Vorlesung
Statistische Methoden der Datenanalyse
Prof. Dr. Dr. Wolfgang Rhode

Entfaltung II

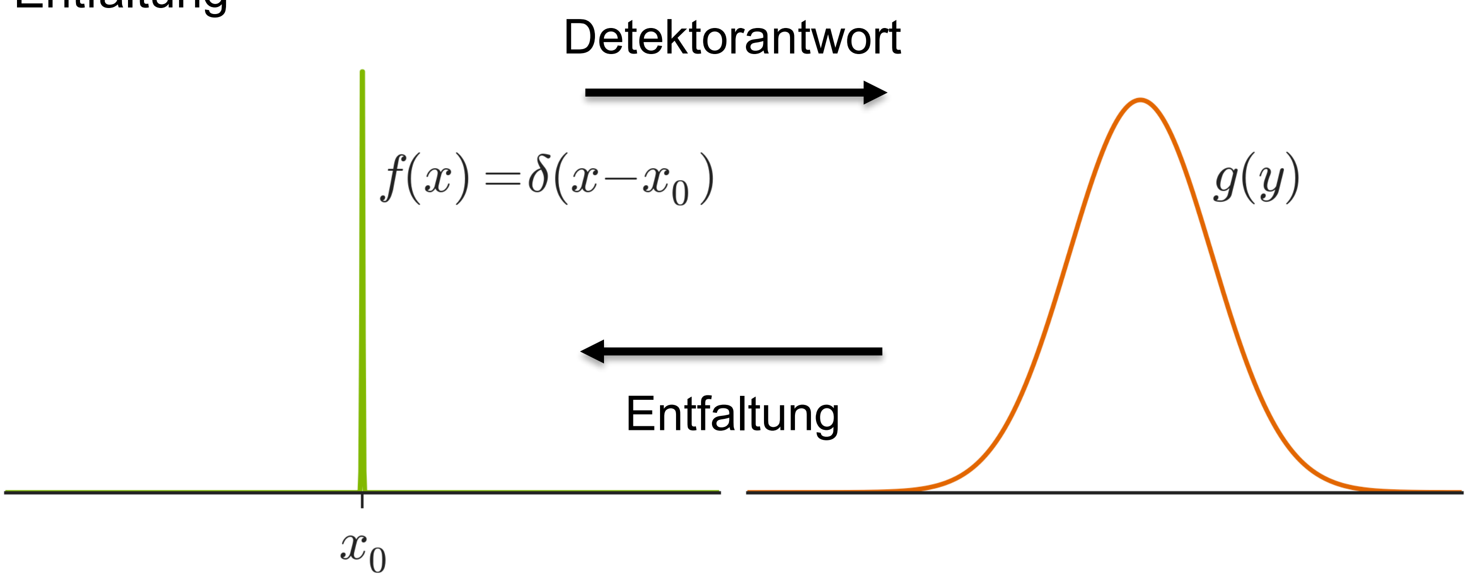
Inhalt

- Wiederholung: Grundlagen der Entfaltung
- Beispiel: Pixeldetektor
- Binning
  - Eindimensionale Verfahren
  - Mehrdimensionale Verfahren
- Likelihood-Ansatz
  - Least-Squares-Ansatz
  - SVD
  - Iterative Bayesian Unfolding
  - Regularisierungsterme
  - Poisson-Ansatz
- Validierung

Zusammenfassung: Entfaltung

- Direkte Messung einer Variable *x* unmöglich: Lediglich korrelierte Observable *y* ist verfügbar, Ergebnis eines stochastischen Prozesses

→ Entfaltung



Zusammenfassung: Entfaltung

- Migration der wahren physikalischen Größe in Observablen beschrieben durch *Fredholm*-Integralgleichung

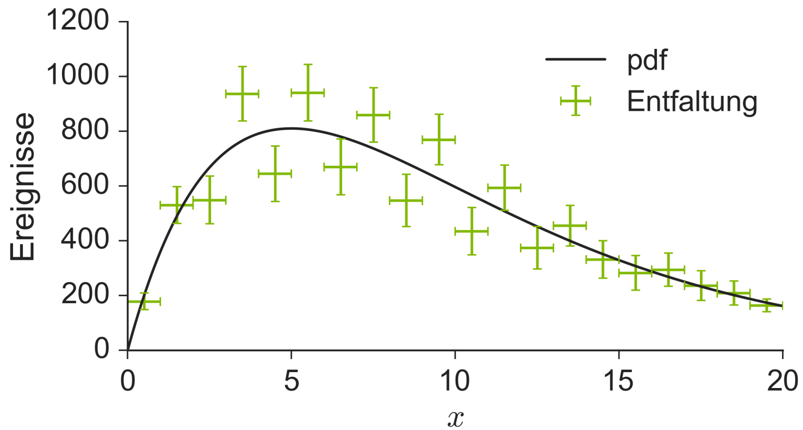
$$g(y) = \int_{\Omega} A(x,y) f(x) \mathrm{d}x + b(y)$$

- So nur schwer lösbar → Diskretisierung → Lineare Algebra

$$g_i = \sum_j A_{ij} f_j + b_i \Longleftrightarrow \mathbf{g} = \mathrm{A} \cdot \mathbf{f} + \mathbf{b}$$

Zusammenfassung: Entfaltung

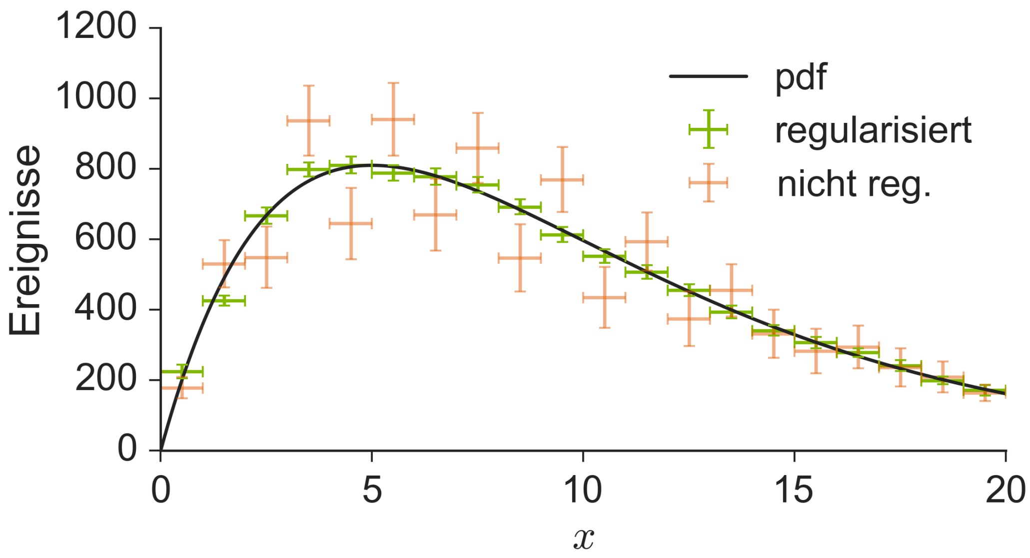
- Problem: Oszillationen → Schlecht konditioniertes Problem



