LAMARR
INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

A Machine Learning Perspective on the Inverse Problem

Mirko Bunse

APP & Philosophy Workshop 2025 – November 13^{th}

Partner institutions:

technische universität dortmund

Fraunhofer IAIS

UNIVERSITÄT BONN

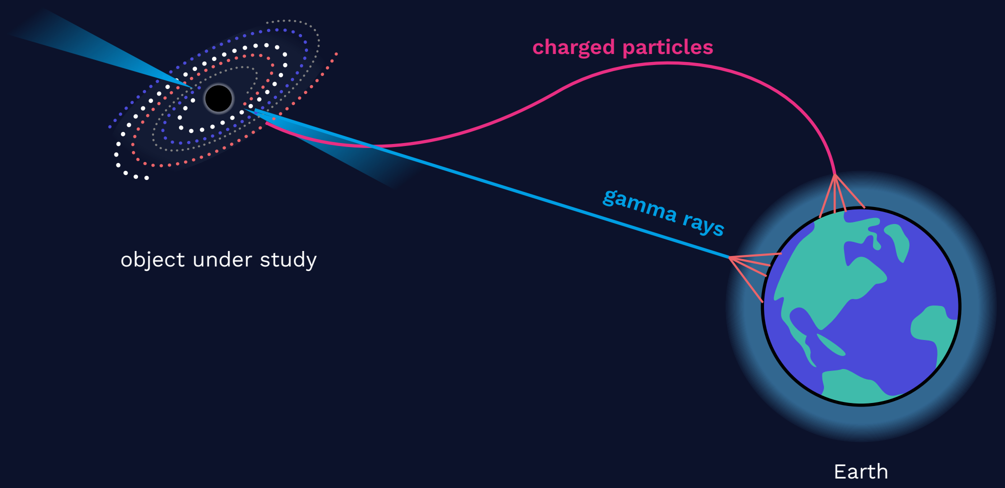
Fraunhofer IML

Institutionally funded by:

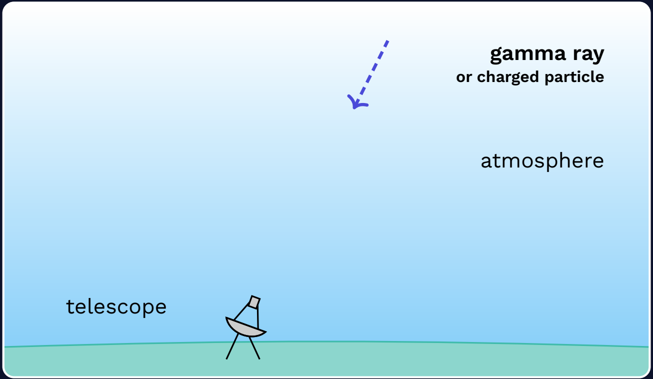
Bundesministerium für Forschung, Technologie und Raumfahrt

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

Example Use Case: Ground-Based Gamma-Ray Observatories

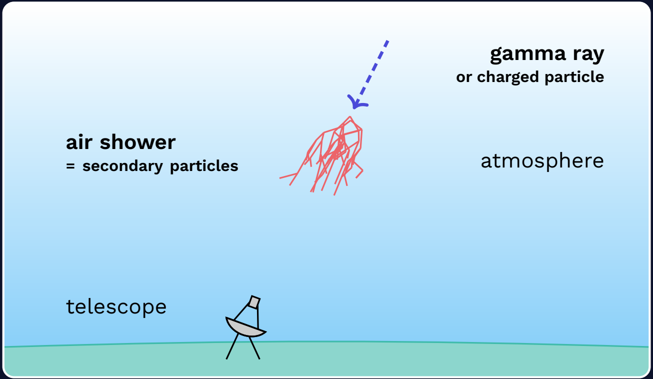


Example Use Case: Ground-Based Gamma-Ray Observatories



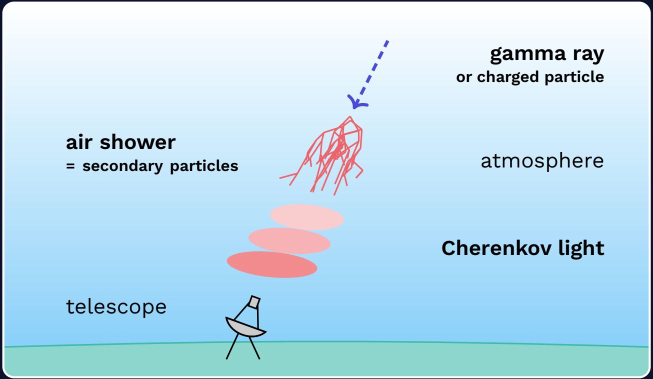
[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Example Use Case: Ground-Based Gamma-Ray Observatories



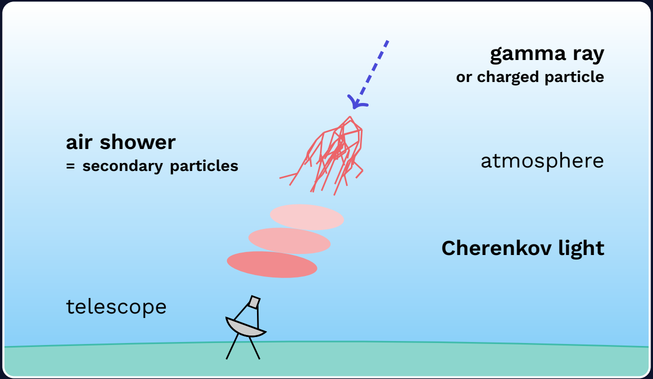
[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Example Use Case: Ground-Based Gamma-Ray Observatories



[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

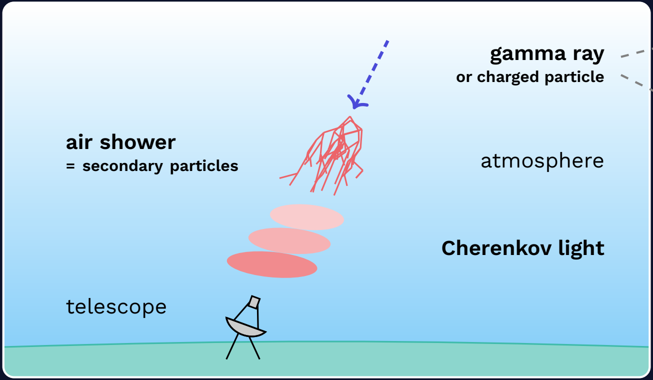
Example Use Case: Ground-Based Gamma-Ray Observatories

