

Westfälische Wilhelms-Universität Münster

Institut für Theoretische Physik
Institut für Biomagnetismus und
Biosignalanalyse

Bachelorarbeit

Analyse nichtstationärer stochastischer Prozesse
am Beispiel von EEG Daten

von
Yvonne Krumbeck*

23. Juli 2015

Gutachter:
Dr. Oliver Kamps
PD Dr. Carsten Wolters

* y_krum01@uni-muenster.de, Matrikel-Nr.: 381 616

Inhaltsverzeichnis

1	Einleitung	1
2	Theoretische Grundlagen	3
2.1	Nichtstationäre stochastische Prozesse	3
2.1.1	Grundlegende Begriffe	3
2.1.2	Die Langevin-Gleichung	4
2.2	Wahrscheinlichkeitsdichteschätzung	5
2.2.1	Histogrammschätzer	6
2.2.2	Kerndichteschätzer	6
2.3	Nichtparametrische Regression	7
2.3.1	Der bedingte Erwartungswert	7
2.3.2	Lokal Polynomiale Kernel-Regression	8
2.3.3	Multivariate Regression	9
2.4	Elektroenzephalografie (EEG)	11
2.4.1	Neurophysiologischer Hintergrund	11
2.4.2	Aufnahme von EEG Daten	12
3	Analyse synthetischer Daten	13
3.1	Erzeugung synthetischer Daten	13
3.2	Vergleich der 1-dimensionalen Regressionsmethoden	13
3.3	Vergleich der 2-dimensionalen Regressionsmethoden	21
3.4	Nichtstationärer Langevin-Prozess	24
4	Anwendung auf EEG Daten	28
5	Zusammenfassung und Diskussion	33

1 Einleitung

Komplexe Systeme setzen sich aus vielen einzelnen Systemen zusammen, deren Zusammenwirken zu einer selbstorganisierten Ordnung auf der makroskopischen Skala führen kann. Diese Ordnungszustände lassen sich meist durch wenige Variablen bzw. Differentialgleichungen beschreiben. Beispiele für solche komplexen Systeme finden sich von der Physik über die Chemie bis hin zur Biologie und Medizin [1, 2]. Auch das neuronale Netzwerk des Gehirns gehört in diese Klasse von Systemen. Erst das Zusammenwirken größerer Neuronenpopulationen erzeugt übergeordnete Strukturen in Form von Gedanken oder der motorischen Steuerung von Muskelgruppen, die sich aus der reinen Superposition des Verhaltens einzelner Neuronen nicht verstehen lassen.

Ein komplexes System kann verschiedene Ordnungszustände einnehmen. Die Übergänge zwischen diesen Zuständen werden auch als Phasenübergänge bezeichnet. Besonders in der Nähe dieser Phasenübergänge kann die Beschreibung komplexer Systeme auf wenige Dimensionen reduziert werden [1, 2].

Man nimmt an, dass auch einer pathologischen Störung wie dem Übergang zwischen normalem Verhalten und einem epileptischen Anfall beim menschlichen Gehirn ein solcher Phasenübergang zugrunde liegt. Mittels statistischer Analyseverfahren kann versucht werden, die deterministischen sowie stochastischen Anteile der niedrigdimensionalen Dynamik des Systems zu charakterisieren [3, 4, 5]. Die Hoffnung ist, dass sich so aus den Messdaten z.B. des EEG-Signals Vorzeichen für einen epileptischen Anfall gewinnen lassen [6, 7].

Die Analyse steht dabei meist vor einem wichtigen Problem: die zur Verfügung stehenden Daten sind oft durch verschiedene Faktoren nur in begrenzten Mengen vorhanden. Eine sich zeitlich schnell ändernde Dynamik bringt zum Beispiel auch weniger Daten pro Zeitintervall mit sich. Dies ist vor allem bei biologischen Vorgängen zu beobachten, wo die einzelnen Prozesse innerhalb von wenigen Millisekunden ablaufen.