- Grund: Kleine, statistisch insignifikante Eigenwerte der Migrationsmatrix

Zusammenfassung: Entfaltung

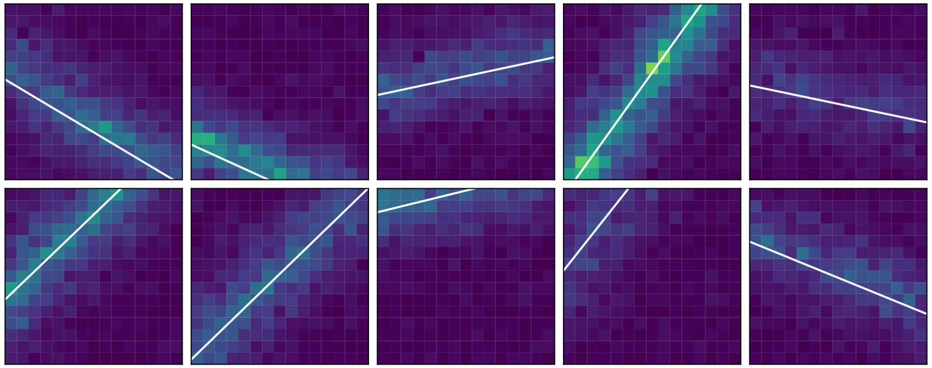
- Lösung: Regularisierung durch Abschneiden von Beiträgen kleiner Eigenwerte



Beispiel: Ein einfacher Pixeldetektor

- Einfaches Beispiel: 2d-Pixeldetektor mit 8x8 Pixeln
- Teilchen fallen aus beliebigen Richtungen mit diversen Energien ein
- Energieverluste der Teilchen etwa proportional zu Energie des Teilchens
- In Folge der Energieverluste werden Photonen im Pixeldetektor deponiert

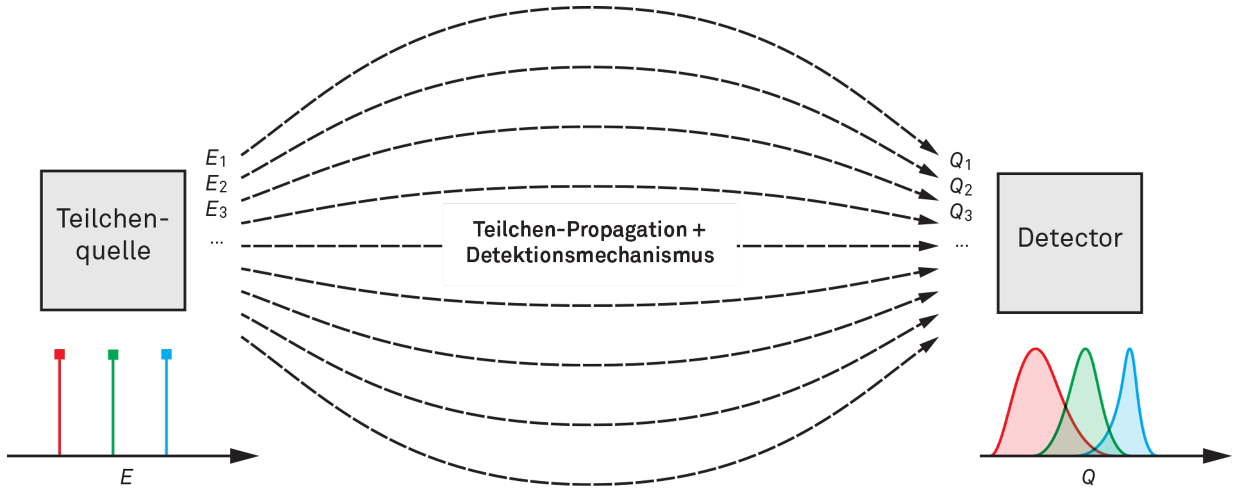
→ Deponierte Ladung im Detektor ist Maß für Energie des Teilchens



10 simulierte Beispielereignisse

Beispiel: Ein einfacher Pixeldetektor

- Problem: Zuordnung zwischen deponierter Ladung und ursprünglichen Energie des Teilchens ist nicht 1:1
- Fragestellung: Welcher Verteilung folgt die Energie des Teilchens an der Quelle?