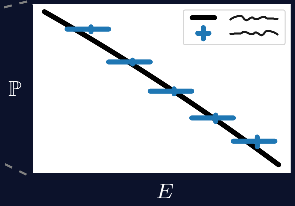


Preprocessing → Background rejection

[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

Example Use Case: Ground-Based Gamma-Ray Observatories





Preprocessing → Background rejection

[1] **Fig:** Bockermann et al., “Online Analysis of High-Volume Data Streams in Astroparticle Physics”, 2015

A Causal Learning View on the Inverse Problem

relevant quantities
(e.g., gamma energy)

measured quantities
(e.g., Cherenkov light)

$$Y \longrightarrow X$$

A Causal Learning View on the Inverse Problem

relevant quantities (e.g., gamma energy)

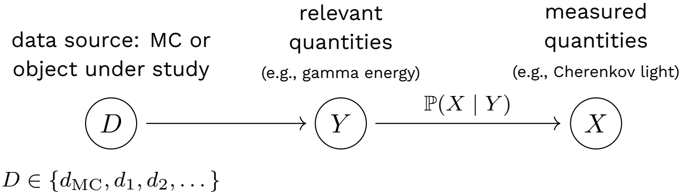
measured quantities (e.g., Cherenkov light)

$$Y \xrightarrow{\mathbb{P}(X \mid Y)} X$$

$\mathbb{P}(Y)$ $\qquad$ $\mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y)$

- **Goal:** estimate $\mathbb{P}(Y)$ from $\mathbb{P}(X) = \mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y)$

A Causal Learning View on the Inverse Problem



- **Goal:** estimate what?

A Causal Learning View on the Inverse Problem

data source: MC or object under study	relevant quantities (e.g., gamma energy)	measured quantities (e.g., Cherenkov light)
D	Y	X
$D \in \{d_{\mathrm{MC}}, d_1, d_2, \dots\}$	$\mathbb{P}(Y \mid D = d_i)$	$\mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$

$$\begin{tikzcd} D \arrow[r] & Y \arrow[r, "\mathbb{P}(X \mid Y)"] & X \end{tikzcd}$$

- **Goal:** estimate $\mathbb{P}(Y \mid D = d_i)$ from $\mathbb{P}(X \mid D = d_i)$

A Causal Learning View on the Inverse Problem

data source: MC or object under study

relevant quantities (e.g., gamma energy)

measured quantities (e.g., Cherenkov light)

$$
\begin{tikzcd}
D \arrow[r] & Y \arrow[r, "\mathbb{P}(X \mid Y)"] & X
\end{tikzcd}
$$

$D \in \{d_{\mathrm{MC}}, d_1, d_2, \dots\}$ $\qquad$ $\mathbb{P}(Y \mid D = d_i)$ $\qquad$ $\mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$

- **Goal:** estimate $\mathbb{P}(Y \mid D = d_i)$ from $\mathbb{P}(X \mid D = d_i)$
  - d_i induce different distributions $\mathbb{P}(Y \mid D = d_i)$
  - d_i induce different measurements $\mathbb{P}(X \mid D = d_i)$
  - use MC data from $\mathbb{P}(X, Y \mid D = d_{\mathrm{MC}})$
- **Key assumption:** $\mathbb{P}(X \mid Y)$ is stable across domains

A Causal Learning View on the Inverse Problem

data source: MC or object under study	relevant quantities (e.g., gamma energy)	measured quantities (e.g., Cherenkov light)	(standard ML estimate)

$$\begin{tikzcd} D \arrow[r] & Y \arrow[r, "\mathbb{P}(X \mid Y)"] & X \arrow[r, dashed, "\widehat{\mathbb{P}}(Y \mid X)"] & \widehat{Y} \neq Y \end{tikzcd}$$

$D \in \{d_{\text{MC}}, d_1, d_2, \dots\}$ $\quad$ $\boxed{\mathbb{P}(Y \mid D = d_i)}$ $\quad$ $\mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$ $\quad$ $\mathbb{P}(\widehat{Y} \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$

- **Goal:** estimate $\boxed{\mathbb{P}(Y \mid D = d_i)}$ from $\mathbb{P}(X \mid D = d_i)$
  - d_i induce different distributions $\mathbb{P}(Y \mid D = d_i)$
  - d_i induce different measurements $\mathbb{P}(X \mid D = d_i)$
  - use MC data from $\mathbb{P}(X, Y \mid D = d_{\text{MC}})$
- **Key assumption:** $\mathbb{P}(X \mid Y)$ is stable across domains
- **Standard machine learning** falsely assumes $\mathbb{P}(Y \mid D = d_i)$ to be stable, too.

A Causal Learning View on the Inverse Problem

data source: MC or object under study

relevant quantities (e.g., gamma energy)

measured quantities (e.g., Cherenkov light)

(standard ML estimate)

$$\begin{tikzcd} D \arrow[r] & Y \arrow[r, "\mathbb{P}(X \mid Y)"] & X \arrow[r, dashed, "\widehat{\mathbb{P}}(Y \mid X)"] & \widehat{Y} \neq Y \end{tikzcd}$$

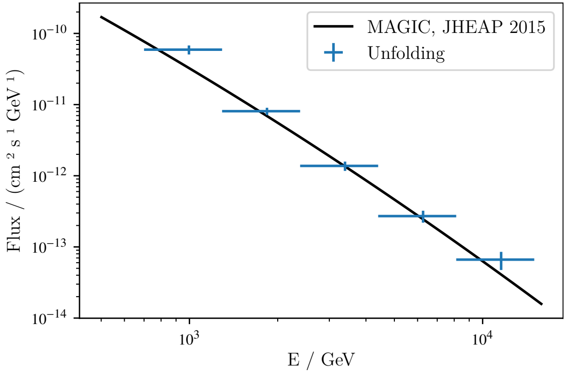
$D \in \{d_{\mathrm{MC}}, d_1, d_2, \ldots\}$ $\quad$ $\mathbb{P}(Y \mid D = d_i)$ $\quad$ $\mathbb{P}(X \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$ $\quad$ $\mathbb{P}(\widehat{Y} \mid Y) \cdot \mathbb{P}(Y \mid D = d_i)$

