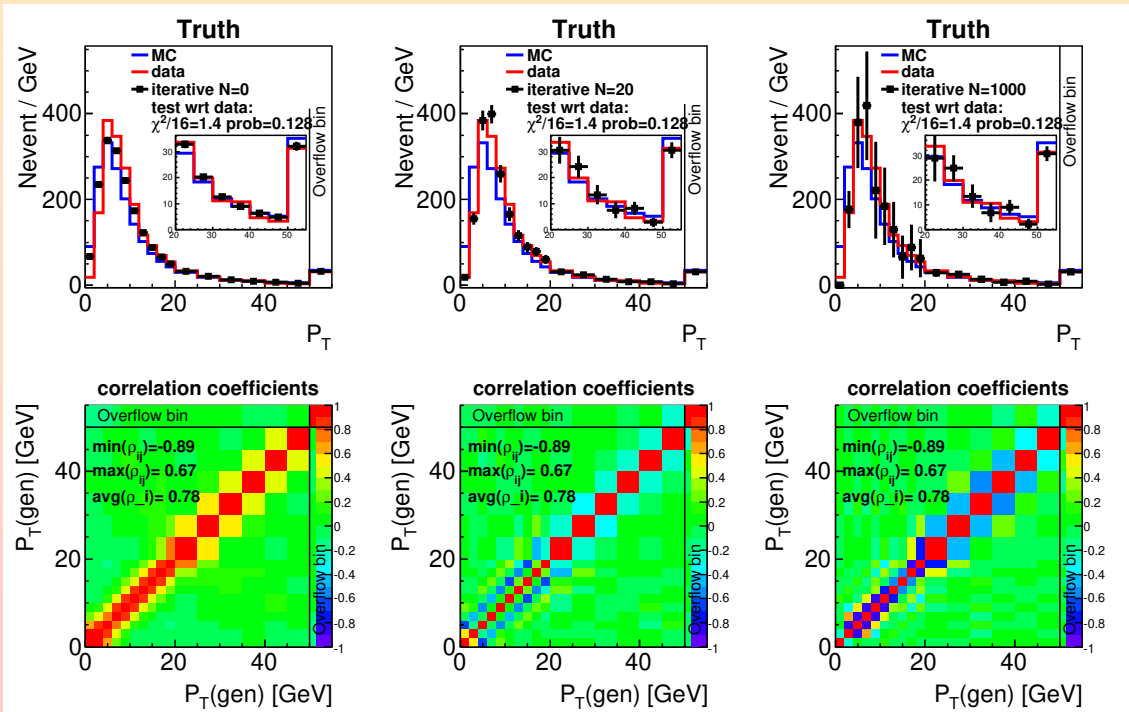
Unfolding: Iterative Unfolding

- Iterative improvement over the result of a previous iteration;

$$x_j^{(n+1)} = x_j^{(n)} \sum_{i=1}^{M} \frac{A_{ij}}{\epsilon_j} \frac{y_i}{\sum_{k=1}^{N} A_{ik} x_k^{(n)} + b_i}$$

  - It converges (slowly, $N_{iter} \sim N_{bins}^2$) to the MLE of the likelihood for independent Poisson-distributed y_i
  - Not necessarily unbiased for correlated data (does not make use of covariance of input data V_{yy})
- In HEP most people don't iterate until convergence
  - Fixed N_{iter} is often used; the dependence on starting values provides regularization
- Intrinsically frequentist method
  - for $N_{iter} \to \infty$ converges to matrix inversion, if all $\hat{x}_j$ from matrix inversion are positive
  - $N_{iter} = 0$ sometimes called improperly "Bayesian" unfolding (the author, D'Agostini, is Bayesian)
- Don't use software defaults!!! (e.g. some software has $N_{iter} = 4$)
  - Minimizing the global ρ is a good objective criterion, but there are others (Akaike information, etc)



