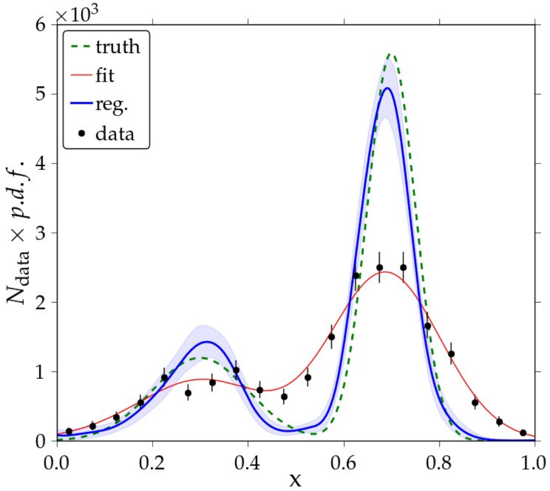
ARU – Towards optimal unfolding

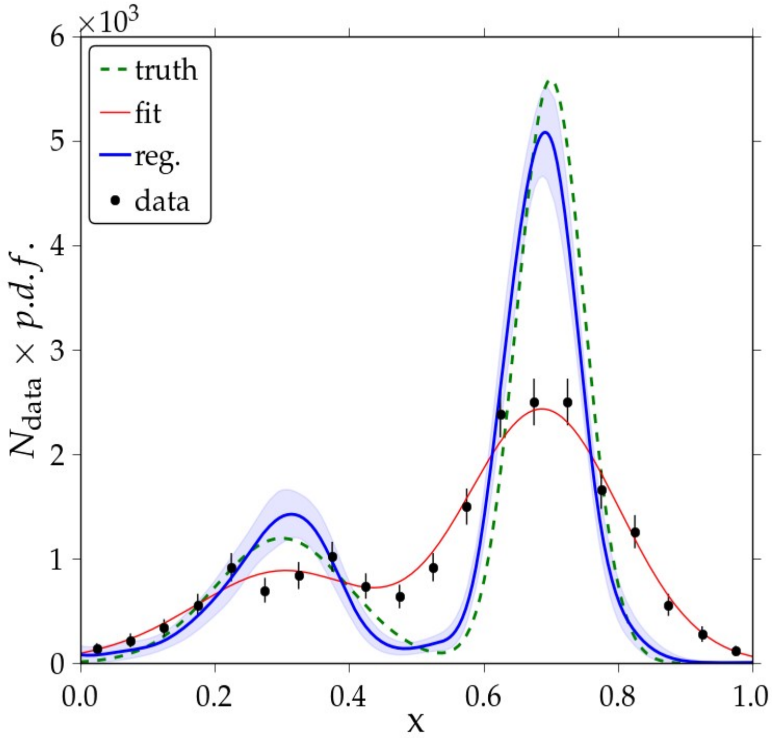
Hans Dembinski, Markus Roth
IEKP, KIT Karlsruhe



PHYSTAT 2011
CERN, Geneva
January, 20th, 2011

Overview

- ARU
  - New algorithm derived from RUN
  - "Automatic" unfolding
  - Un-binned analysis
  - Analytical uncertainty calculation
- This talk
  - Concepts and status
  - MC study



Motivation

Common unfolding chain

what kind of regularization?
how to choose regularization weight?

raw data → histogram → unfolding → unfolded histogram

how to preserve relevant structure?
bins with zero entries?

binned result
often misunderstood
by non-experts

Philosophy: Unfolding should...

be based entirely on statistical concepts (invariant to u → f(u))
→ regularization based on **entropy (statistical information)**

not remove information
→ don't bin at all, use an **un-binned analysis**

be straight forward for non-experts and "objective"
→ **avoid user interaction**

be comprehensible

ARU

ARU is a regularized fit and very similar to RUN (Blobel, 1984)

Only user input: Kernel function and raw data

Statistical term

fit folded solution to data

Solution

suppress fluctuations

Regularization term with RUN-like weight

Flexible function

B-splines

$$f(y) = \int \mathrm{d}x \, K(y,x) \, g(x)$$

User-supplied kernel function

(resolution, efficiency, non-linear calibration)

Maximimize Neg-Log-Likelihood function

$$L(\mathbf{c}) = L_1(\mathbf{c}) + w L_2(\mathbf{c})$$

Statistical term

“Extended likelihood”, also fits normalization of *g(y)*

Regularization term

Variant of Kullback-Leibler divergence

Flexible function

$$f(y) = \int \mathrm{d}x\, K(y, x)\, g(x)$$

Flexible function
how to implement?