- **Goal:** estimate $\mathbb{P}(Y \mid D = d_i)$ from $\mathbb{P}(X \mid D = d_i)$
  - d_i induce different distributions $\mathbb{P}(Y \mid D = d_i)$
  - d_i induce different measurements $\mathbb{P}(X \mid D = d_i)$
  - use MC data from $\mathbb{P}(X, Y \mid D = d_{\mathrm{MC}})$
- **Key assumption:** $\mathbb{P}(X \mid Y)$ is stable across domains
- **Standard machine learning** falsely assumes $\mathbb{P}(Y \mid D = d_i)$ to be stable, too.
- **Solution:** proper modeling of the measurement process & no false assumptions $\rightarrow$ Fredholm equation

Inverse Problems: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

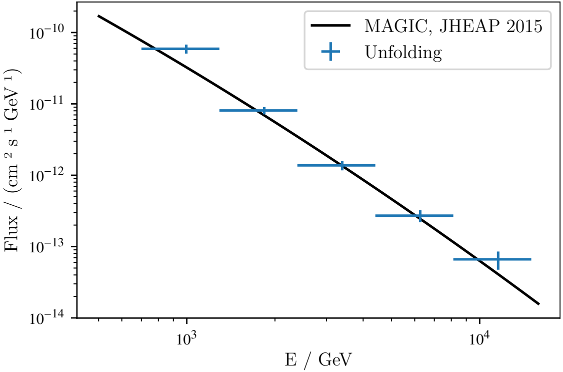
Inverse Problems: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a system of linear equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \begin{cases} \mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

Inverse Problems: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

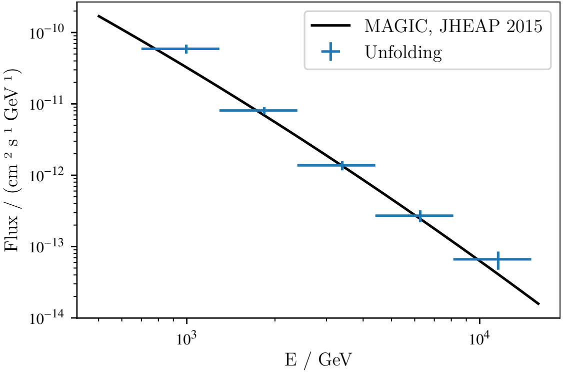
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a system of linear equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \begin{cases} \mathbf{q} &= \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i &= \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$

and solve it by minimizing some loss, i.e.,

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta}{\arg\min} \; \ell\,(\mathbf{p};\; \mathbf{q}, \mathbf{M})$$



[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

Inverse Problems: the Reconstruction of Spectra

Goal: reconstruct the spectrum $p(y)$ of some quantity y from a measurement $q(x)$.

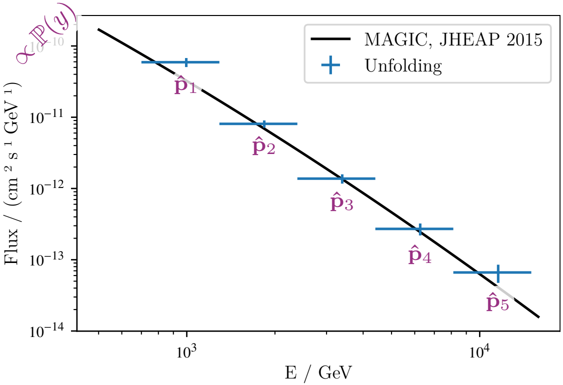
$$\underbrace{q(x)}_{\text{measurement}} = \int \underbrace{M(x \mid y)}_{\text{transfer}} \cdot \underbrace{p(y)}_{\text{target}} \, \mathrm{d}y$$

Approach: set up a system of linear equations

$$\mathbf{q} = \mathbf{M}\mathbf{p} \qquad \text{where} \quad \begin{cases} \mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum_{x \in \mathrm{B}} \phi(x) \\ \mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum_{x \in \mathrm{D}_i} \phi(x) \end{cases}$$

and solve it by minimizing some loss, i.e.,

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta}{\arg\min} \; \ell\,(\mathbf{p};\; \mathbf{q}, \mathbf{M})$$

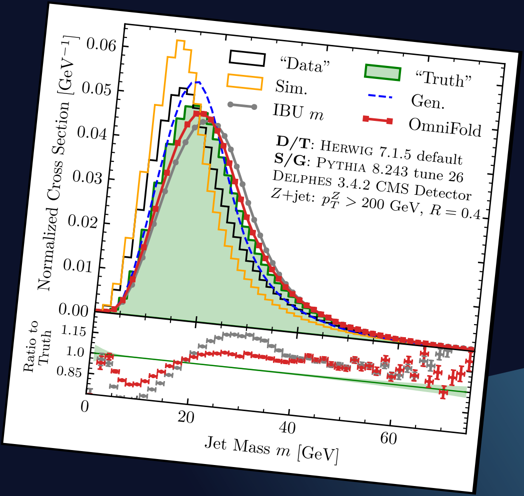


unknown $y \in [1, 5]$

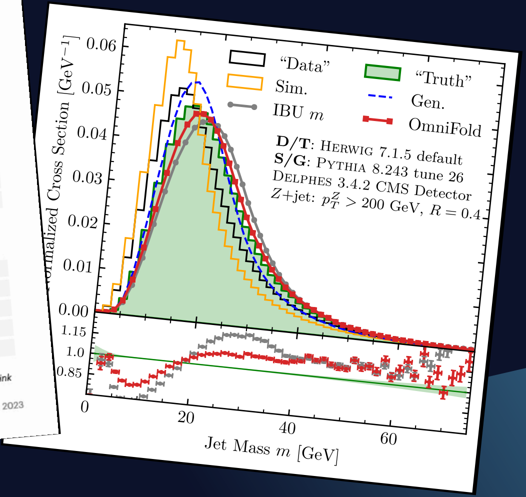
$\equiv$ class label!