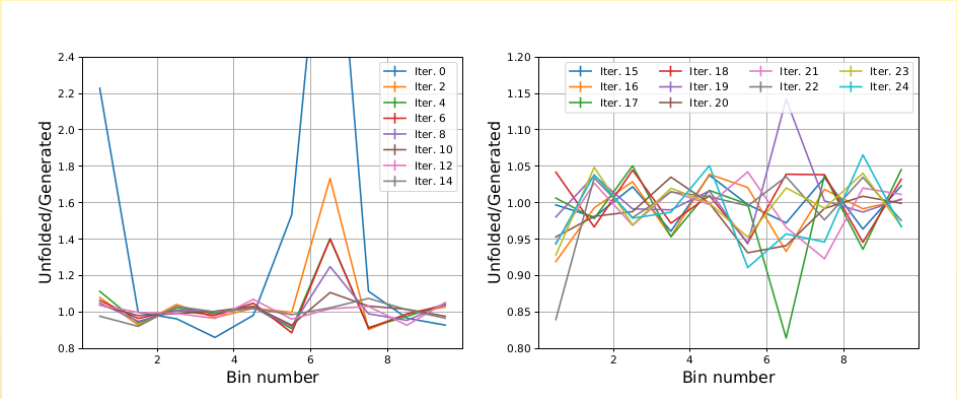


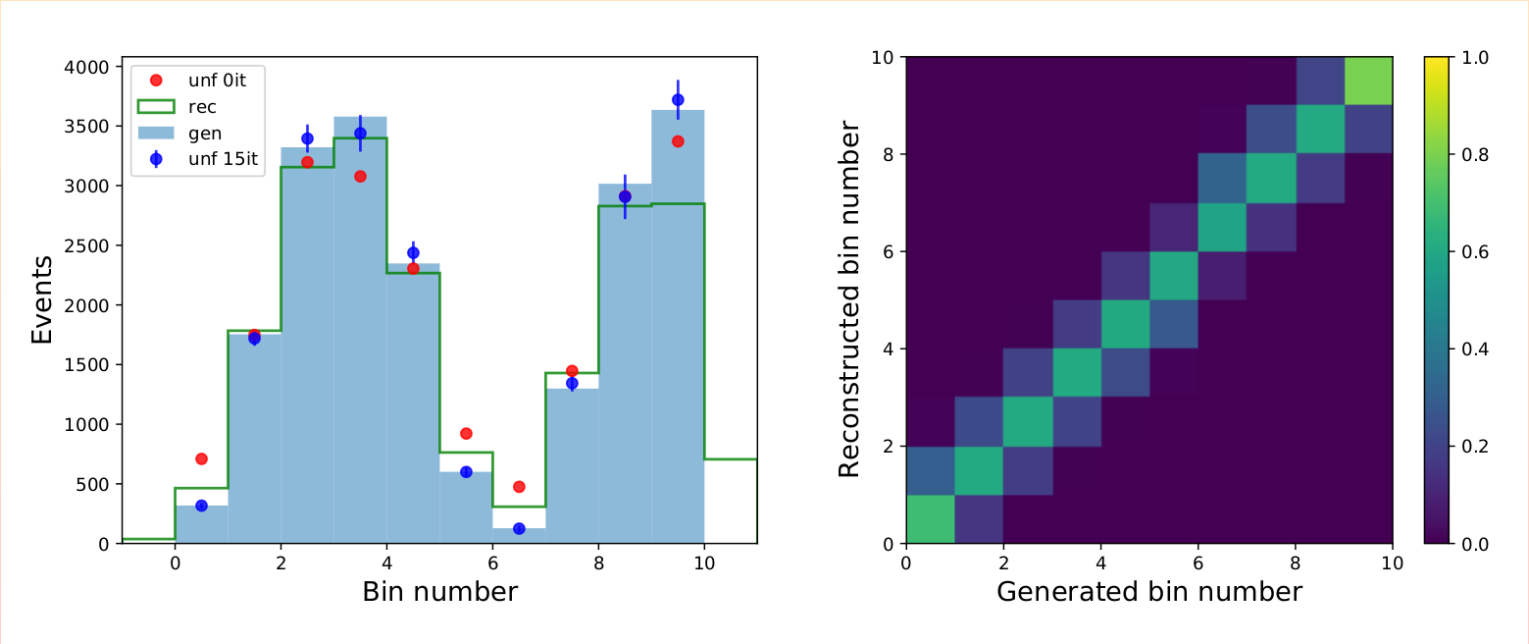
Plots from arXiv:1611.01927

UCLouvain
Institut de recherche
en mathématique et physique

Unfolding with ML

- 1712.01814 frames the problem as a classification one
  - Assign to each reconstructed data event the most probable truth bin
- Iterative process: train using flat inputs, apply to data to obtain updated binned distribution, use result to resample, retrain
- True space discretization regularizes the problem even before performing unfolding
- Coverage studies not (yet?) performed





Two proposed methods (shipped with the next arXiv posting)

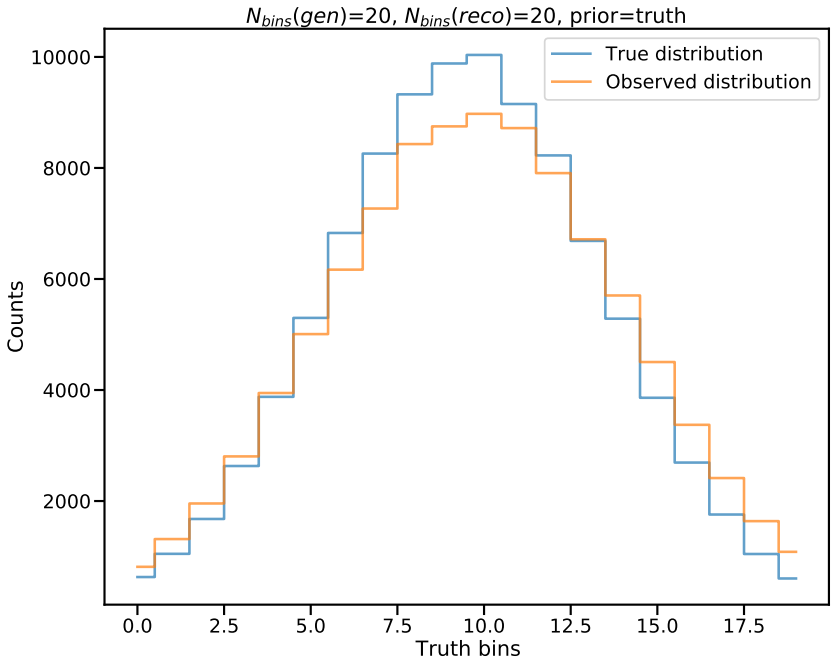
UCLouvain
Institut de recherche
en mathématique et physique

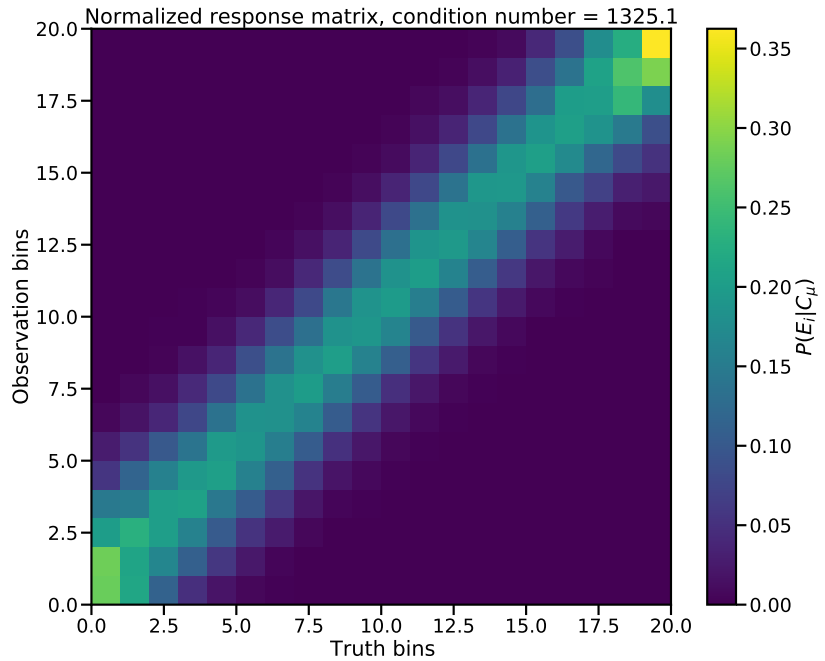
- Density estimation with normalizing flows
  - Estimate the transformation from true to smeared space (i.e. the response matrix) with masked autoregressive flows
  - Obtain an invertible transformation that can be used to transform arbitrary distributions
  - Unbinned method! Binning is performed only for displaying purposes (can use unbinned, or study regularization induced by binning)
  - Can unfold many variables at the same time (under study)
- Unfold without unfolding
  - Start from a distribution in true space (either a simulation, or a flat distribution)
  - Create proposal by resampling according to the statistical uncertainty in each bin
  - Fold the proposal using the untouched response matrix
  - Stochastically travel by accepting proposals if their folded realization is closer to the data than the current proposal
  - Purely binned method, further regularization schemes shown to be necessary