Die vorliegende Arbeit befasst sich mit dem Vergleich vorhandener Methoden, wie die histogramm- und kernelbasierte Regression. Dieser wurde bereits von Lamouroux und Lehnertz in [6] durchgeführt. Infolge der Ergebnisse wurde die Regression mittels eines *lokal konstanten Kernschätzers* als effektivere Methode - vor allem bei einer geringen Datenmenge - eingestuft. In dieser Arbeit wird zusätzlich ein erweiterndes Konzept, die *lokal lineare Regression*, untersucht und mit den vorherigen Methoden verglichen. Zudem wird die Zeitreihenanalyse zu einem zweidimensionalen Problem erweitert, sodass im Gegensatz zu Lamouroux und Lehnertz die Zeitfenster, innerhalb derer der vorliegende Prozess als stationär betrachtet wird, ebenfalls mit einer Kernelgewichtung in die Analyse eingehen. Damit soll eine mögliche Verbesserung der Datenanalyse bei sich zeitlich schnell ändernden Dynamiken

und geringer Datendichte erreicht werden.

Zur Auswertung der Vergleiche werden zunächst durch numerische Simulation generierte Daten verwendet. Anschließend sollen die erweiterten Methoden beispielhaft auf EEG Datenreihen angewendet werden, um daran mögliche Erkenntnisgewinne für die weitere Forschung zu zeigen.

2 Theoretische Grundlagen

2.1 Nichtstationäre stochastische Prozesse

Komplexe Systeme lassen sich wegen der hohen Freiheitsgrade nicht mikroskopisch beschreiben. Stattdessen geht man von einem *stochastischen* Verhalten aus, über dessen zugrunde liegende Dynamik mittels statistischer Analyseverfahren Kenntnisse gewonnen werden können.

Einige dynamische Prozesse verändern sich auf sehr kurzen Zeitskalen. Beispielsweise finden Wechselwirkungen der Neuronen im menschlichen Gehirn in Bruchteilen von Millisekunden statt. Einem solchen System liegen also sich fortlaufend ändernde Dynamiken zugrunde, die dementsprechend nur auf kürzeren Zeitskalen mit weniger Daten analysiert werden können. Aus diesem Grund beschäftigt sich diese Arbeit mit *nichtstationären stochastischen Prozessen*.

2.1.1 Grundlegende Begriffe

Im Folgenden werden zunächst Begriffe und Konzepte erläutert, die dieser Arbeit zugrunde liegen.

Zeitreihe: Als *Zeitreihe* bezeichnet man eine zeitlich geordnete Menge von Datenpunkten $X_j = X(t_j)$, die in diskreten Abständen zum Zeitpunkt t_j gemessen oder erhoben worden sind.

Stochastischer Prozess: Zeitreihen eines nicht-deterministischen Systems bezeichnet man als *stochastischen Prozess*. Dabei handelt es sich um Vorgänge mit zufälligem Zuwachs $\epsilon_j = \epsilon(t_j)$. Die Realisationen des Prozesses folgen dem deterministischen *Drifterm* $D(x)$, der durch die stochastische *Diffusion* gestört wird.

Ein einfaches Beispiel für einen solchen Prozess ist die *Brownsche Bewegung* [8]:

$$\ddot{x} = -\gamma\dot{x} + \epsilon \tag{1}$$

Sie beschreibt die Trajektorie eines Teilchens innerhalb eines Fluids, welches durch Stöße in eine zufällige Richtung abgelenkt wird.

Nichtstationärer Prozess: Bei einem *nichtstationären Prozess* ist zusätzlich eine Zeitabhängigkeit vorhanden. Dadurch ist die Entwicklung der Realisationen nicht gleichbleibend (stationär) sondern folgt einer sich zeitlich ändernden Dynamik. Als Beispiel ist eine solche Zeitreihe in Abbildung 1 dargestellt.