[2] Fig.: Morik and Rhode, *Machine Learning under Resource Constraints – Discovery in Physics*, 2023

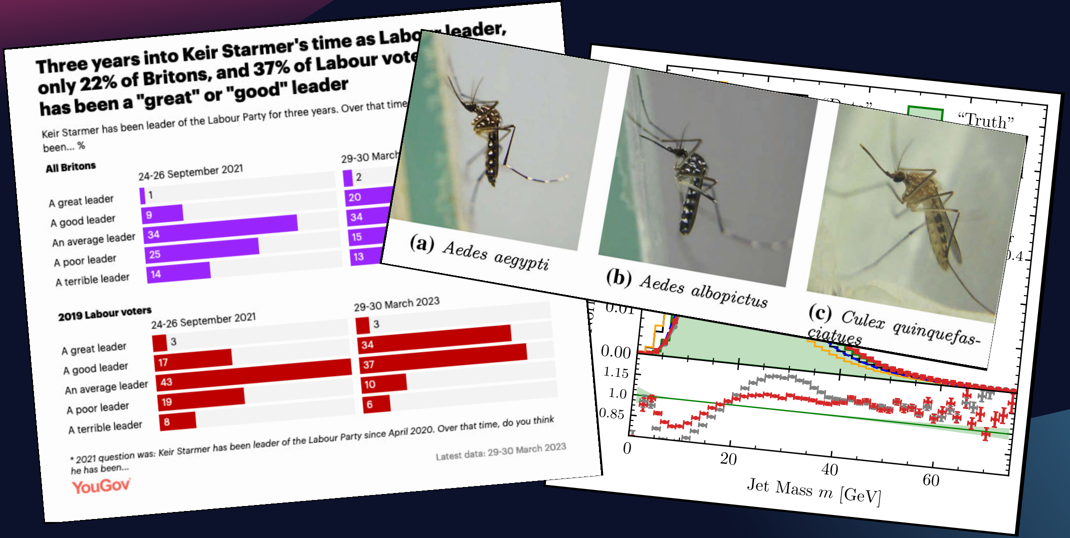
Inverse Problems Everywhere: Count + Predict



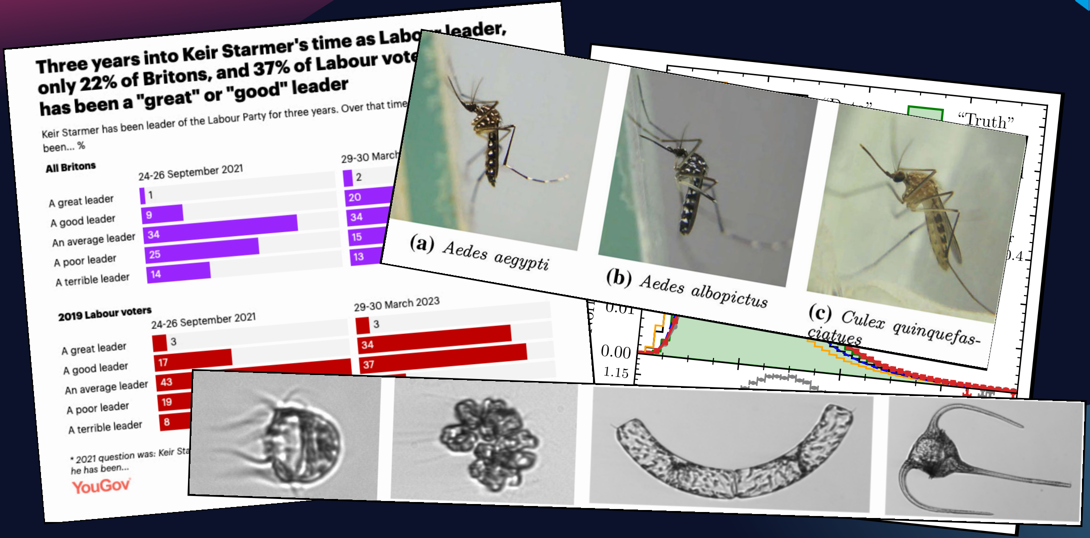
Inverse Problems Everywhere: Count + Predict



Inverse Problems Everywhere: Count + Predict



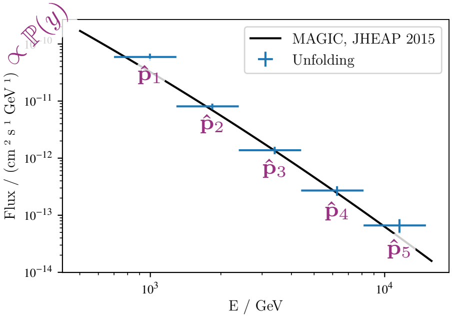
Inverse Problems Everywhere: Count + Predict



Cognitive Interests of Different Fields

Physics: physical implications of the result $\hat{\mathbf{p}}$

- accurate observation of particle sources
- validation of theories about:
  - particle acceleration
  - dark matter
  - ...
- key assumption: $\hat{\mathbf{p}} \approx \mathbf{p}^*$



$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p} \in \Delta} \ell\left(\mathbf{p};\ \mathbf{q}, \mathbf{M}\right)$$

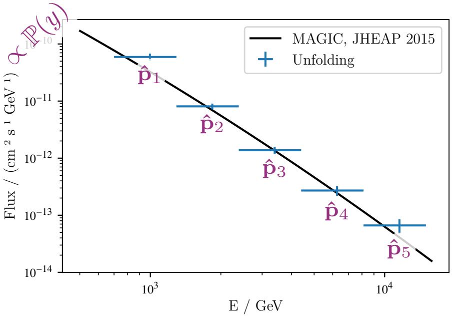
$$\mathbf{q} = \frac{1}{|\mathrm{B}\ |} \sum\nolimits_{x \in \mathrm{B}} \ \phi(x)$$

$$\mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum\nolimits_{x \in \mathrm{D}_i} \phi(x)$$

Cognitive Interests of Different Fields

Physics: physical implications of the result $\hat{\mathbf{p}}$

- accurate observation of particle sources
- validation of theories about:
  - particle acceleration
  - dark matter
  - ...
- key assumption: $\hat{\mathbf{p}} \approx \mathbf{p}^*$



Computer Science / ML: implications of the choice of ϕ and ℓ on $\hat{\mathbf{p}} \to \mathbf{p}^*$

- statistical consistency
- convergence
- robustness and data quality
- key assumption: MC data is sufficiently informative

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta}{\arg\min}\, \ell\left(\mathbf{p};\ \mathbf{q}, \mathbf{M}\right)$$

$$\mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum\nolimits_{x \in \mathrm{B}} \ \phi(x)$$

$$\mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum\nolimits_{x \in \mathrm{D}_i} \phi(x)$$

Addressing the Central Questions of This Workshop

What Can Be Measured? / Consequences of Ill-Posedness

Most methods[3] are both Fisher-consistent[4] and asymptotically consistent.[5]

[3] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.

[4] Tasche, “Fisher consistency for prior probability shift”, 2017.

[5] Dussap, Blanchard, and Chérief-Abdellatif, “Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching”, 2023.

What Can Be Measured? / Consequences of Ill-Posedness

Most methods[3] are both Fisher-consistent[4] and asymptotically consistent.[5]

quality improves
with data volume

[3] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.

[4] Tasche, “Fisher consistency for prior probability shift”, 2017.

[5] Dussap, Blanchard, and Chérief-Abdellatif, “Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching”, 2023.