UCLouvain
Institut de recherche
en mathématique et physique

The scenarios

- Very simple scenarios based on Gaussian distributions smeared with random noise, 100k events per scenario
- Alternative data (representing scenarios where the data don't come from the assumed true distribution) generated by further smearing and shifting
- Different scenarios: less diagonal matrices will induce need for regularization)
  - Diagonality from max and min singular values in SVD decomposition of the response matrix $cond(R) := \frac{\sigma_{\max}}{max(0,\sigma_{\min})}$
- Different scenarios: $N_{bins}(gen)$ larger, smaller, or equal to $N_{bins}(reco)$

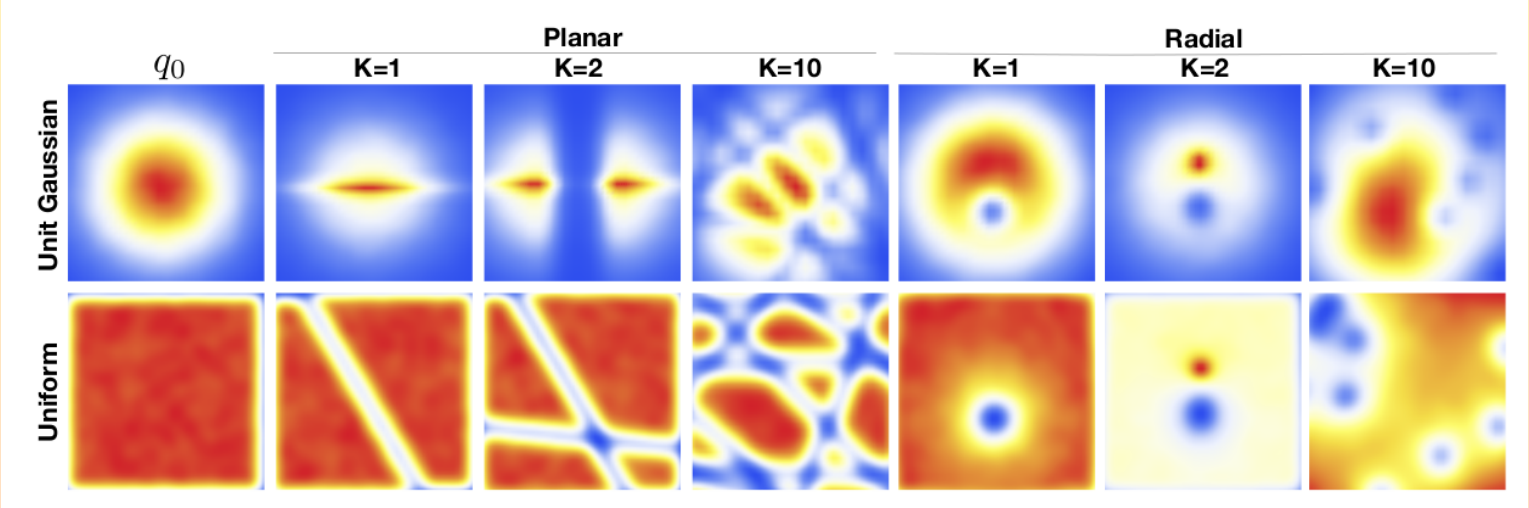


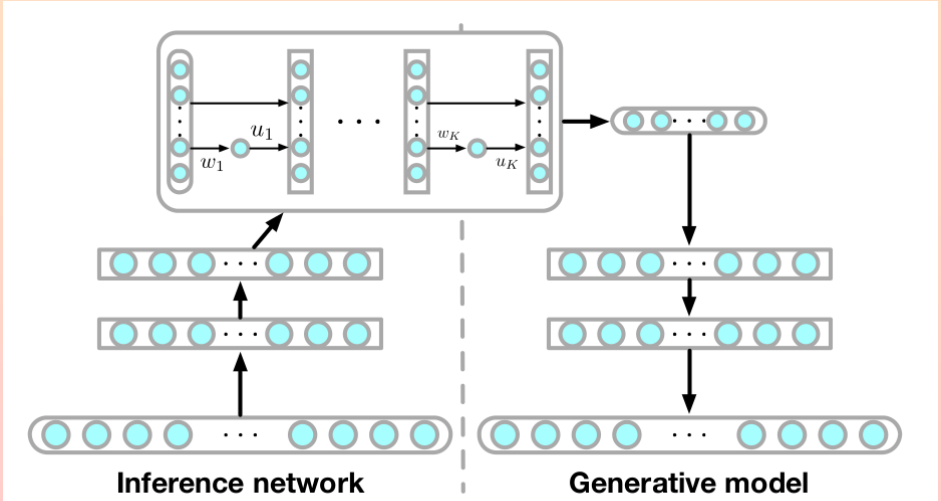


UCLouvain
Institut de recherche
en mathématique et physique

Normalizing flows

- Variational inference (Rezende and Mohamed, 1505.05770)
  - Start from a simple density
  - Apply a sequence of invertible transformations until the desired complexity is reached
- Combination of inference and generative model
  - Inference network maps observations to the parameters of the model
  - Generative model receives the posterior samples from the inference network at training time





UCLouvain
Institut de recherche
en mathématique et physique

Gaussian autoregressive models

- Decompose a joint density into a product of conditional densities
  - Condition on *previous* variables (time series, or lower-index coordinates for some ordering
  - Predicted value of features depend on past values of the same feature, rather than on other predictors
- Used for density estimation
  - Take some variable with some implicit ordering (e.g. tensor)
  - Output a mean and standard deviation for each element of the input, conditioned on previous elements
- In a sense, a Bayesian network