Express f(x) by a sum of scaled basis functions $g(x) = \sum_k c_k\, b_k(x)$

Exploit linearity $f(y) = \sum_k c_k \underbrace{\int \mathrm{d}x\, K(y, x)\, b_k(x)}_{\text{compute numerically}} = \sum_k c_k\, f_k(y)$

RUN's approach

B-splines $b_k(x)$

well established in approximation theory

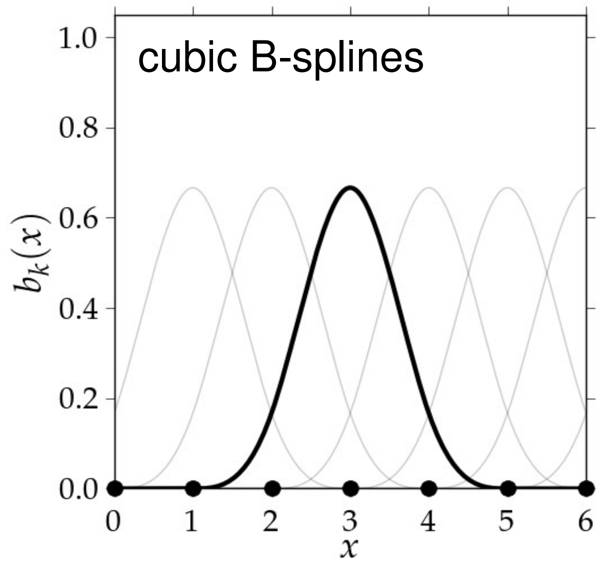
local → do not suffer from Runge phenomenon

easy to adapt to kernel resolution via knot intervals

evaluation fast and numerical stable

easy to enforce $g(x) \geq 0$

analytic integral and derivatives



Statistical term

Karlsruhe Institute of Technology

Standard unbinned max-likelihood fit ignores normalization

Extended likelihood method also includes normalization v(**c**)

$$v(\mathbf{c}) = \int \mathrm{d}y\, f(y) = \sum_k c_k\, F_k \quad \text{with} \quad F_k = \int_{-\infty}^{\infty} \mathrm{d}y\, f_k(y)$$

standard likelihood

$$\ln P(\mathbf{y}|\hat{f}(y)) = \sum_i \ln \hat{f}(y_i|\mathbf{c}) + \text{const.}$$

and

Poisson

$$P(N|v) = \frac{v^N\, e^{-v}}{N!}$$

$$L_1(\mathbf{c}) = -\sum_i \ln \frac{\sum_k c_k f_k(y_i)}{v(\mathbf{c})} - N \ln v(\mathbf{c}) + v(\mathbf{c}) + \text{const.}$$

$$= -\sum_i \ln \sum_k c_k f_k(y_i) + v(\mathbf{c}) + \text{const.}$$

Regularization term

Karlsruhe Institute of Technology

Variant of Kullback-Leibler divergence

$$L_2(\mathbf{c}) = \int \mathrm{d}x\, g(x)\, \ln \frac{g(x)}{h(x)} - v(\mathbf{c})$$

h(x) is **corrected** kernel density estimate of raw data
corrects for calibration and efficiency, but not for resolution

h(x) is best a priori guess about solution

Features

principle of minimum discriminant information

if h(x) = 1 → principle of maximum entropy (minimal information) (Schmelling, 1994)

pulls solution towards h(x) → regularization! $f(x|\mathbf{c}) \to h(x) : L_2 \to 0$

integrand depends on ratio g(x)/h(x)
→ g(x) constrained more where h(x) is small (less data)

if K(y,x) = δ(y,x) (no smearing), impact of regularization vanishes and unfolding reduces to standard fit