What Can Be Measured? / Consequences of Ill-Posedness

Most methods[3] are both Fisher-consistent[4] and asymptotically consistent.[5]

quality improves with data volume

errors can be estimated

[3] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.

[4] Tasche, “Fisher consistency for prior probability shift”, 2017.

[5] Dussap, Blanchard, and Chérief-Abdellatif, “Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching”, 2023.

What Can Be Measured? / Consequences of Ill-Posedness

Most methods[3] are both Fisher-consistent[4] and asymptotically consistent.[5]

quality improves with data volume

errors can be estimated

(perfect measurement if data was infinite)

[3] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.
[4] Tasche, “Fisher consistency for prior probability shift”, 2017.
[5] Dussap, Blanchard, and Chérief-Abdellatif, “Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching”, 2023.

What Can Be Measured? / Consequences of Ill-Posedness

Most methods[3] are both Fisher-consistent[4] and asymptotically consistent.[5]

quality improves with data volume ✓

errors can be estimated ✓

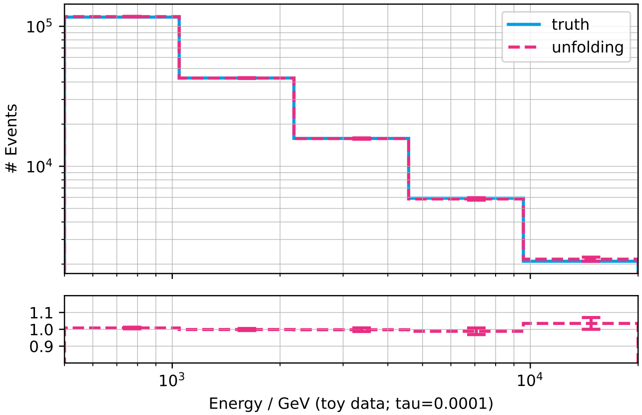
(perfect measurement if data was infinite) ✓

Key Assumption: $\mathbb{P}(X \mid Y)$ is stable = no data-MC mismatch besides the spectrum ✗

[3] Bunse, "Unification of Algorithms for Quantification and Unfolding", 2022.
[4] Tasche, "Fisher consistency for prior probability shift", 2017.
[5] Dussap, Blanchard, and Chérief-Abdellatif, "Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching", 2023.

What Can Be Measured? / Consequences of Ill-Posedness

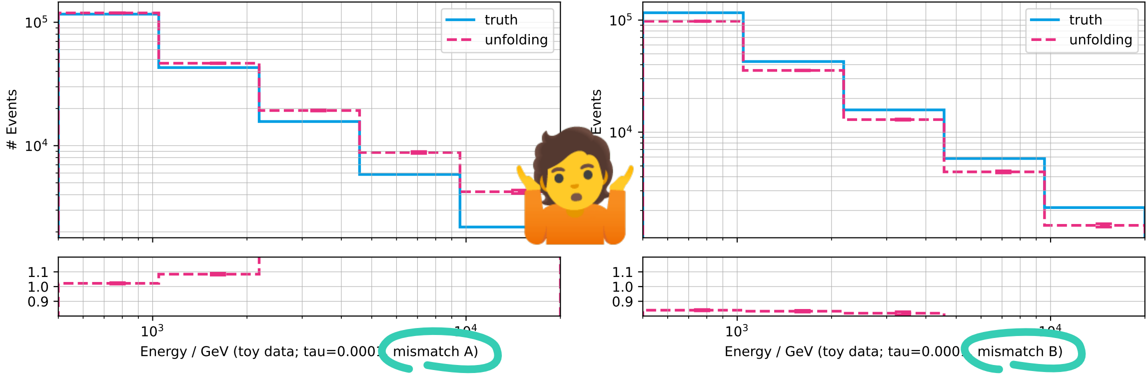
Fig: Perfect unfolding; no data-MC mismatches & abundant data.



✓

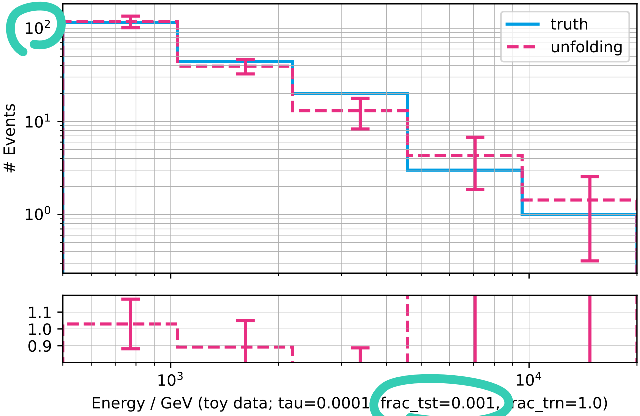
What Can Be Measured? / Consequences of Ill-Posedness

Fig: Unfolding with different data-MC mismatches.



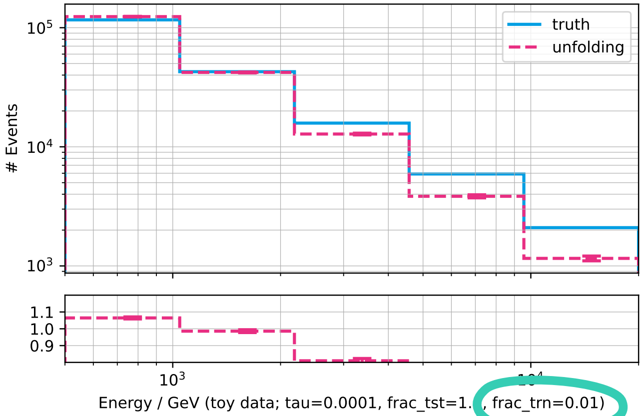
What Can Be Measured? / Consequences of Ill-Posedness

Fig: Unfolding with little measured data.



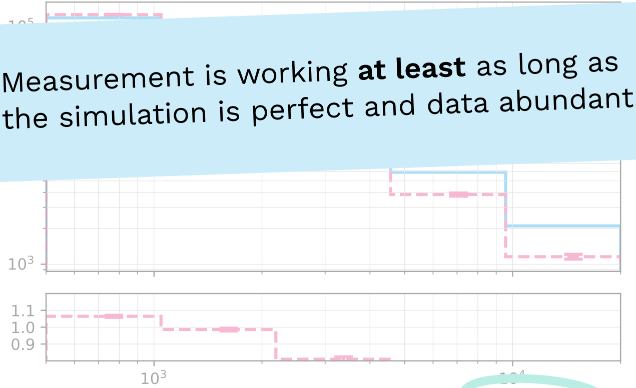
What Can Be Measured? / Consequences of Ill-Posedness

Fig: Unfolding with little MC data.