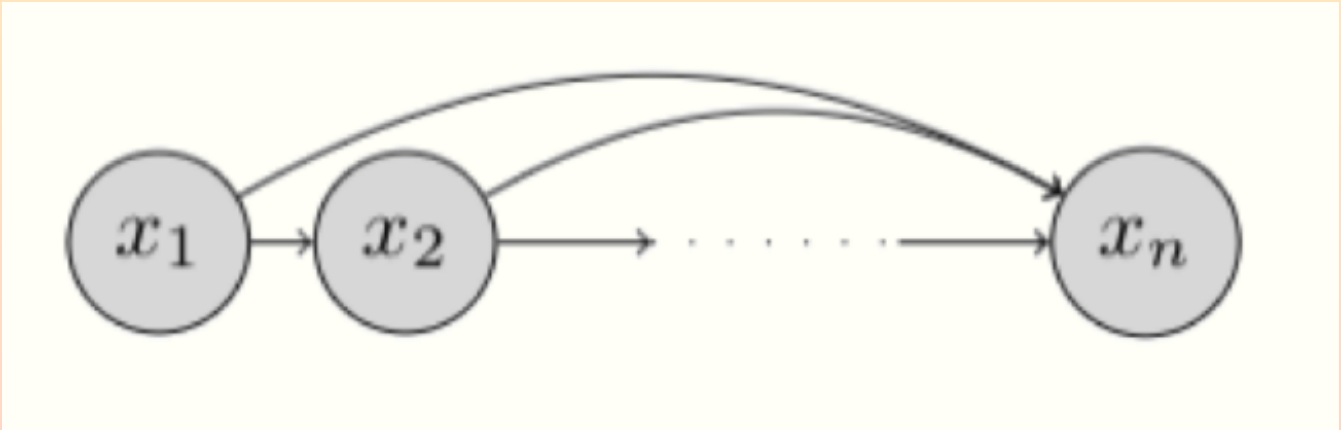
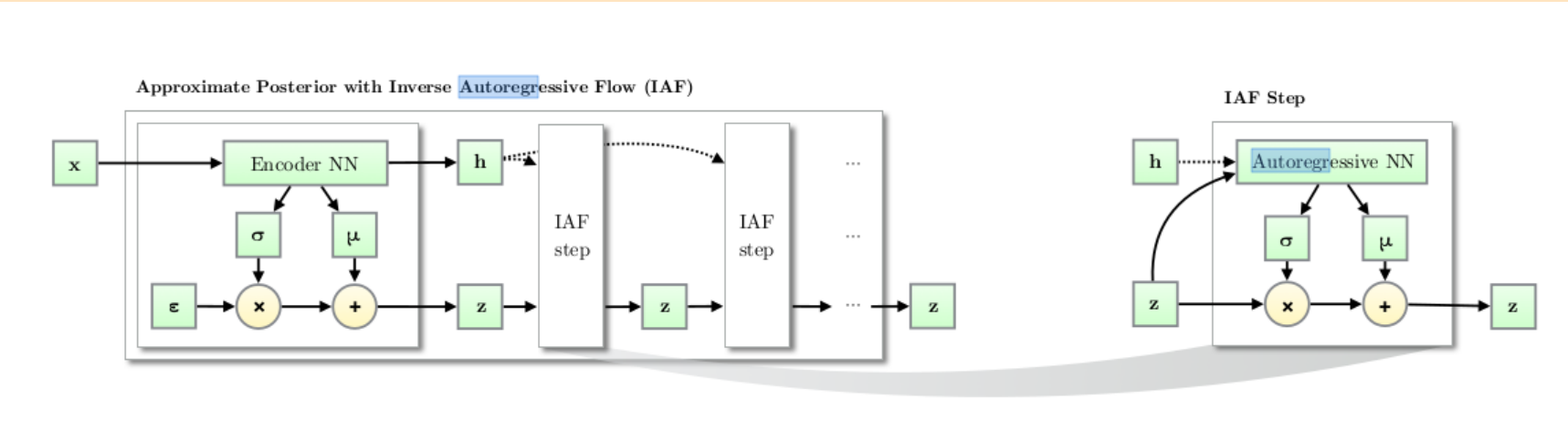


Image from https://deepgenerativemodels.github.io

UCLouvain
Institut de recherche
en mathématique et physique

Autoregressive models are normalizing flows (if invertible)!

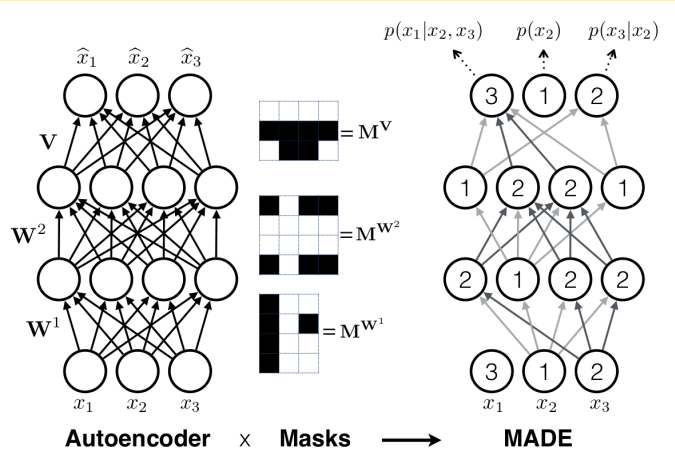
- Kingma and colleagues 1606.04934
  - Autoregressive models result in tranformations that by construction have a tractable Jacobian
  - If the transformation is also invertible, then the model is equivalent to a normalizing flow
- Can therefore be used for density estimation, e.g. Inverse Autoregressive Flow
- Decompose the operation of sampling from a posterior (typical Bayesian operation, e.g. in MCMC or ABC)
  - Sample from a very simple gaussian distribution
  - Apply chain of nonlinear invertible transformations
  - Obtain posterior sample

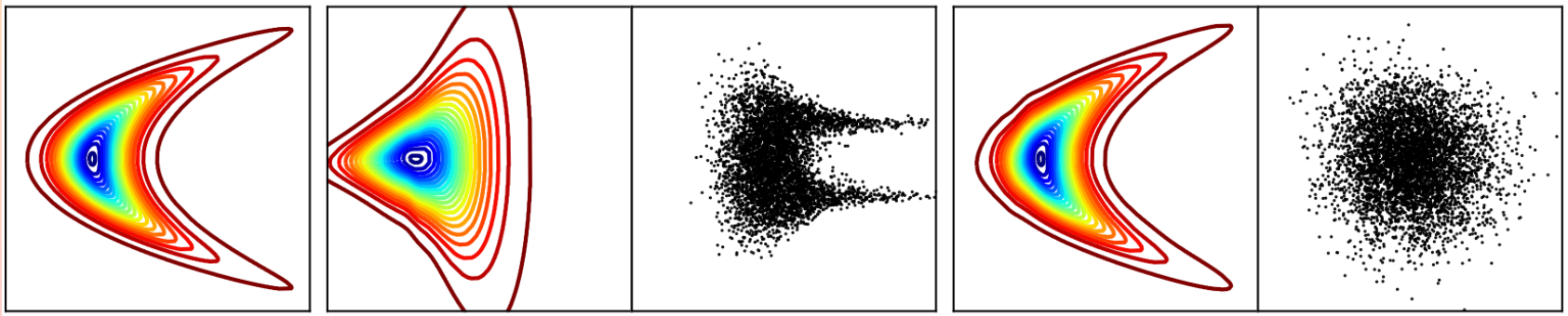


UCLouvain
Institut de recherche
en mathématique et physique

MADE and MAF

- MADE Germain and colleagues, 1502.03509
  - Masked Autoencoders for Distribution Estimation
  - Vectorized architecture for density estimation based on autoencoders
  - Fast and reliable
  - *Masked*: deactivate inputs to enforce autoregressive model!
- Masked Autoregressive Flows (MAF), Papamakarios and colleagues, 1705.07057
  - If you stack several autoregressive models that each learn a conditional density, you obtain a normalizing flow!