Summary: likelihood function

minimize for given weight *w* $$L(\mathbf{c}) = L_1(\mathbf{c}) + w L_2(\mathbf{c})$$

$$L_1(\mathbf{c}) = -\sum_i \ln f(y_i) + v(\mathbf{c})$$

$$L_2(\mathbf{c}) = \int \mathrm{d}x\, g(x)\, \ln \frac{g(x)}{h(x)} - v(\mathbf{c})$$

$$g(x) = \sum_k c_k\, b_k(x)$$

$$f(y) = \sum_k c_k \int \mathrm{d}x\, K(y,x)\, b_k(x) = \sum_k c_k\, f_k(y)$$

$$v(\mathbf{c}) = \int \mathrm{d}y\, f(y) = \sum_k c_k\, F_k$$

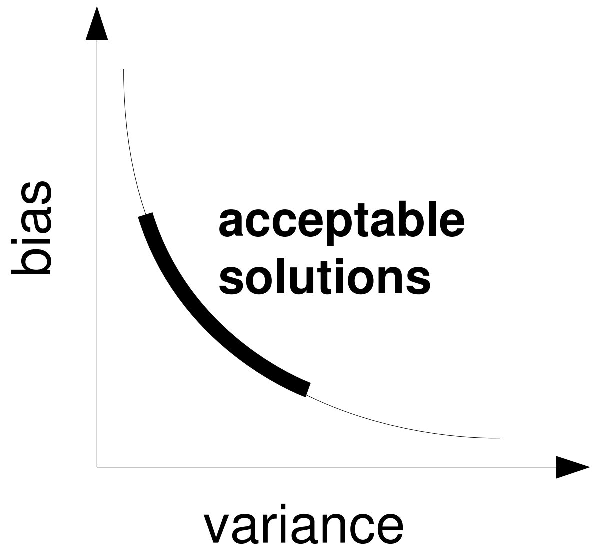
integral calculated numerically, but analytical treatment seems possible

Analytical derivatives available, good for numerical minimization with MINUIT 2

$$h_{1\ell} = \frac{\partial L_1}{\partial c_\ell} \qquad h_{2\ell} = \frac{\partial L_2}{\partial c_\ell} \qquad H_{1\ell m} = \frac{\partial^2 L_1}{\partial c_\ell\, \partial c_m} \qquad H_{2\ell m} = \frac{\partial^2 L_2}{\partial c_\ell\, \partial c_m}$$

both H_1 and H_2 are positive definite → **fit always converges to unique solution**

Weight criterion



there is no *single* optimal weight

bias cannot be avoided

Philosophy: prefer bias over artificial features added by unfolding

ARU uses RUN's weight criterion

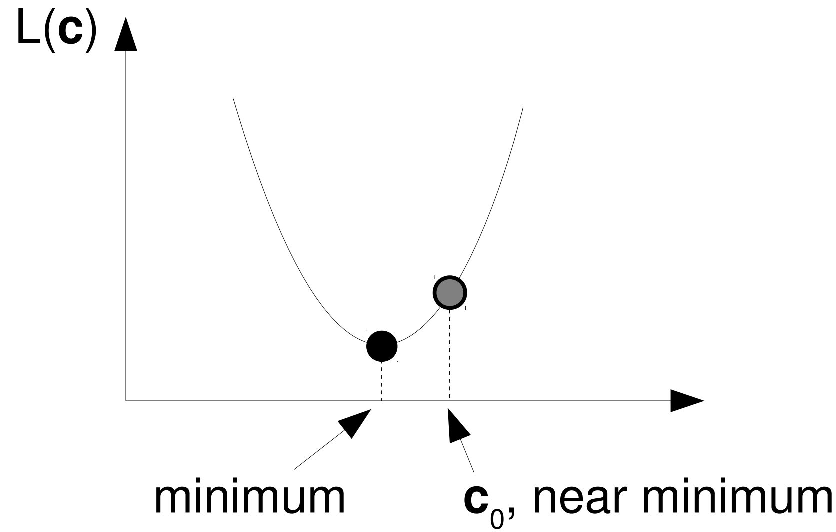
i) Transform to eigensystem of solution with **diagonal covariance matrix**

ii) Choose weight so that regularization dampens all coefficients that are **not significantly different from zero** (noise)

Regularization effect

Taylor expansion around minimum

$$L(\boldsymbol{c}) \simeq L(\boldsymbol{c}_0) + \boldsymbol{c}^T(\boldsymbol{h}_1 + w\boldsymbol{h}_2) + \frac{1}{2}(\boldsymbol{c}-\boldsymbol{c}_0)^T\boldsymbol{H}_1(\boldsymbol{c}-\boldsymbol{c}_0) + w\frac{1}{2}(\boldsymbol{c}-\boldsymbol{c}_0)^T\boldsymbol{H}_2(\boldsymbol{c}-\boldsymbol{c}_0)$$