What Can Be Measured? / Consequences of Ill-Posedness

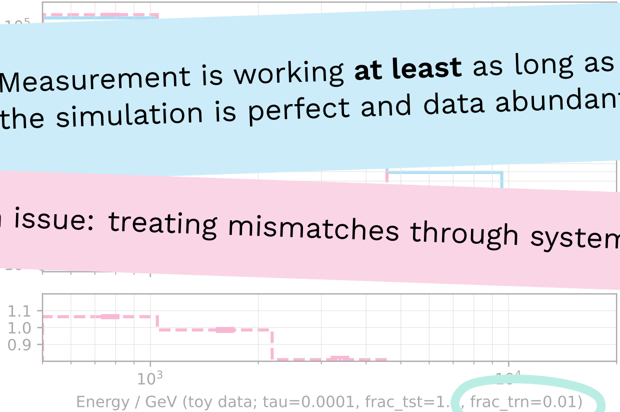
Fig: Unfolding with little MC data.



Measurement is working **at least** as long as the simulation is perfect and data abundant.

What Can Be Measured? / Consequences of Ill-Posedness

Fig: Unfolding with little MC data.



Measurement is working **at least** as long as the simulation is perfect and data abundant.

Open issue: treating mismatches through systematics.

To What Extent is the Measurement Theory-Laden?

Regularization: introduces preference towards certain properties of the spectrum

- needed for maximum accuracy
- sacrifices sensitivity to deviations from these properties
- impact can be controlled by setting a scalar parameter value

To What Extent is the Measurement Theory-Laden?

Regularization: introduces preference towards certain properties of the spectrum

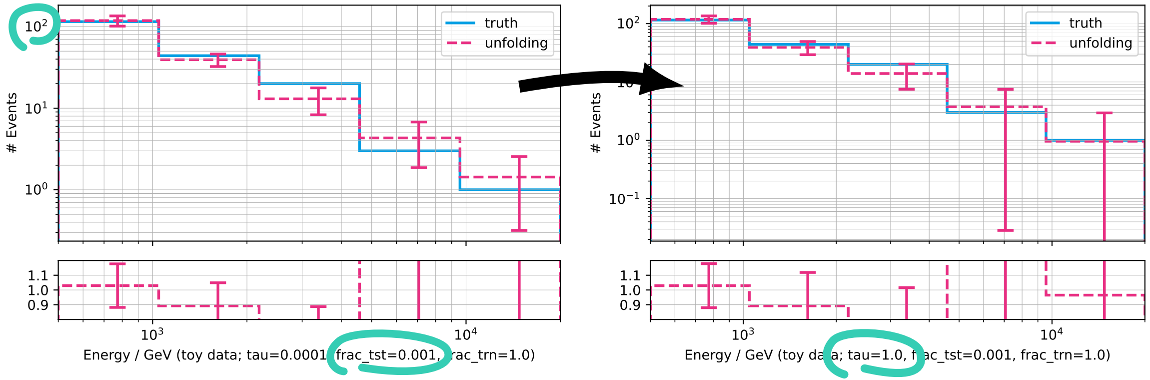
- needed for maximum accuracy
- sacrifices sensitivity to deviations from these properties
- impact can be controlled by setting a scalar parameter value

Example: Tikhonov regularization linearizes $\hat{\mathbf{p}}$ (Question 4: assumptions about $\mathbf{p}^*$)

- increases smoothness of measurements
- reduces sensitivity to bumps and kinks

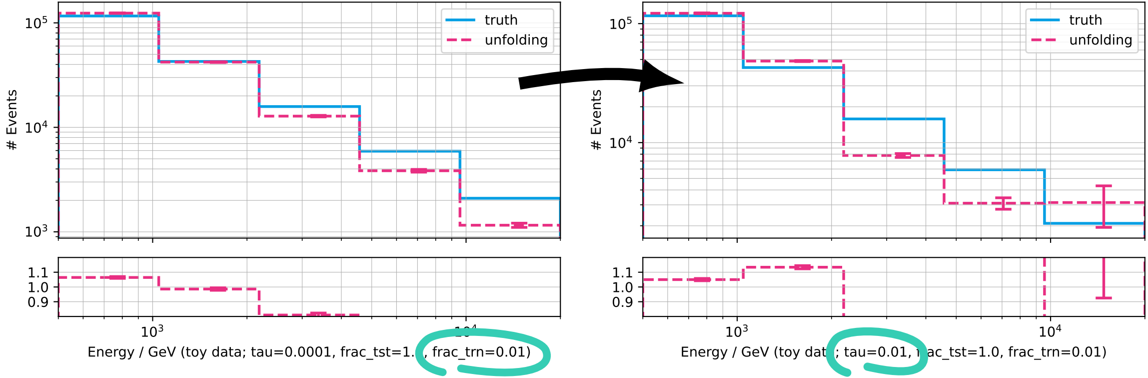
To What Extent is the Measurement Theory-Laden?

Fig: Unfolding with little measured data, revisited with less regularization.



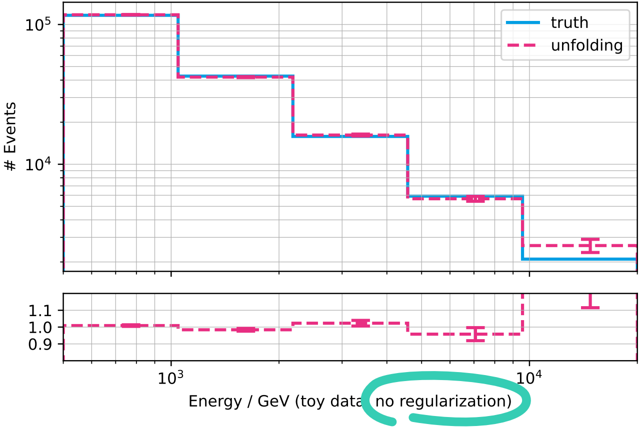
To What Extent is the Measurement Theory-Laden?

Fig: Unfolding with little MC data, revisited with less regularization.