(a) Target density (b) MADE with Gaussian conditionals (c) MAF with 5 layers

Learn response matrix with MAF

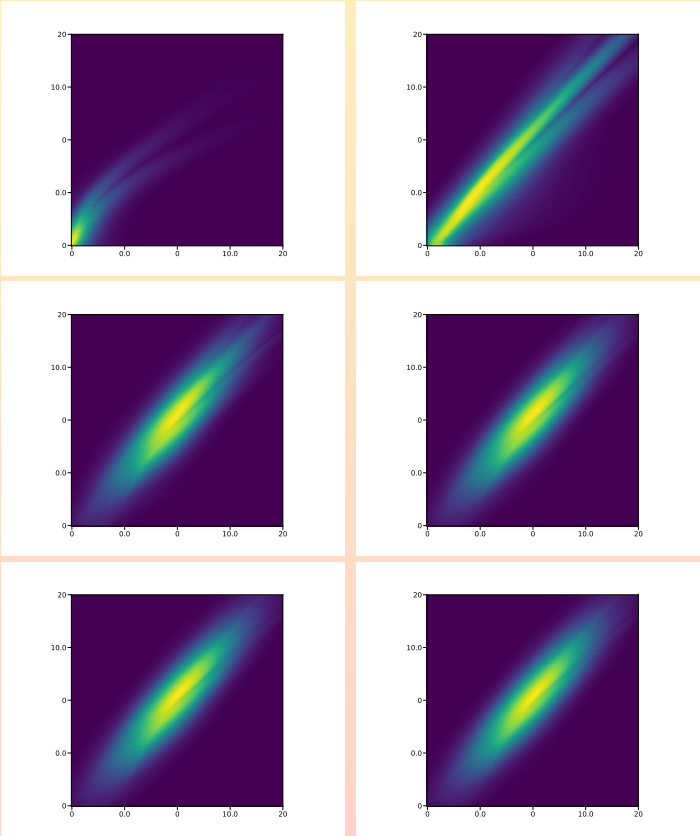
UCLouvain
Institut de recherche
en mathématique et physique

- MAF composed of 10 MADE layers, with permutation after each one
  - Each model is essentially a copy of the first set of coordinates and a transformation of the second set of coordinates
  - Permutation ensures that in the next step the transformed coordinates are those that were simply copied at the previous step
- I use a simple Gaussian distribution as an input to the learning process

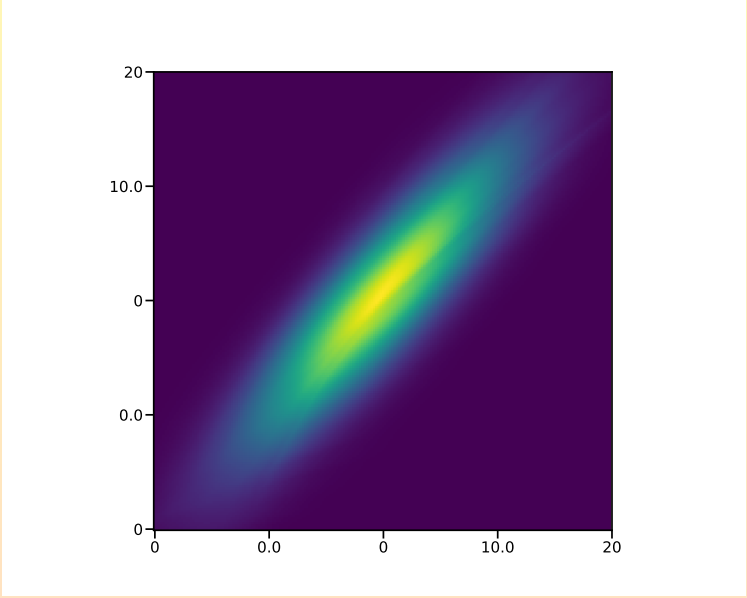
UCLouvain
Institut de recherche
en mathématique et physique