Markov-Prozess: Die besondere Eigenschaft von *Markov-Prozessen* ist die stochastische Verteilung “ohne Gedächtnis”. Das bedeutet, dass der zufällige Zuwachs

unabhängig vom bisherigen Verlauf des Prozesses ist. Ausgedrückt mit der Realisationswahrscheinlichkeit $P_{i,j}(x_i, t_i | x_j, t_j)$, die angibt mit welcher Wahrscheinlichkeit x_i zur Zeit t_i realisiert wird unter der Bedingung, dass $x_j(t_j)$ bereits realisiert worden ist, erhält man formal:

$$P_{n,1}(x_n, t_n | x_{n-1}, t_{n-1}; \dots; x_1, t_1) = P_{n,n-1}(x_n, t_n | x_{n-1}, t_{n-1}) \quad (2)$$

für beliebige n bei einer Zeitreihe mit $t_1 < t_2 < \dots < t_n$. Somit ist also die Realisation von $x_n(t_n)$ ausschließlich abhängig von der Realisation x_{n-1} zum Zeitpunkt t_{n-1} und nicht von den vorherigen $t_i < t_{n-1}$ [3, 4].

Makroskopisch gesehen können für viele stochastische Prozesse Markoveigenschaften angenommen werden, daher konzentriert sich diese Arbeit auf Markov-Prozesse.

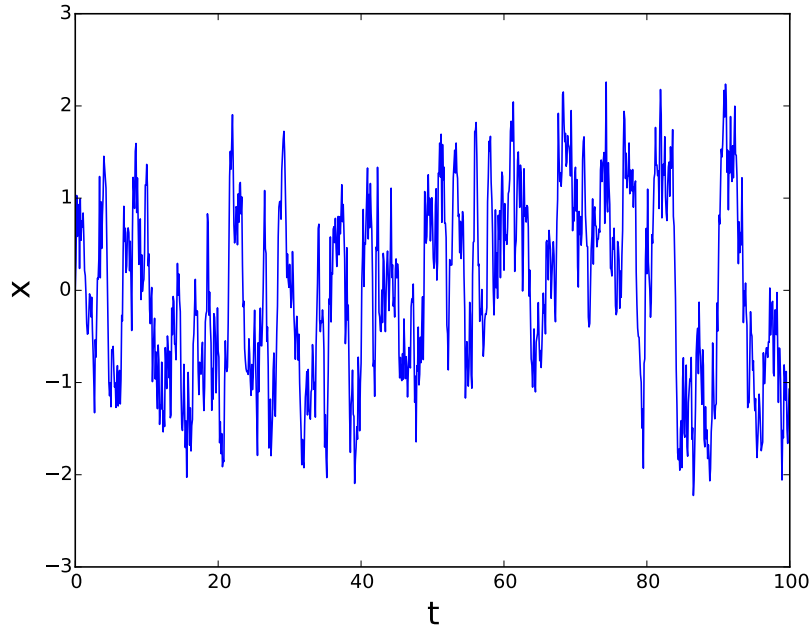


Abbildung 1: Beispiel einer generierten Zeitreihe mit zeitabhängigem Drift.

2.1.2 Die Langevin-Gleichung

Stochastische Prozesse, wie sie in dieser Arbeit behandelt werden, lassen sich durch die (eindimensionale) Langevin-Gleichung¹ beschreiben [5, 7] :

$$\dot{x}(t) = D^{(1)}(x, t) + \sqrt{2D^{(2)}(x, t)}\epsilon(t). \quad (3)$$

¹In mancher Literatur findet man, dass $\sqrt{2}$ mit in den Diffusionsterm eingebracht wird oder die Bedingungen für ϵ entsprechend verändert werden. Dabei ist zu beachten, dass sich die Beziehung zur Fokker-Planck-Gleichung (5) ändert.