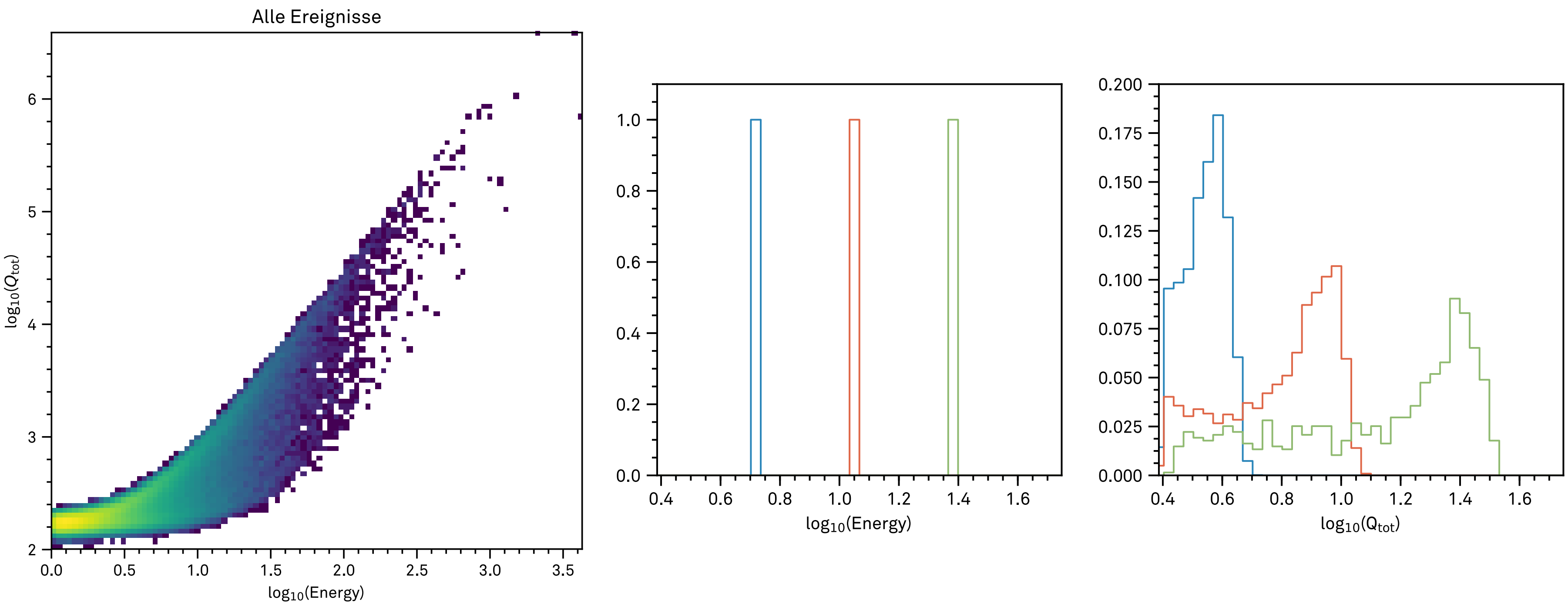
Beispiel: Ein einfacher Pixeldetektor

- Observable: Summe der deponierten Ladung im Detektor
  → Messbare Größe im echten Experiment

- Entfaltungsvariable: Ursprüngliche Energie des detektierten Teilchens
  → Nur verfügbar in Simulationen

→ Nutze Simulationen um Migrationsmatrix zu berechnen

Beispiel: Ein einfacher Pixeldetektor



Binning: Binning des Observablenraumes

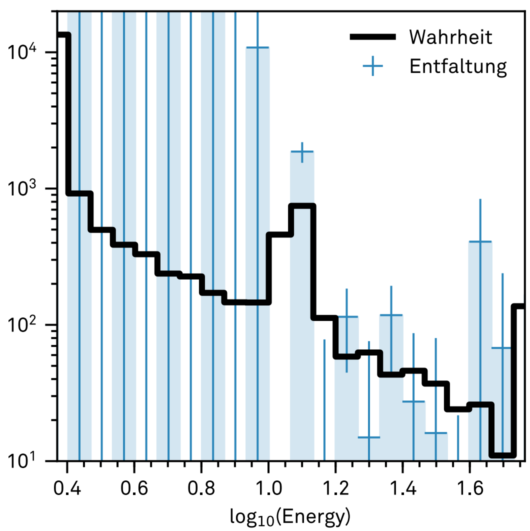
Generelle Bemerkungen:

- Je feiner der Observablenraum gebinnt ist, desto besser bestimmt ist mein Entfaltungsproblem
- Je gröber der Observablenraum gebinnt wird, desto kleiner sind die relativen statistischen Unsicherheiten der Bininhalte

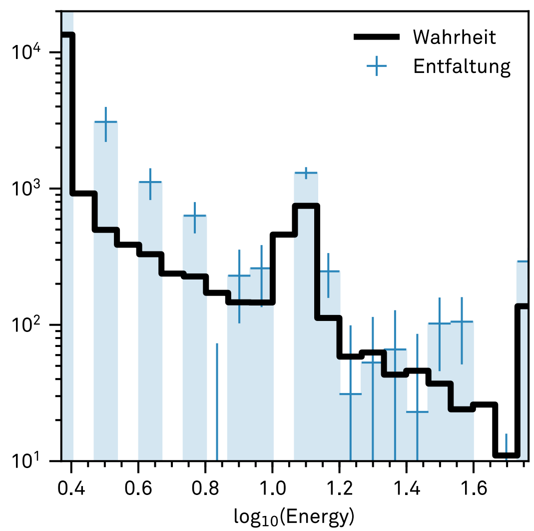
→ Wahl des Binnings ist ein Kompromiss!

Binning: Binning des Observablenraumes

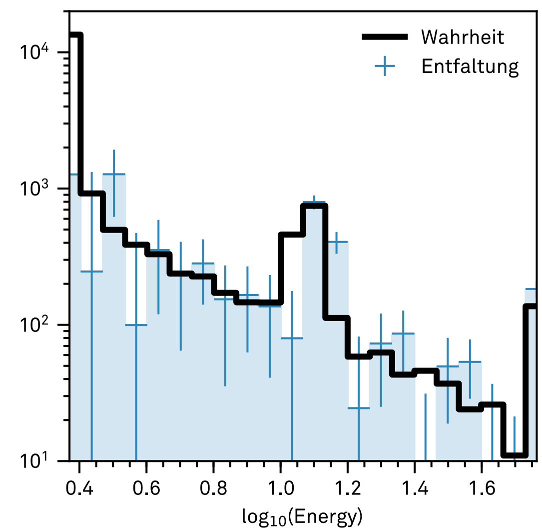
20 Bins



100 Bins



1000 Bins

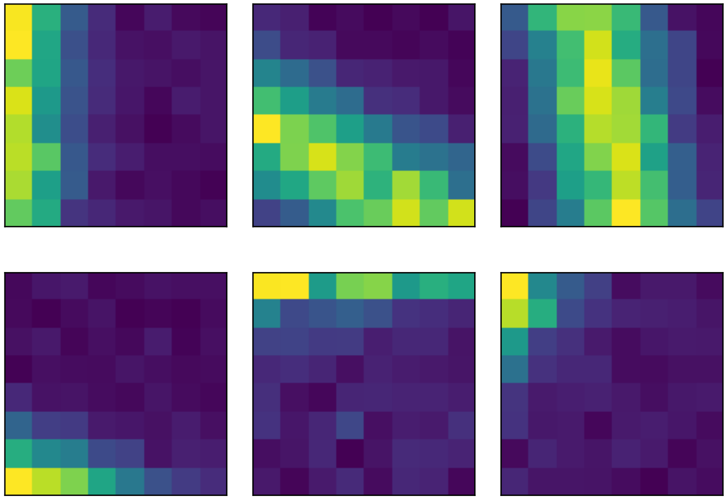


Binning: Mehrdimensionales Binning

- Oft ist es nützlich nicht nur die eine Observable zu betrachten, sondern mehrere Observablen

- Events haben alle die gleiche Energie, aber deutlich unter-schiedliche Gesamtladungen

- Zusammenhang zwischen Gesamtladung und Energie ist abhängig von der Position des Tracks!