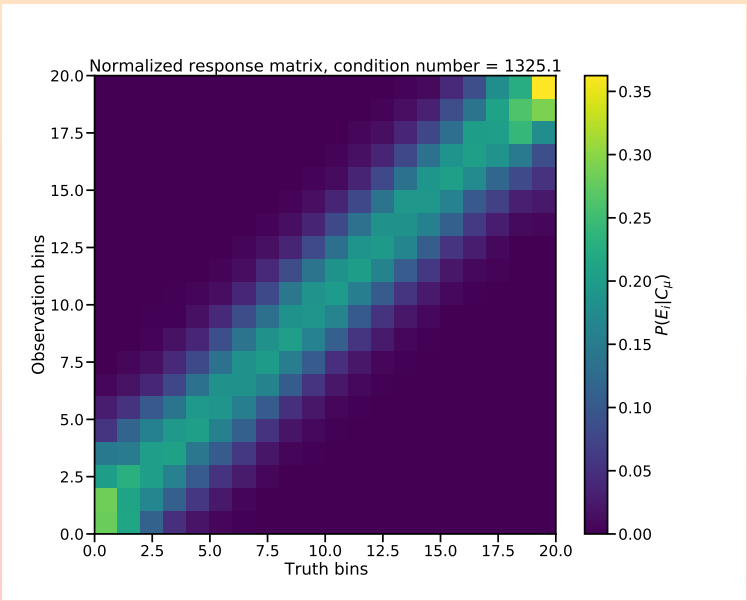
Learned response matrix

- 100 epochs (if early stopping, regularization)
- Density at epoch 0, 10,.. 50



- Final learned density: need to improve at boundaries

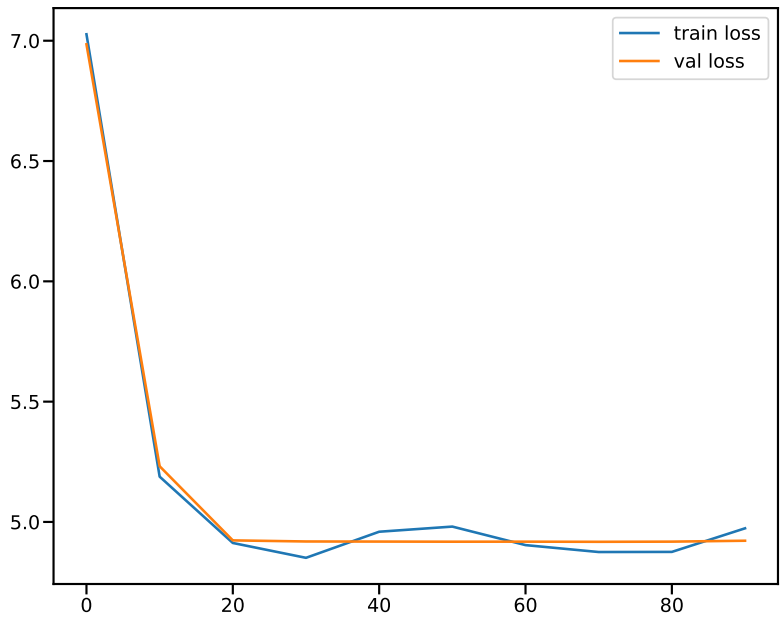


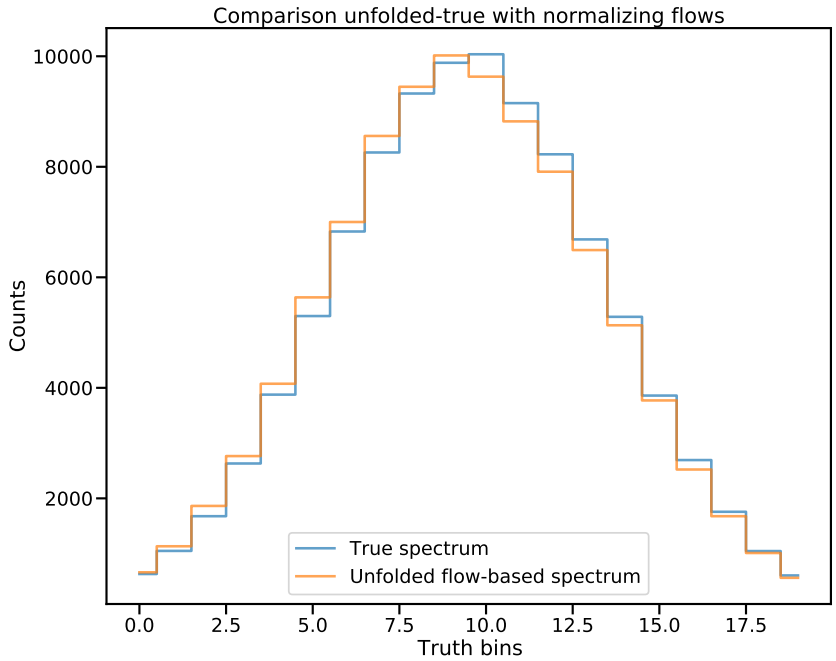


UCLouvain
Institut de recherche
en mathématique et physique

Use inverted MAF transformation as inverse of response matrix

- Unfolding by transforming data using the MAF density suffers from the density not completely learned
- More studies ongoing with more complex models (gaussian-to-gaussian might be too simple) and using manifold learning





Unfolding by folding

UCLouvain
Institut de recherche
en mathématique et physique

- Why should we invert the response matrix, when we can resample?
  - Start from binned distributions (but implicit regularization!)
  - We can resample as many times as we want in the true space
  - Assume a distribution, then resample assuming a Poisson pdf in each bin
  - Fold the resulting distribution, and compare it with data with a statistical test
  - If the test hints at better compatibility, accept proposal, otherwise keep the current one
- Turned out that the general idea was also developed by Landweber, 1951, but with gradient descent
- My proposal is then essentially a stochastic optimization version of the Landweber method

UCLouvain
Institut de recherche
en mathématique et physique

Algorithm

- Can set threshold on the compatibility test for early stopping (regularization)
- Algorithm rather robust with respect to changes of statistical test
  - Pearson's χ^2 and its *minbin* modification by Porter, Bhattacharyya distance, Kullback-Leibler divergence, Wasserstein Distance etc
- Uncertainties from building a full distribution using best proposals (similar to ABC modification under study)

Algorithm 1: The algorithm to resample in the truth-level space until the folded distribution of the sample coincides with the data distribution. I have to put an early stopping and remove the maximum number of iterations.

input : A truth distribution $\boldsymbol{x}_s$;
the data distribution $\boldsymbol{y}$;
the response matrix R;
the number of iterations $N_{samples}$;
a statistical test *test* and an optional condition *oc* for similarity between histograms;
an optional early-stopping condition *sc*.

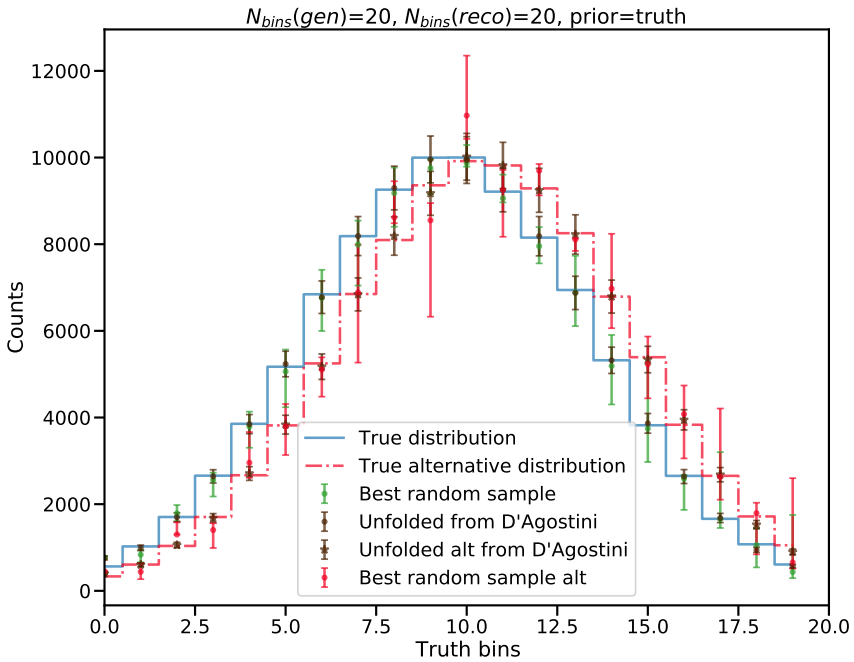
output: The random sample $\hat{\boldsymbol{x}}_{best}$ whose folded version matches the better with the data distribution.