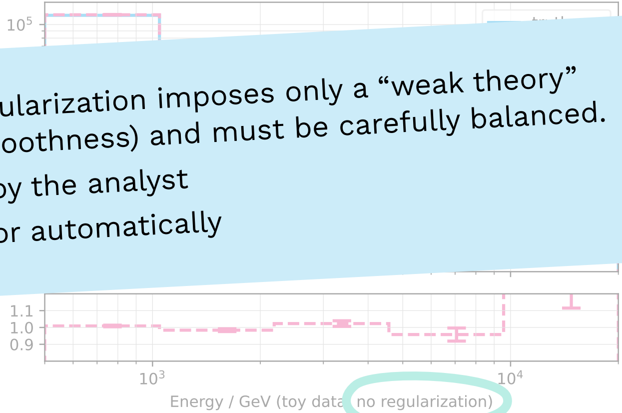
To What Extent is the Measurement Theory-Laden?

Fig: “Perfect unfolding” from before, but entirely without regularization.



To What Extent is the Measurement Theory-Laden?

Fig: “Perfect unfolding” from before, but entirely without regularization.



Regularization imposes only a “weak theory” (smoothness) and must be carefully balanced.

- by the analyst
- or automatically

The Role of the Background

Removal of background: before unfolding, using binary classification.

The Role of the Background

Removal of background: before unfolding, using binary classification.

Balance between purity and sensitivity?

- with background, the model becomes $\mathbf{q}' = \mathbf{q} + \mathbf{b} = \mathbf{M}\mathbf{p} + \mathbf{b}$
- hence, $\mathbf{b}$ has to be removed from both sides

The Role of the Background

Removal of background: before unfolding, using binary classification.

Balance between purity and sensitivity?

- with background, the model becomes $\mathbf{q}' = \mathbf{q} + \mathbf{b} = \mathbf{M}\mathbf{p} + \mathbf{b}$
- hence, $\mathbf{b}$ has to be removed from both sides
- purity & sensitivity do not affect the unfolding
- potential difficulty (again): data-MC mismatches

Background does not impose additional difficulties

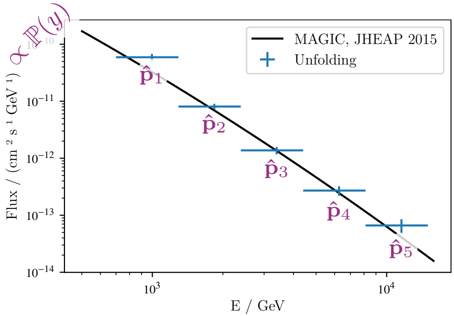
Recap: Cognitive Interests of Different Fields

Physics: physical implications of the result $\hat{\mathbf{p}}$

- accurate observation of particle sources
- validation of theories about:
  - particle acceleration
  - dark matter
  - ...
- key assumption: $\hat{\mathbf{p}} \approx \mathbf{p}^*$

Computer Science / ML: implications of the choice of ϕ and ℓ on $\hat{\mathbf{p}} \rightarrow \mathbf{p}^*$

- statistical consistency
- convergence
- robustness and data quality
- key assumption: MC data is sufficiently informative



$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p} \in \Delta} \ell\left(\mathbf{p};\, \mathbf{q}, \mathbf{M}\right)$$

$$\mathbf{q} = \frac{1}{|\mathrm{B}\,|} \sum\nolimits_{x \in \mathrm{B}} \; \phi(x)$$

$$\mathbf{M}_i = \frac{1}{|\mathrm{D}_i|} \sum\nolimits_{x \in \mathrm{D}_i} \phi(x)$$

Conclusion: ML Perspective on the Inverse Problem

What can be measured? / Consequences of ill-posedness:

- measurement is working, as long as the simulation is perfect
- open issue: treatment of mismatches through systematics

Conclusion: ML Perspective on the Inverse Problem

What can be measured? / Consequences of ill-posedness:

- measurement is working, as long as the simulation is perfect
- open issue: treatment of mismatches through systematics

To what extend is the measurement theory-laden?

- careful balance is necessary (by user & by algorithms)

Conclusion: ML Perspective on the Inverse Problem

What can be measured? / Consequences of ill-posedness:

- measurement is working, as long as the simulation is perfect
- open issue: treatment of mismatches through systematics

To what extend is the measurement theory-laden?

- careful balance is necessary (by user & by algorithms)

The role of the background:

- no additional difficulties

Central epistemological problems: MC quality & data volume

Backup Slides:
5 Learnings From
Quantification Research

1st Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to “unlimited data”), it would return the true class prevalences:

$$\underbrace{h'\big(\mathbb{Q}(X)\big)}_{\substack{\text{population}\\ \text{analogue}\\ \text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

[6] Blobel, “An unfolding method for high energy physics experiments”, 2002, .

[7] Bunse, “Unification of Algorithms for Quantification and Unfolding”, 2022.

[8] Gövert, “Fisher-Konsistenz für Quantification-Algorithmen”, 2023.

1st Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to "unlimited data"), it would return the true class prevalences:

$$\underbrace{h'\big(\mathbb{Q}(X)\big)}_{\substack{\text{population}\\\text{analogue}\\\text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

- can also be defined for other types of data set shift
- not a sufficient but certainly **a necessary criterion** for quantifier selection

[6] Blobel, "An unfolding method for high energy physics experiments", 2002, .

[7] Bunse, "Unification of Algorithms for Quantification and Unfolding", 2022.

[8] Gövert, "Fisher-Konsistenz für Quantification-Algorithmen", 2023.

1st Learning: Statistical Consistency

Definition (Fisher Consistency for Prior Probability Shift):

If a consistent quantifier had access to the entire population $\mathbb{Q}(X)$ (i.e., to "unlimited data"), it would return the true class prevalences:

$$\underbrace{h'\Big(\mathbb{Q}(X)\Big)}_{\substack{\text{population} \\ \text{analogue} \\ \text{of } h(\mathrm{B})}} = \mathbb{Q}(Y) \quad \underbrace{\forall\ \mathbb{Q} : \mathbb{Q}(X \mid Y) = \mathbb{P}(X \mid Y)}_{\text{for any } \mathbb{Q} \text{ with PPS}}$$

- can also be defined for other types of data set shift
- not a sufficient but certainly **a necessary criterion** for quantifier selection

1) RUN[6] **/ TRUEE** (and others) *are* Fisher consistent[7] ✓

2) DSEA & DSEA+ are *not* Fisher consistent[8] ✗

[6] Blobel, "An unfolding method for high energy physics experiments", 2002, .

[7] Bunse, "Unification of Algorithms for Quantification and Unfolding", 2022.

[8] Gövert, "Fisher-Konsistenz für Quantification-Algorithmen", 2023.