After dropping all constant terms

$$L(\boldsymbol{c}) = \boldsymbol{c}^T(\boldsymbol{h}_1 + w\boldsymbol{h}_2 - (\boldsymbol{H}_1 + w\boldsymbol{H}_2)\,\boldsymbol{c}_0) + \frac{1}{2}\boldsymbol{c}^T\boldsymbol{H}_1\boldsymbol{c} + w\frac{1}{2}\boldsymbol{c}^T\boldsymbol{H}_2\boldsymbol{c}$$

Transform into eigensystem with $\bar{\boldsymbol{c}}^T = \boldsymbol{c}^T\boldsymbol{M}^T = \boldsymbol{c}^T\boldsymbol{U}_1^T\boldsymbol{D}_2^{-1/2}\boldsymbol{U}_2^T$

$$\boxed{L(\bar{\boldsymbol{c}}) = \bar{\boldsymbol{c}}^T\boldsymbol{M}^T(\boldsymbol{h}_1 + w\boldsymbol{h}_2 - (\boldsymbol{H}_1 + w\boldsymbol{H}_2)\,\boldsymbol{c}_0) + \frac{1}{2}\bar{\boldsymbol{c}}^T\boldsymbol{S}\,\bar{\boldsymbol{c}} + w\frac{1}{2}\bar{\boldsymbol{c}}^T\bar{\boldsymbol{c}}}$$

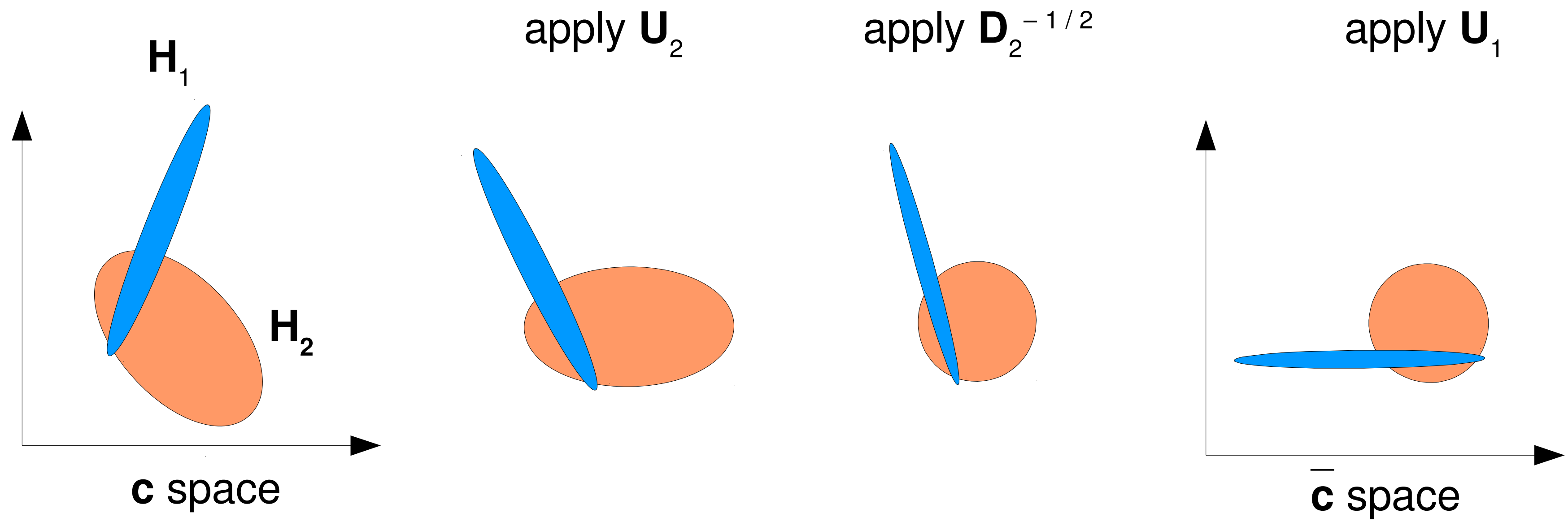
Transformations

Use two rotations and one scaling transformation to turn $\mathbf{H}_2$ to $\mathbf{1}$ and diagonalize $\mathbf{H}_1$

$$\boldsymbol{D}_2^{-1/2}\boldsymbol{U}_2^T\boldsymbol{H}_2\boldsymbol{U}_2\boldsymbol{D}_2^{-1/2} = \boldsymbol{D}_2^{-1/2}\boldsymbol{D}_2\boldsymbol{D}_2^{-1/2} = \boldsymbol{1}$$

$$\boldsymbol{U}_1^T\boldsymbol{D}_2^{-1/2}\boldsymbol{U}_2^T\boldsymbol{H}_1\boldsymbol{U}_2\boldsymbol{D}_2^{-1/2}\boldsymbol{U}_1 = \boldsymbol{U}_1^T\tilde{\boldsymbol{H}}_1\boldsymbol{U}_1 = \boldsymbol{S}$$