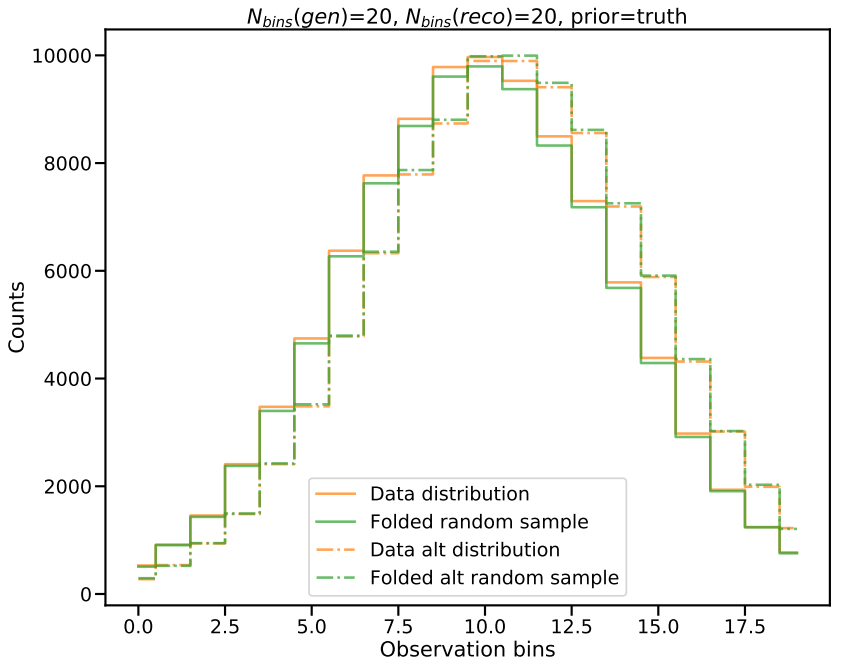
$\hat{\boldsymbol{x}}_{best} \leftarrow \boldsymbol{x}_s$;
$TS_{best} \leftarrow 10^5$;
for $i \leftarrow 0$ **to** $N_{samples}$ **do**
  $\boldsymbol{x}_i \sim Pois(\boldsymbol{x}_{best})$;
  $\boldsymbol{y}_i \leftarrow R\boldsymbol{x}_i$;
  $TS \leftarrow test(\boldsymbol{y}_i, \boldsymbol{y})$;
  if $TS < TS_{best}$ *and* *oc* **then**
    $\hat{\boldsymbol{x}}_{best} \leftarrow \boldsymbol{x}_i$;
    $TS_{best} \leftarrow TS$;
  end
  if *sc* **then**
    return
  end
end

Dependence on parameters

UCLouvain
Institut de recherche
en mathématique et physique

- Use d'Agostini unfolding as a reference for comparison
- The method performs very well for diagonal matrices...

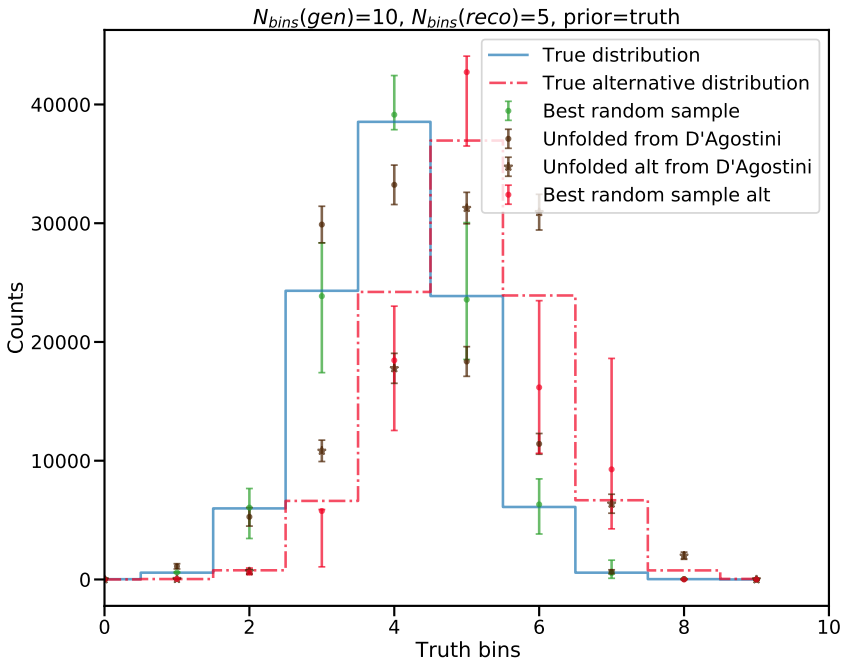


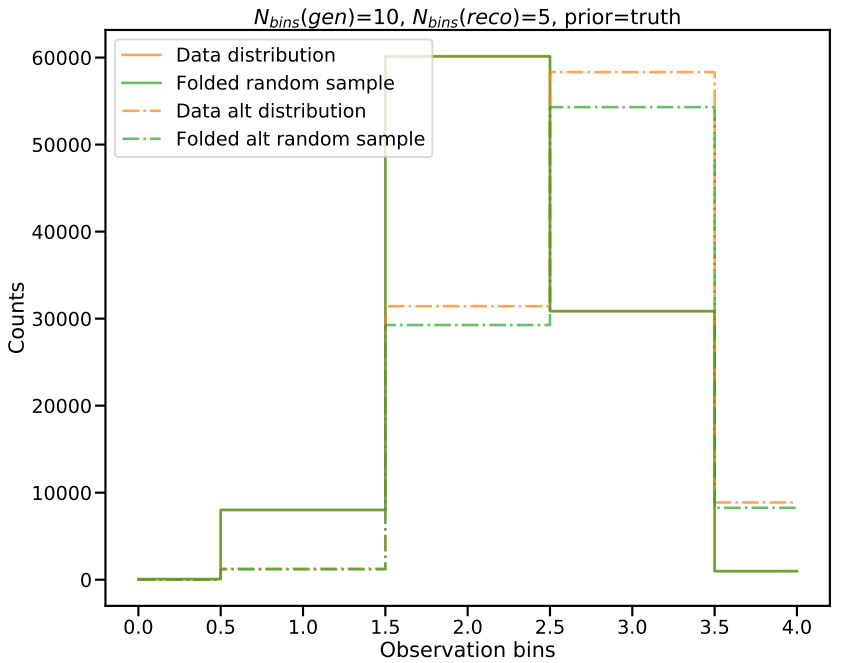


Dependence on parameters

UCLouvain
Institut de recherche
en mathématique et physique

- ...and particularly well for usually ill-defined problems ($N_{bins}(gen) > N_{bins}(reco)$!!!