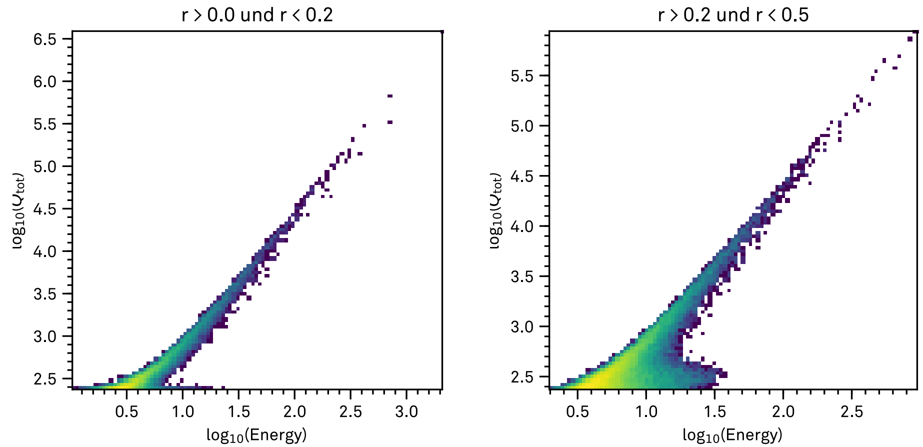


Binning: Mehrdimensionales Binning

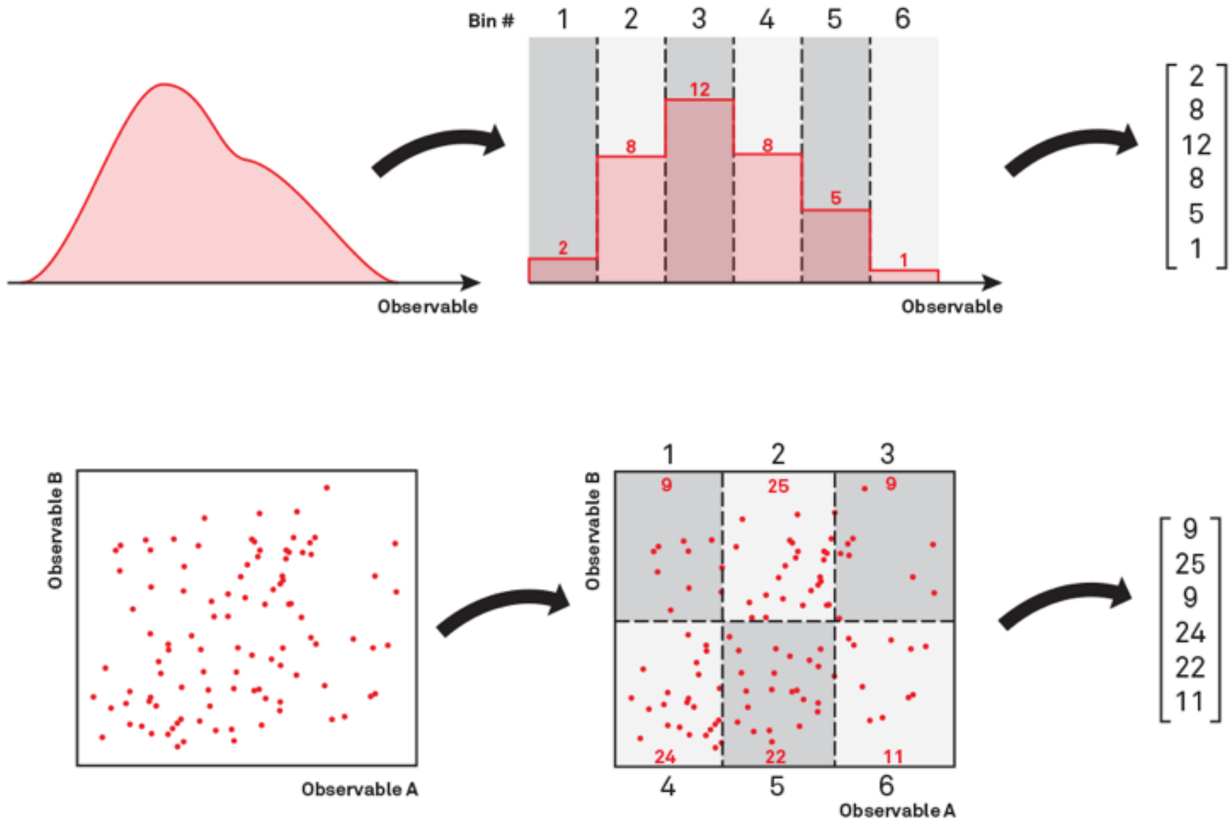
- Oft ist es nützlich nicht nur die eine Observable zu betrachten, sondern mehrere Observablen

- Events haben alle die gleiche Energie, aber deutlich unter-schiedliche Gesamtladungen

- Zusammenhang zwischen Gesamtladung und Energie ist abhängig von der Position des Tracks!