Dabei stellt $\epsilon(t)$ ein gaußverteiltes statistisches Rauschen dar, mit folgenden Eigenschaften [3, 9]:

$$\langle \epsilon(t) \rangle = 0 \quad (4a)$$

$$\langle \epsilon(t)\epsilon(t') \rangle = \delta(t - t'). \quad (4b)$$

Die Langevin-Gleichung lässt sich äquivalent über die Wahrscheinlichkeitsdichte $f(x, t)$ (siehe Abschnitt 2.2) mit der Fokker-Planck-Gleichung ausdrücken [8, 4]:

$$\frac{\partial}{\partial t} f(x, t) = \left[-\frac{\partial}{\partial x} D^{(1)}(x, t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x, t) \right] f(x, t). \quad (5)$$

Dabei werden die *Kramers-Moyal-Koeffizienten* $D^{(1)}(x, t)$ und $D^{(2)}(x, t)$ im Bezug auf Langevin-Prozesse als Drift- und Diffusionsfunktionen bezeichnet und stehen für den deterministischen bzw. stochastischen Beitrag [5, 10]. Nach [8, 4, 5] können die Kramers-Moyal-Koeffizienten für endliche Zeitschritte τ direkt aus einer gemessenen Zeitreihe geschätzt werden durch:

$$\hat{D}^{(n)}(x, t) = \lim_{\tau \rightarrow 0} \frac{1}{n! \tau} \langle [X(t + \tau) - X(t)]^n | X(t) = x \rangle \quad (6)$$

wobei $n = 1, 2$.

Um den bedingten Erwartungswert

$$m(x) = \langle Y | X = x \rangle \quad (7)$$

zu bestimmen, wird eine Regression basierend auf Histogramm- und Kernschätzern, die im Folgenden erläutert werden, durchgeführt.

2.2 Wahrscheinlichkeitsdichteschätzung

Seien $X_1, \dots, X_N$ Datenpunkte einer Zeitreihenmessung und $f(x)$ die zugrundeliegende Wahrscheinlichkeitsdichtefunktion (engl.: probability density function, kurz: pdf). Liegen keine Informationen über $f(x)$ vor, lässt sich diese mit Hilfe von nicht-parametrischen Dichteschätzern, wie dem Histogramm- und Kernschätzer, aus der Datenreihe ermitteln.

Die folgenden Abschnitte (2.2 und 2.3) folgen, wenn nicht anders angegeben, der Darstellung in [11].

2.2.1 Histogrammschätzer

Für die Konstruktion eines Histogramms wählt man zunächst einen Startpunkt x_0 und definiert n Intervalle, sogenannte *Bins*, der Breite h :

$$B_l = [x_0 + (l - 1)h; x_0 + lh], \quad l \in \mathbb{Z}. \quad (8)$$

Nun bestimmt man die Anzahl aller Datenpunkte, die sich jeweils in dem Bin B_l befinden. Um die relativen Häufigkeiten zu erhalten, wird die Anzahl jeweils mit Nh dividiert. Formal erhält man dann für die geschätzte pdf:

$$\hat{f}_H(x) = \frac{1}{Nh} \sum_{j,l=1}^{N,n} I(x \in B_l) I(X_j \in B_l) \quad (9)$$

mit der Indikatorfunktion:

$$I(x \in B_l) = \begin{cases} 1 & \text{wenn } x \in B_l, \\ 0 & \text{sonst.} \end{cases} \quad (10)$$

Die Qualität der Schätzung hängt stark von der gewählten Binbreite h ab. Sie entscheidet über die Glättung von $\hat{f}_H(x)$, wobei eine zu groß gewählte Breite die Auflösung möglicher Strukturen von $\hat{f}_H(x)$ verhindert (vgl. [11, S. 25]). Nimmt man eine Gaußsche Normalverteilung an, gilt:

$$h_{opt} \approx 3.49\hat{\sigma}N^{-\frac{1}{3}} \quad (11)$$

wobei $\hat{\sigma}$ eine Schätzung der Standardabweichung ist. Dies wird in vielen Fällen als Daumenregel verwendet, auch wenn die Verteilungsfunktion zunächst nicht bekannt ist [3].