2nd Learning: The Anatomy of Prediction Errors

Prediction Error Bound:[9] describes the *impact of* and the *interplay between* causes of errors.

$$\underbrace{\left\| h(\mathrm{B}) - \mathbf{p}^* \right\|_2}_{\text{prediction error}} \leq \underbrace{\frac{2k(2 + \sqrt{2 \log \frac{2C}{\delta}})}{\sqrt{\lambda_2}}}_{\text{representation } \phi} \cdot \left(\underbrace{\left\| \frac{\mathbf{p}^*}{\mathbf{p}_{\mathrm{trn}}} \right\|_2}_{\text{shift}} \cdot \underbrace{\frac{1}{\sqrt{|\mathrm{D}|}}}_{\text{volume D}} + \underbrace{\frac{1}{\sqrt{|\mathrm{B}|}}}_{\text{volume B}} \right)$$

where

- $h(\mathrm{B})$ is the solution of $\mathbf{q} = \mathbf{M}\mathbf{p}$
- k is a constant s.t. $\|\phi(x)\|_2 \leq k \ \forall \, x \in \mathcal{X}$
- λ_2 is the second-smallest eigenvalue of some particular $\mathbf{G}$
- δ is the desired probability

[9] Dussap, Blanchard, and Chérief-Abdellatif, “Label Shift Quantification with Robustness Guarantees via Distribution Feature Matching”, 2023, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[6]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\, \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗

[10] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[11] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[6]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\, \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗
TRUEE[10]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\, \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗

[10] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[11] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[6]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗
TRUEE[10]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗
Constrained[11]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓

[10] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[11] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

3rd Learning: Improved Optimization Techniques

Algorithm	Estimate	Validity
RUN[6]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^C}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗
TRUEE[10]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \geq \mathbf{0}}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	invalid: $\hat{\mathbf{p}} \notin \Delta$ ✗
Constrained[11]	$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \Delta}{\arg\min} \; \ell(\mathbf{p};\; \mathbf{q}, \mathbf{M})$	valid ✓
Soft-Max[11]	$\hat{\mathbf{p}} = \sigma(\mathbf{l}^*), \quad \mathbf{l}^* = \underset{\mathbf{l} \in \mathbb{R}^{C-1}}{\arg\min} \; \ell(\sigma(\mathbf{l});\; \mathbf{q}, \mathbf{M})$	valid ✓

[10] Milke et al., “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”, 2013.
[11] Bunse, “On Multi-Class Extensions of Adjusted Classify and Count”, 2022, .

4th Learning: Methods Are Numerous

Most methods are combinations of

- a data representation $\phi : \mathcal{X} \rightarrow \mathcal{Z}$
- a loss function $\ell : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$
- an optimization algorithm

These components can be recombined to even more methods.

[12] Forman, “Quantifying counts and costs via classification”, 2008, .
[13] Bella et al., “Quantification via Probability Estimators”, 2010, .
[14] González-Castro, Alaíz-Rodríguez, and Alegre, “Class distribution estimation based on the Hellinger distance”, 2013, .
[15] Börner et al., “Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era”, 2020, .
[16] Kawakubo, Plessis, and Sugiyama, “Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance”, 2016, .

4th Learning: Methods Are Numerous

Most methods are combinations of

- a data representation $\phi : \mathcal{X} \to \mathcal{Z}$
- a loss function $\ell : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$
- an optimization algorithm

These components can be recombined to even more methods.

github.com/mirkobunse/qunfold

Representations: hard[12] & soft[13] classification, histograms[14], tree-based binnings[15], kernel means[16], ...

Loss Functions: least squares[12,13], Hellinger distance[14], energy distance[16], Poisson likelihood[6], ...

How to Choose? axiomatic approach required but up to you (consistency, simplicity, transparency, ...?)

[12] Forman, "Quantifying counts and costs via classification", 2008, .
[13] Bella et al., "Quantification via Probability Estimators", 2010, .
[14] González-Castro, Alaíz-Rodríguez, and Alegre, "Class distribution estimation based on the Hellinger distance", 2013, .
[15] Börner et al., "Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era", 2020, .
[16] Kawakubo, Plessis, and Sugiyama, "Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance", 2016, .

5th Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \; \forall \, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])

[17] Bunse et al., "Regularization-based Methods for Ordinal Quantification", 2024.

5th Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \ \forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[9])

[17] Bunse et al., "Regularization-based Methods for Ordinal Quantification", 2024.

5^{th} Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[9])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \;\exists\, i \neq j$

[17] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5th Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \ \forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[9])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \ \exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$

[17] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5^{th} Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[9])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \;\exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$
- data-MC mismatches / concept shift: $\mathbb{Q}(\mathbf{x} \mid y) \neq \mathbb{P}(\mathbf{x} \mid y)$ (in addition to PPS)

[17] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

5th Learning: Open Issues in Quantification Research

Complications of Experimental Physics:

- ordinality: $y_i \prec y_{i+1} \;\forall\, i \in \mathcal{Y}$ (to be covered through regularization for ordinal plausibility[17])
- background: $\mathbb{Q}(\mathbf{x}) = \mathbb{Q}(\mathbf{x}, \varnothing) + \sum_{y=1}^{C} \mathbb{Q}(\mathbf{x}, y)$ (PPS with a noise class[9])
- acceptance / class-conditional selection bias: $\mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_i) \neq \mathbb{Q}(\mathbf{x} \in \mathrm{B} \mid y_j) \;\exists\, i \neq j$
- changing environment: $\mathbb{Q}(\mathbf{x}, y) = \sum_{e \in \mathcal{E}} \mathbb{Q}(\mathbf{x}, y, e)$
- data-MC mismatches / concept shift: $\mathbb{Q}(\mathbf{x} \mid y) \neq \mathbb{P}(\mathbf{x} \mid y)$ (in addition to PPS)
- inspect contributions of individual data items $\mathbf{x} \in \mathrm{B}$ to $h(\mathrm{B})$ (data selection, human in the loop)

Hence, there are substantial opportunities for quantification-related research in Computer Science.

[17] Bunse et al., “Regularization-based Methods for Ordinal Quantification”, 2024.

Conclusion: 5 Learnings From Quantification

Understanding of the Problem Statement:

1) Consistency is a necessary criterion for algorithm selection

2) The prediction error is governed by the representation, the amount of shift, and the data volumes

Conclusion: 5 Learnings From Quantification

Understanding of the Problem Statement:

1) Consistency is a necessary criterion for algorithm selection

2) The prediction error is governed by the representation, the amount of shift, and the data volumes

Improvements of the Methods:

3) Contraints must be implemented, either explicitly or via soft-max

4) Many methods—or aspects thereof—have a potential for improving physics analyses

5) Physics applications motivate further developments in quantification research