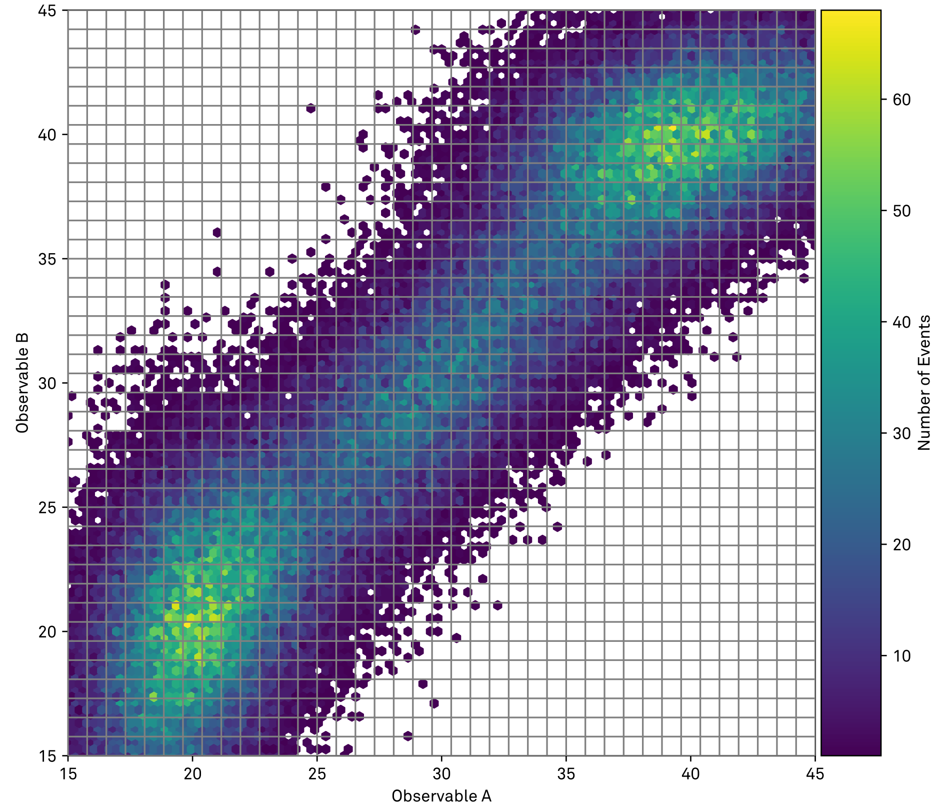
Binning: Mehrdimensionales Binning



Binning: Mehrdimensionales Binning

- Problem: *Curse of Dimensionality*
- Binnt man *d* Observablen in *n* äquidistante Bins, so erhält man insgesamt n^d Bins
- Die meisten dieser Bins sind nur sehr dünn besiedelt, dadurch steigt die statistische Unsicherheit in der Migrationsmatrix

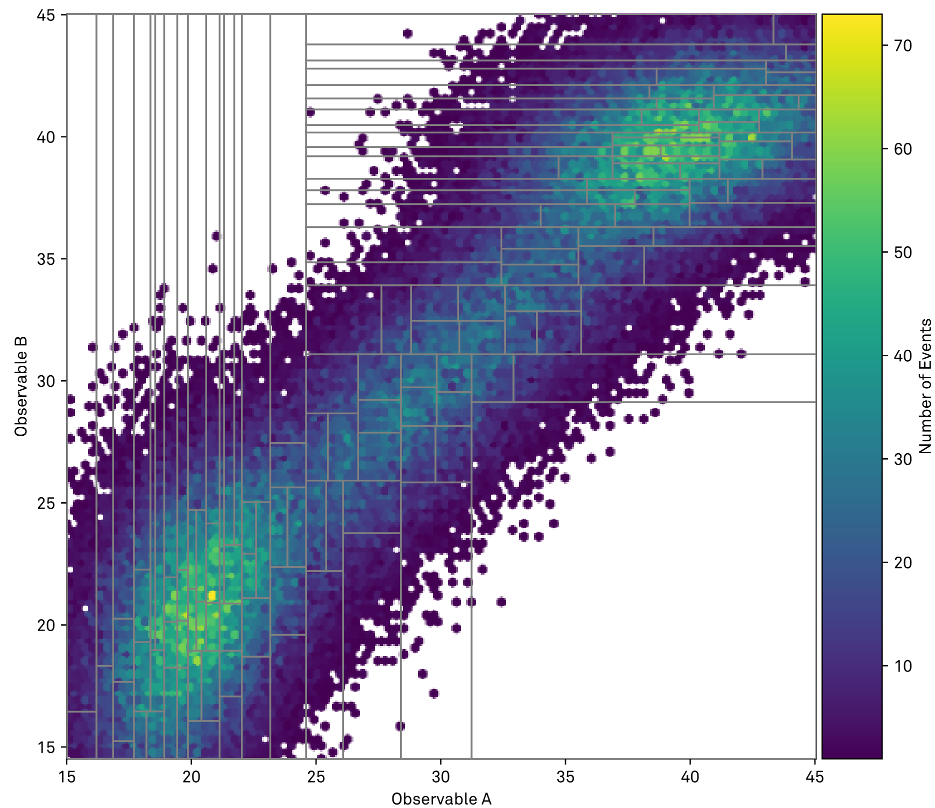
→ Keine skalierbare Lösung



Binning: Mehrdimensionales Binning

Alternative: Binning mit Entscheidungsbäumen

- Nutze einen Entscheidungsbaum um den Observablenraum zu unterteilen
- Die Blätter des Entscheidungsbaum können dann als Bins verwendet werden



Entfaltung als Maximum-Likelihood-Problem

- Schlechte Kondition des Problems macht eine Regularisierung notwendig
- Ein flexibler Ansatz ist notwendig mit dem sich A-Priori-Wissen in statistisch sinnvoller Weise nutzen lässt

→ Maximierung einer Likelihood-Funktion

- **Idee**: Gegeben der statistische Prozess dem die Messung unterliegt, was ist die wahrscheinlichste *wahre Physik*, die diese Messung hervorgerufen hat?

Likelihood-Methoden: Likelihood-Funktion

- Die generelle Form einer Likelihood-Funktion im Falle der Entfaltung ist

$$\mathcal{L} = \mathcal{L}_{\mathrm{stat}} + \mathcal{L}_{\mathrm{reg}}$$

- Der erste Term soll hierbei generell die Übereinstimmung der Messung mit der vorhergesagten Detektor-Antwort gegeben eines Entfaltungsergebnisses bemessen.

- Der zweite Term beinhaltet Regularisierungsterme, die verschiedene Formen haben können. Generell stellen diese Terme Annahmen über die Lösung dar, z.B. Glattheit der Lösung oder unplausible Wertebereiche.

Likelihood-Methoden: Erster Term

- Im allgemeinsten Sinne beschreibt der Term den statistischen Prozess mit dem der Messprozess modelliert wird.
- Dies kann prinzipiell jede Statistik sein, in der Praxis jedoch sind einige Modelle von besonders großer Bedeutung.
- Die Summe der quadratischen Abweichungen:

$$\mathcal{L}_{\mathrm{MSE}} = \sum_i (\mathbf{g} - \mathrm{A} \cdot \mathbf{f})_i^2 = (\mathbf{g} - \mathrm{A} \cdot \mathbf{f})^\top (\mathbf{g} - \mathrm{A} \cdot \mathbf{f})$$

- Hierbei wird der Messprozess als Gaußscher Prozess angenommen, wobei die Kovarianz eine Einheitsmatrix ist und der Erwartungswert die vorhergesagte Detektorantwort.