Geometrical interpretation



Solution in eigensystem

$$L(\bar{\boldsymbol{c}}) = \bar{\boldsymbol{c}}^T \boldsymbol{M}^T (\boldsymbol{h}_1 + w\boldsymbol{h}_2 - (\boldsymbol{H}_1 + w\boldsymbol{H}_2)\,\boldsymbol{c}_0) + \frac{1}{2}\bar{\boldsymbol{c}}^T \boldsymbol{S}\,\bar{\boldsymbol{c}} + w\frac{1}{2}\bar{\boldsymbol{c}}^T\bar{\boldsymbol{c}}$$

Find solution of approximated *L* with $\nabla L(\bar{\mathbf{c}}) = 0$

Solution can be expressed as combination of **limiting cases**

$$\bar{\boldsymbol{c}} = (\boldsymbol{S} + w\mathbf{1})^{-1}\,(\boldsymbol{S}\bar{\boldsymbol{c}}_1 + w\,\bar{\boldsymbol{c}}_2)$$

$\bar{\boldsymbol{c}}_1 = \boldsymbol{S}^{-1}\boldsymbol{M}^T(\boldsymbol{H}_1\boldsymbol{c}_0 - \boldsymbol{h}_1)$ for $w \longrightarrow 0$ **no** regularization

$\bar{\boldsymbol{c}}_2 = \boldsymbol{M}^T(\boldsymbol{H}_2\boldsymbol{c}_0 - \boldsymbol{h}_2)$ for $w \longrightarrow \infty$ **only** regularization

$$\bar{c}_k = \frac{S_{kk}}{S_{kk} + w} c_{1k} + \frac{w}{S_{kk} + w} c_{2k}$$

linear interpolation

Weight criterion

$$\bar{c}_k = \underbrace{\frac{S_{kk}}{S_{kk} + w}}_{\text{suppression matrix (SM)}} c_{1k} + \frac{w}{S_{kk} + w}\, c_{2k}$$

$$S_{kk} = \sigma_k^{-2}$$

inverse of variance of c_{1k}

many $S_{kk}^{1/2} c_{1k} \ll 1$

→ noise!

Approach

i) Get number of statistically significant coefficients of $\mathbf{c}_1$ → n_{sc}

ii) Choose w so that effective rank of SM = n_{sc}

$$\sum_k \frac{S_{kk}\, c_{1k}{}^2}{S_{kk}\, c_{1k}{}^2 + n_\sigma{}^2} = n_{\mathrm{sc}} \xrightarrow{\text{then}} \sum_k \frac{S_{kk}}{S_{kk} + w} = n_{\mathrm{sc}}$$

calculate $n_{sc}(n_\sigma)$, e.g. $n_\sigma = 2$ rejects noise with 95 % confidence

solve numericaly for w with given n_{sc}

Uncertainty of solution

Uncertainty of solution g(x) can be calculated analytically (error propagation)

$$V[g(x)] = \sum_i \sum_j \frac{\partial g(x)}{\partial c_i} \frac{\partial g(x)}{\partial c_j} V[\mathbf{c}]_{ij}$$

$$V[\mathbf{c}] = V_1[\mathbf{c}] + V_2[\mathbf{c}]$$

contribution from fit **and** from regularization
(KDE also has uncertainty)