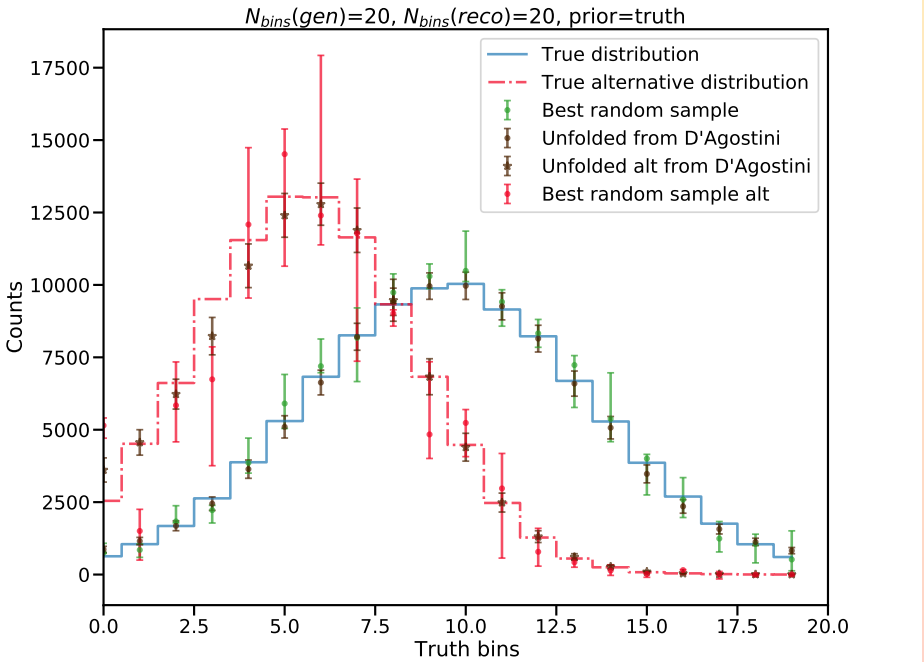


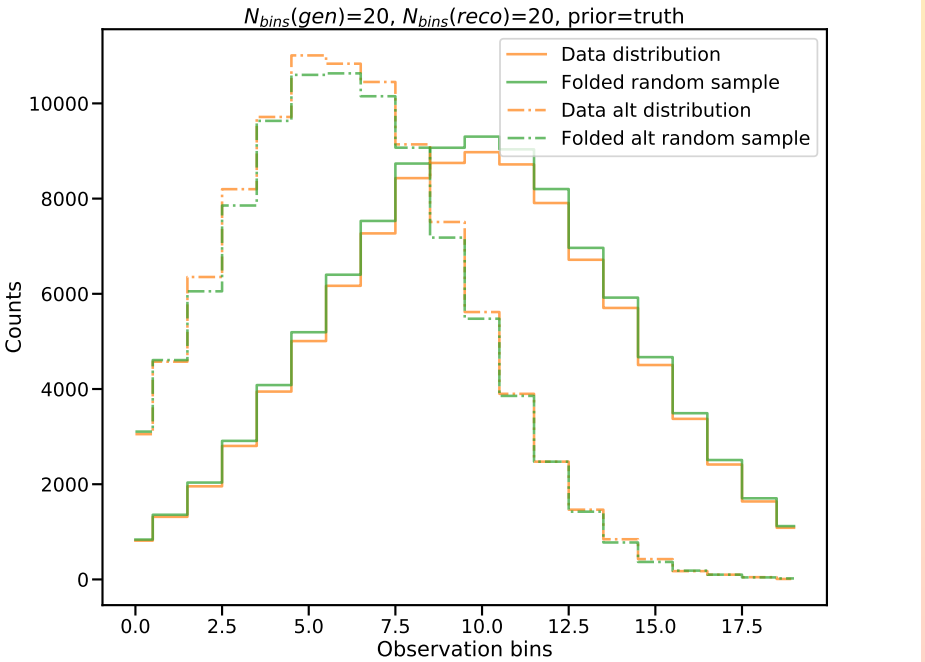


UCLouvain
Institut de recherche
en mathématique et physique

Dependence on parameters

- For less diagonal cases, further regularization is needed
  - More aggressive early stopping
  - Smoothness conditions on the proposals
  - Rebinning (unfold with finer bins, then rebin to larger bins)
- Coverage to be studied!





Bottom line test

UCLouvain
Institut de recherche
en mathématique et physique

- Comparison between imperfect data and folded truth in folded space, and between best random from imperfect data and truth-level truth in unfolded space
- The idea is that the inference in the unfolded space should not be better than that in the folded space (Cousins and colleagues, 1607.07038
- Here using the Pearson χ^2, divided by $NDOF$
  - For the moment, the regular NDOF is used, its computation neglecting the covariance matrix.

Configuration	Algo1 $\chi^2/NDOF_{foldedSpace}$ $(altData, foldedTruth)$		Algo1 $\chi^2/NDOF_{unfoldedSpace}$ $(bestRandomFromAltData, truth)$		D'Agostini $\chi^2/NDOF_{unfoldedSpace}$ $(d'Agostiniunfolded, truth)$	
	init:truth	init:flat	init:truth	init:flat	init:truth	init:flat
$N_{bins}(gen) = 20, N_{bins}(reco) = 20$	280.753	280.753	306.196	322.679	307.210	307.210
$N_{bins}(gen) = 10, N_{bins}(reco) = 20$	285.260	285.260	592.488	596.413	603.189	603.189
$N_{bins}(gen) = 20, N_{bins}(reco) = 10$	542.366	542.366	302.102	358.552	304.772	304.772

Perspectives for uncertainties (ABC and MCMC)

UCLouvain
Institut de recherche
en mathématique et physique

- Uncertainties in the response matrix dealt with as commonly in particle physics
  - Vary response matrix by applying systematic shifts to it (e.g. change jet energy scale, obtain new response matrix)
  - Re-unfold using the alternative matrix
  - Difference between the unfolded spectra is taken as the uncertainty
- Uncertainty of the method
  - Taking the samples that are better compatible (when folded) with the data is questionable
  - Switching to a full ABC implementation
  - Testing a full MCMC-based unfolding version

UCLouvain
Institut de recherche
en mathématique et physique

Summary

- Briefly outlined the unfolding problem and the current standard methods
- Proposed two methods, which will appear on the arxiv in a couple days
- Unfolding with normalizing flows
  - Learn an invertible function representing the response matrix
  - Use it to convert smeared distributions to unfolded distributions
  - No binning needed
  - Can unfold multiple observables at the same time!
- Unfolding by folding
  - Build proposals by sampling in true space, folding them, and comparing with data
  - Stochastic optimization to find best solution
  - Build distribution to estimate uncertainty, or use full ABC/MCMC
  - Implicitly regularized (samples binned distributions), more regularization needed for non-diagonal cases
- Performing bias and coverage studies for both methods

Two Approaches to Unfolding: Invert the Problem using Normalizing Flows, or Stochastically Fold without Actually Inverting any Matrix

Pietro Vischia

Institut de recherche en Mathématique et Physique,
Université catholique de Louvain
pietro.vischia@cern.ch

UCLouvain
Institut de recherche
en mathématique et physique

THANK YOU VERY MUCH FOR ATTENDING!!

UCLouvain
Institut de recherche
en mathématique et physique

THANKS FOR THE ATTENTION!

UCLouvain
Institut de recherche
en mathématique et physique

Backup