Likelihood-Methoden: Erster Term

- Der häufigste Anwendungsfall der Entfaltung sind Zählexperimente. Hierbei stellen die Elemente des Vektors **g** Zählraten dar.
- Die naheliegenste statistische Beschreibung eines solchen Experiments ist die Poisson-Statistik:

$$\mathcal{L}_{\mathrm{Poisson}} = \sum_i g_i \log\left((\mathrm{A}\cdot\mathbf{f})_i\right) - (\mathrm{A}\cdot\mathbf{f})_i$$

Least Squares: Analytische Lösung

- Die Least-Squares-Likelihood hat dabei eine analytische Lösung. Für quadratische Matrizen ist diese trivial:

$$\nabla \log \mathcal{L} = 2(A \cdot \mathbf{f} - \mathbf{g}) \stackrel{!}{=} 0 \Longrightarrow \mathbf{f} = A^{-1}\mathbf{g}$$

- Für nicht-quadratische Matrizen muss man einen kleinen Umweg gehen:

$$\begin{aligned}\nabla \log \mathcal{L} = 2(A \cdot \mathbf{f} - \mathbf{g}) \stackrel{!}{=} 0 &\Longrightarrow A^\top A \cdot \mathbf{f} = A^\top \mathbf{g} \\ &\Longrightarrow \mathbf{f} = (A^\top A)^{-1} A^\top \mathbf{g} = A^+ \mathbf{g}\end{aligned}$$

Least Squares: Singular Value Decomposition (SVD)

- A^+ wird auch als Pseudoinverse bezeichnet und lässt sich als Verallgemeinerung der Inversen verstehen
- Ähnlich kann man auch die Eigenwertzerlegung verallgemeinern, die sog. Singulärwertzerlegung (Singular Value Decomposition, SVD)

$$A \cdot \mathbf{f} = \mathrm{U}\Sigma\mathrm{V}^{\top} \cdot \mathbf{f} = \mathbf{g} \Longleftrightarrow \mathrm{V}\Sigma^{+}\mathrm{U}^{\top} \cdot \mathbf{g} = \mathbf{f}$$

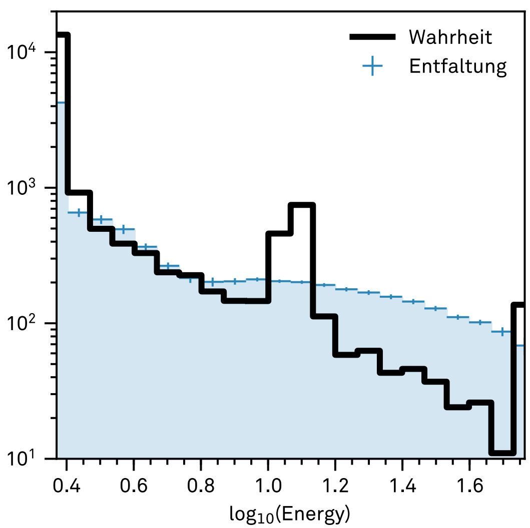
Least Squares: Singular Value Decomposition (Regularisierung)

- In dieser Zerlegung lässt sich nun eine Regularisierung einführen, indem kleine Singulärwerte unterdrückt werden:

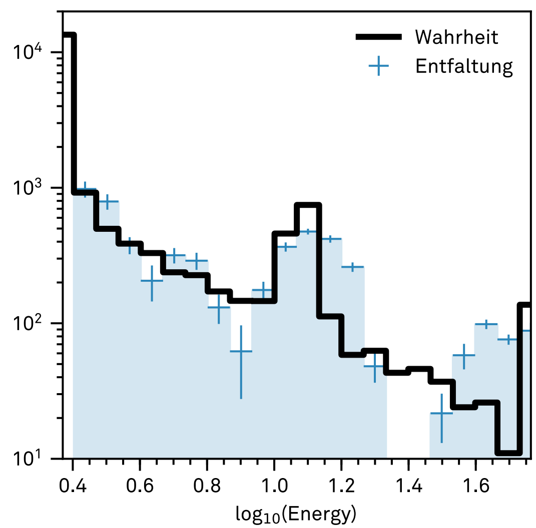
$$A^{+} \cdot \mathbf{g} = \mathrm{V}\Sigma^{+}\mathrm{U}^{\top} \cdot \mathbf{g} = \mathbf{f} \longrightarrow A^{+}_{\mathrm{reg}} \cdot \mathbf{g} = \mathrm{V}\Sigma^{+}\mathrm{diag}(\tau)\mathrm{U}^{\top} \cdot \mathbf{g} = \mathbf{f}$$

SVD-Entfaltung

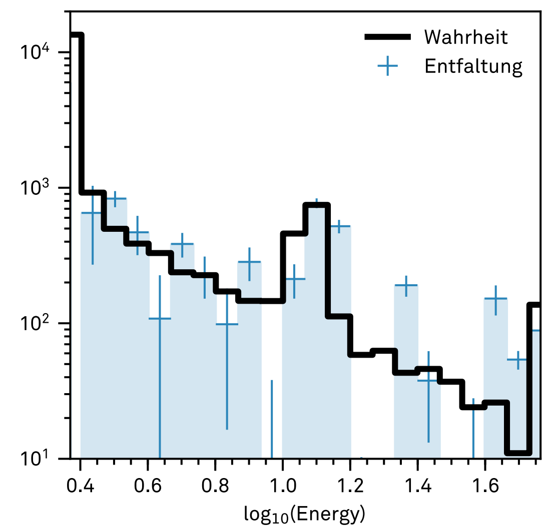
Größte 5 Singulärwerte



Größte 10 Singulärwerte



Größte 15 Singulärwerte



Regularisierungsstärke

Iterative Bayesian Unfolding

- In diesem Ansatz wird der Satz von Bayes benutzt um die Inverse der Migrationsmatrix iterativ anzunähern (siehe Vorlesung zum Thema *Schätzen*)
- Hierzu schreiben wir die Entfaltungsgleichung wie folgt um:

$$\mathbf{g} = \mathrm{A} \cdot \mathbf{f} \Longleftrightarrow g(y_j) = \sum_j A(y_j|x_i) f(x_i)$$

- Die Migrationsmatrix A_{ij} wird mit Hilfe von Simulationsdaten berechnet, wobei durch die spaltenweise Normierung diese explizit modellunabhängig ist:

$$A(y_j|x_i) = \frac{A(x_i, y_j)}{\sum_j A(x_i, y_j)} = \frac{A(x_i, y_j)}{f(x_i)}$$

Iterative Bayesian Unfolding

- In dieser Formulierung kann man die Lösung der Entfaltungsgleichung wie folgt schreiben:

$$f(x_i) = \sum_i B(x_i|y_j)g(y_j) \tag{1}$$

- Hierbei tritt jedoch das Problem auf, dass B_{ji} nicht mehr modellunabhängig ist, d.h. die Lösung fällt unterschiedlich aus, je nachdem mit welchen Annahmen über den Fluss simuliert wurde.