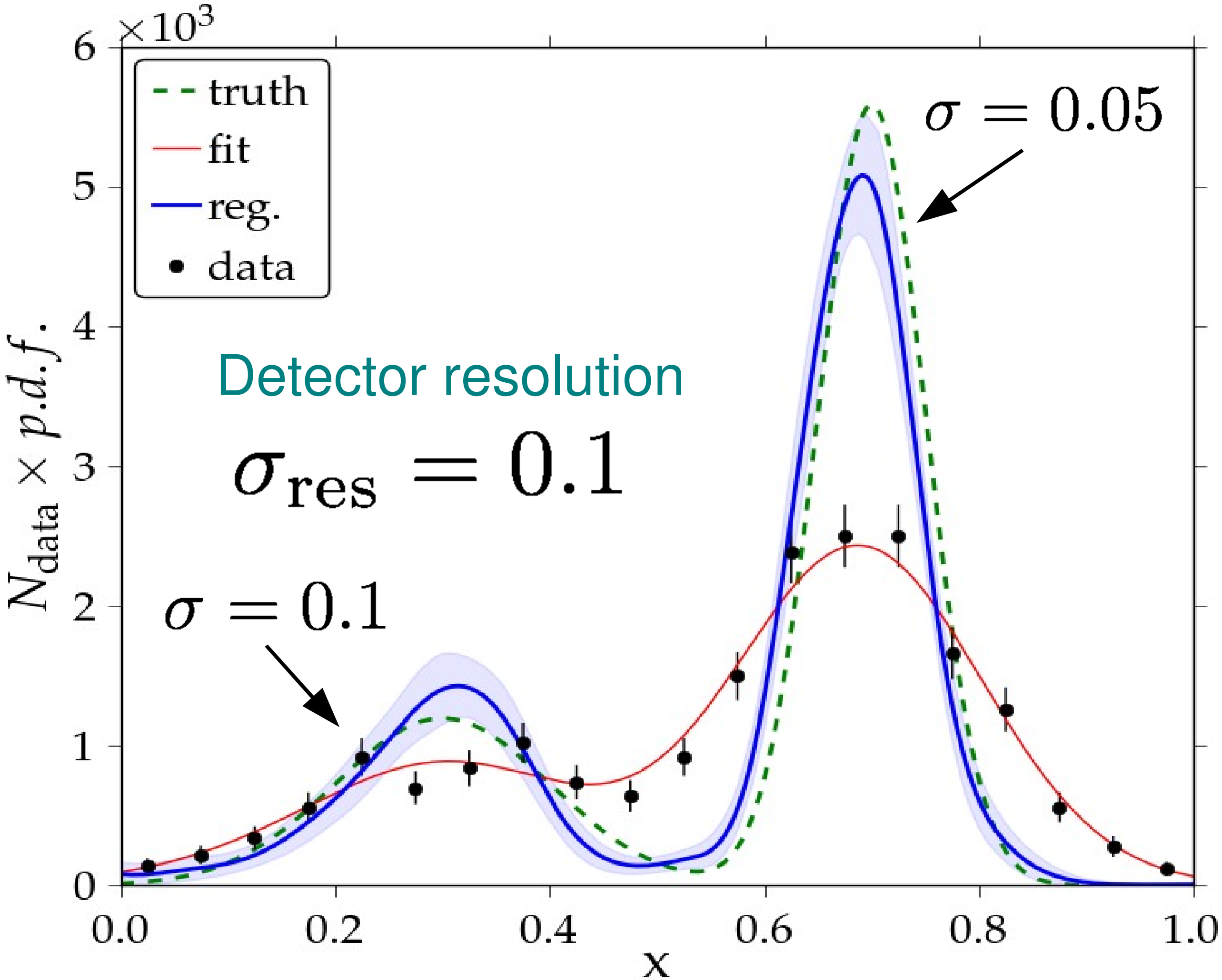
$$V_1[\mathbf{c}]_{ij} = \sum_k M_{ik} M_{jk} \frac{S_{kk}}{(S_{kk} + w)^2}$$

with $\mathbf{M} = \mathbf{U}_2 \mathbf{D}^{-1/2} \mathbf{U}_1$

$$V_2[\mathbf{c}]_{ij} = \sum_k \sum_l \sum_m \sum_p H^{-1}_{ik} H^{-1}_{jl} \frac{\partial h_k}{\partial d_m} \frac{\partial h_l}{\partial d_p} V[\mathbf{d}]_{mp}$$