Iterative Bayesian Unfolding

- Nutze nun den Satz von Bayes mit dem *f* aus der vorherigen Berechnung als Prior, um erneut die Matrix *B* zu berechnen:

$$B(x_i|y_j) = \frac{A(y_j|x_i)\hat{f}(x_i)}{\sum_i A(y_j|x_i)\hat{f}(x_i)} \qquad (2)$$

- Dieses Verfahren kann nun iterativ fortgeführt werden, indem die Matrix aus (2) wieder in (1) eingesetzt wird um die nächste Iteration von *f* zu berechnen.

Iterative Bayesian Unfolding

- Dieses Vorgehen bezeichnet man als *Iterative Bayesian Unfolding:*

$$B^{(k+1)}(x_i|y_j) = \frac{A(y_j|x_i)f^{(k)}(x_i)}{g(y_j)}$$

$$f^{(k+1)}(x_i) = \sum_i B^{(k+1)}(x_i|y_j)g(y_j)$$

- Die Abbruchsbedingung ist entweder Konvergenz oder eine festgelegte Schrittzahl

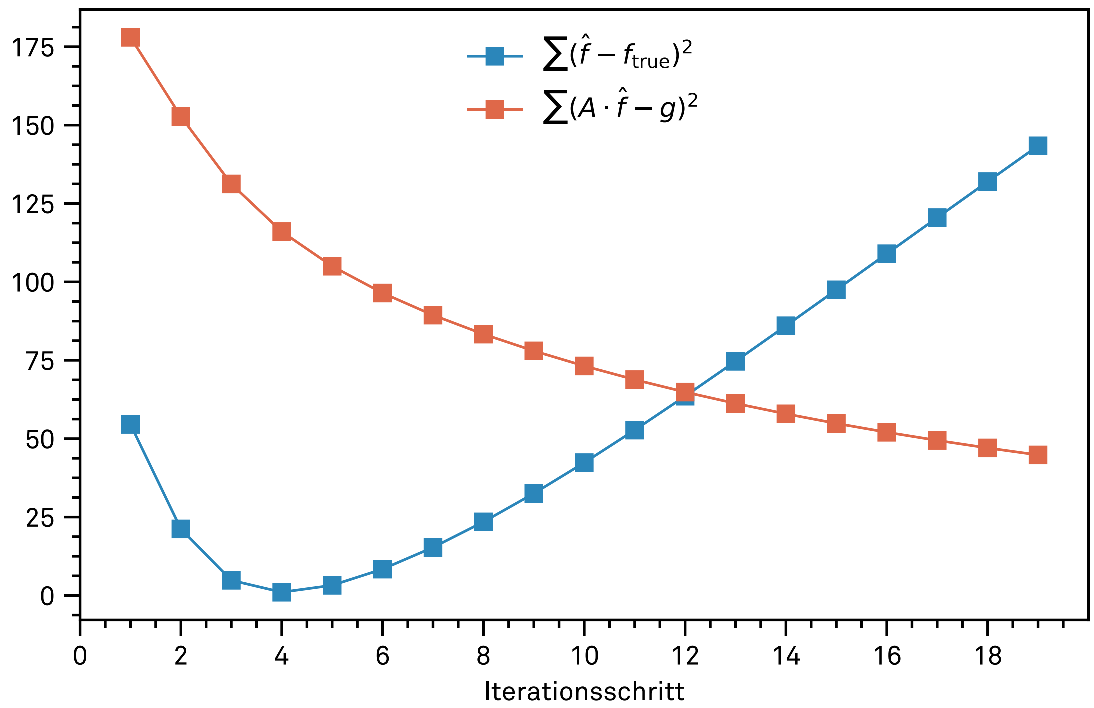
Iterative Bayesian Unfolding: Konvergenz?

- Der Algorithmus lässt sich auch auffassen als Minimierung einer einfachen Least-Squares-Likelihood

- Das hat zur Folge, dass bei Konvergenz des Algorithmus die gleichen Probleme wie bei der unregularisierten SVD-Entfaltung auftreten (Oszillationen etc)

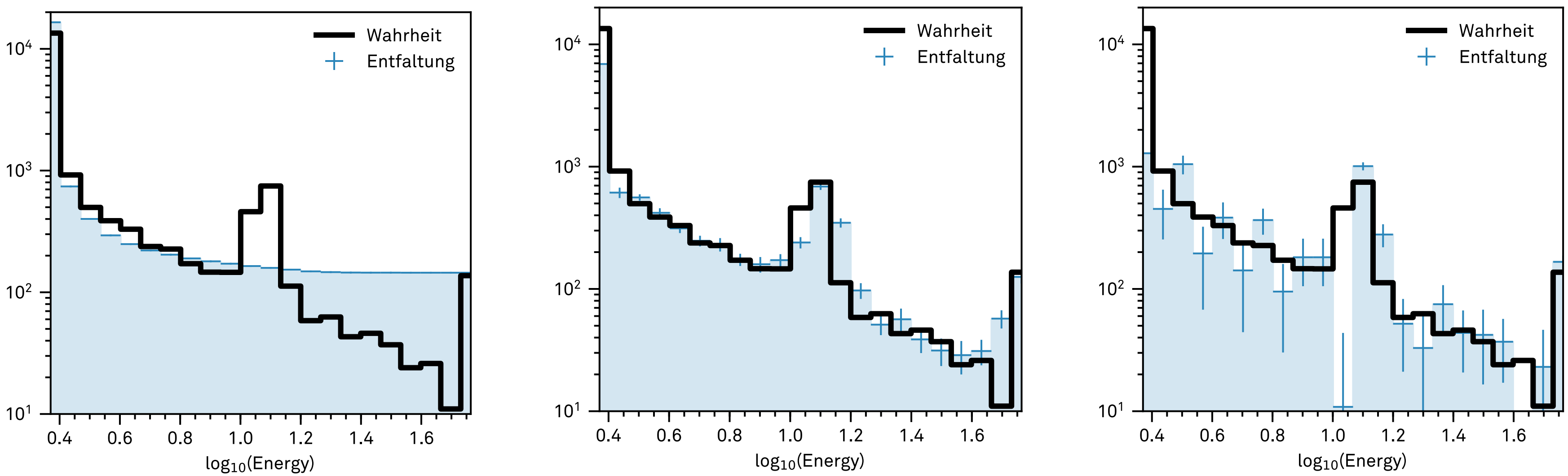
- Jedoch konvergieren die großen Eigenwerte schneller, als die Kleinen
  → Früher Abbruch führt zu Vermeidung von Artefakten