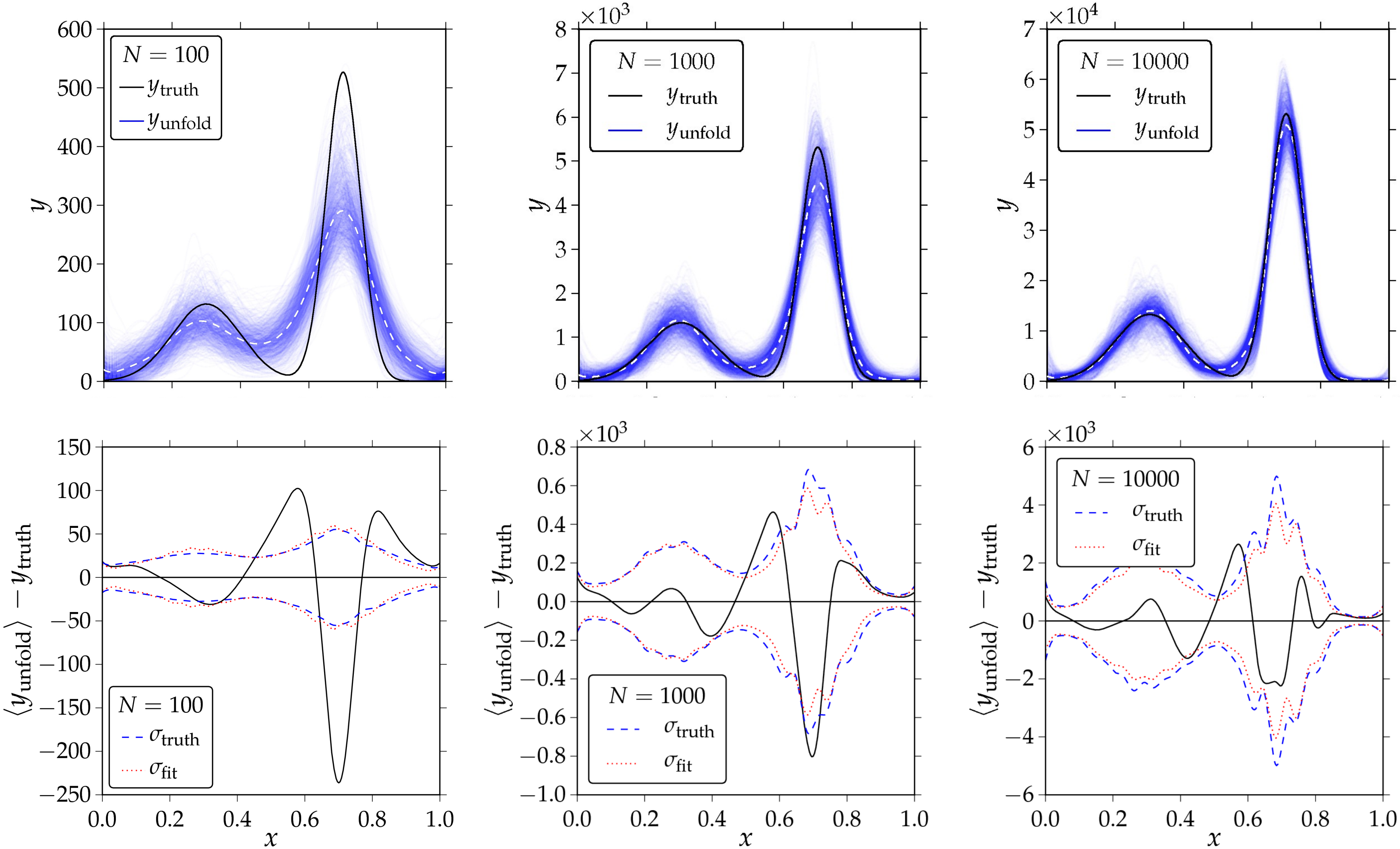
d_m are coefficients of spline approximation of KDE

MC study: example



100, 1000, 10000 data points

Results



$N = 100$
y_{truth}
y_{unfold}
$N = 1000$
$N = 10000$
$\times 10^3$
$\times 10^4$
y
$\langle y_{\text{unfold}} \rangle - y_{\text{truth}}$
σ_{truth}
σ_{fit}
x

Summary

- ARU unfolding algorithm
  - Complete procedure to unfold 1D distributions (C++ code available)
  - No user input apart from raw data and Kernel function K(y,x)
  - Un-binned analysis
  - Regularization based on adding only minimal information to kernel density estimate of original folded data distribution
  - Analytical uncertainty calculation

C++ implementation available (requirements: ROOT, Einspline library)

- Things for the future
  - How to choose knot intervals of spline parametrization in an optimal way
  - Extension to multiple dimensions
  - Estimate bias of unfolded solution
  - Analytical calculation of integral in L_2