Iterative Bayesian Unfolding



Iterative Bayesian Unfolding

k = 1 k = 5 k = 200



Regularisierungsstärke

Likelihood-Methoden: Zweiter Term

- Die Einführung zusätzlicher Annahmen über die Lösung des Entfaltungsproblems kann helfen, das Problem besser zu konditionieren.
- Generell gilt die Regel: So wenige Annahmen wie möglich, so viele wie nötig.
- Eine Möglichkeit ist es, den Wertebereich der Lösung einzuschänken, z.B. weil bekannt ist, dass physikalische Flüsse nicht negativ sind:

$$\mathcal{L}_{+}(\mathbf{f}) = \begin{cases} \infty & \text{wenn mindestens ein } f_i < 0 \\ 0 & \text{sonst} \end{cases}$$

Likelihood-Methoden: Tikhonov-Regularisierung

- Oft ist es hilfreich eine glatte Lösung zu fordern. Hierzu definiert man "Glattheit" als eine durchweg kleine zweite Ableitung.
- Numerisch kann die zweite Ableitung eines Vektors durch Anwendung einer Matrix (s. regularisierte kleinste Quadrate) erreicht werden:

$$C = \begin{bmatrix} 1 & -1 & 0 & \dots & & \\ -1 & 2 & -1 & \dots & & \\ 0 & -1 & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & -1 & 0 \\ & & \dots & -1 & 2 & -1 \\ & & \dots & 0 & -1 & 1 \end{bmatrix}$$

Likelihood-Methoden: Tikhonov-Regularisierung

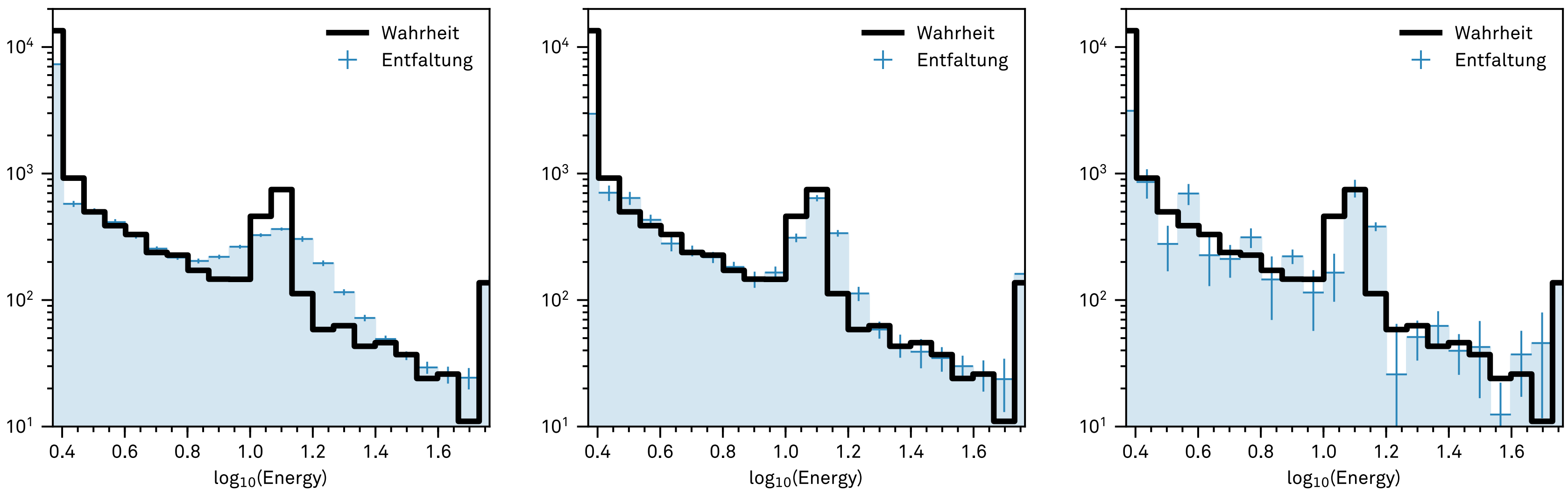
- Anschließend wird die Summe der Quadrate der zweiten Ableitungen als Maß für die Glattheit der Lösung berechnet:

$$\mathcal{L}_{\mathrm{tikh}}(\mathbf{f}) = \frac{\tau}{2} \sum_i \|\mathrm{C} \cdot \mathbf{f}\|_i^2 = \frac{\tau}{2} \mathbf{f}^\top \mathrm{C}^\top \mathrm{C}\mathbf{f}$$

- Der Vorfaktor τ wird als Regularisierungsstärke bezeichnet und ist ein Maß dafür, wie stark Nicht-Glattheit bestraft wird.
- Der Term lässt sich auch als Gaußscher Prior auf die zweiten Ableitungen verstehen. (→ Bayesianische Statistik)

Likelihood-Methoden: Numerische Minimierung

τ = 100 τ = 6 τ = 0



Regularisierungsstärke

Validierung: Varianz und Bias

- Regularisierungen sind Annahmen, die man über das Spektrum macht, die dazu führen, dass bestimmte Bereiche der Lösung ausgeschlossen oder unterdrückt werden

- Da das Spektrum im Regelfall unbekannt ist, passen die Annahmen, die man einführt nicht zwingend zum vorliegenden Datensatz

→ Falsche Annahmen über das Ergebnis sorgen für einen Bias

→ Einschränkung der Lösung führt zu einer Unterschätzung der Fehler

Validierung: Varianz und Bias

- Um abzuschätzen wie ausgeprägt diese systematischen Effekte sind, kann man auf Grundlage von Simulationen das Entfaltungsergebnis mit der Wahrheit vergleichen. Hier zu kann folgende Teststatistik berechnet werden:

$$\Lambda_i = \frac{\hat{f}_i - f_i}{\sigma_{\hat{f}}}$$

- Diese Abweichung des entfalteten Bininhalts von der Wahrheit, normiert auf den geschätzten Fehler, sollte im Fall eines erwartungstreuen Schätzers standardnormalverteilt sein.

→ Nur nicht-regularisierte Entfaltung ist erwartungstreu

Validierung: Varianz und Bias