Es existiert außerdem eine Abhängigkeit vom gewählten Startpunkt x_0 . Die Form der geschätzten pdf kann stark mit der Verschiebung der festgelegten Bins variieren (vgl. [11, S. 30]).

2.2.2 Kerndichteschätzer

Anders als beim Histogramm werden keine Intervalle vorher festgelegt. Vielmehr wird um jede Stelle x der zu schätzenden Funktion $\hat{f}_K(x)$ ein Intervall gelegt. Somit ergibt sich bereits ein Vorteil gegenüber dem Histogrammschätzer, da die geschätzte Funktion nun unabhängig von einem zu wählenden Startpunkt ist.

Nun wird für jedes X_j mit Hilfe einer sogenannten Kernfunktion $K_h(x - X_j)$ ein

Gewicht bestimmt und dem jeweiligen Intervall um x zugeordnet. Dabei ist

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right) \quad (12)$$

mit der Bandbreite h . Wie die Binbreite beim Histogrammschätzer muss diese zuvor gewählt werden. Auch hier hängt die schließlich geschätzte pdf $\hat{f}_K(x)$ von der Wahl für h ab. Hierzu wird folgende Daumenregel von Silverman verwendet:

$$h_{opt} \approx 1.06\hat{\sigma}N^{\frac{1}{5}}. \quad (13)$$

Wie in (11) basiert diese auf der Annahme einer Gaußschen Normalverteilung.

Für $K(u)$ können beliebige, üblicherweise symmetrische Funktionen mit der Eigenschaft

$$\int_{-\infty}^{\infty} K(u) du = 1 \quad (14)$$

benutzt werden. Der *Epanechnikov-Kern* hat sich dabei als besonders effizient erwiesen und wird auch in dieser Arbeit verwendet. Er ist wie folgt definiert:

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{für } |u| \leq 1, \\ 0 & \text{sonst.} \end{cases} \quad (15)$$

Der Kernschätzer ist schließlich definiert als:

$$\hat{f}_K(x) = \frac{1}{Nh} \sum_{j=1}^N K\left(\frac{x - X_j}{h}\right).$$

Im Gegensatz zum Histogramm erhält man nun eine stetige Funktion, die zudem die Eigenschaft der Differenzierbarkeit von der Kernfunktion erhält.

2.3 Nichtparametrische Regression

2.3.1 Der bedingte Erwartungswert

Wie zuvor erwähnt, wird zur Schätzung der Kramers-Moyal-Koeffizienten der bedingte Erwartungswert benötigt (siehe (6)). Dieser gibt an, wie die üblicherweise durch eine Zufallsvariable ϵ gestörte abhängige Variable Y einer Datenreihe mit der unabhängigen X zusammenhängt. Formal lässt sich schreiben:

$$Y = m(X) + \epsilon \quad (16a)$$

$$m(x) = \langle Y | X = x \rangle. \quad (16b)$$

Es gilt also für die Datenreihe im Mittel die Beziehung $Y = m(X)$, sofern $X = x$ [4, 11].

Um den bedingten Erwartungswert berechnen zu können wird folgende Definition verwendet:

$$m(x) = \int y \frac{f(x, y)}{f_X(x)} dy \quad (17)$$

wobei $f(x, y)$ die gemeinsame Dichtefunktion und $f_X(x)$ die Randdichte im Bezug auf X sind. Sie können, wie in Abschnitt 2.2 beschrieben, aus der Datenreihe geschätzt werden.

Für die Kernel basierte Regression ergibt sich somit der *Nadaraya-Watson Schätzer* (siehe [11, S. 89]):

$$\hat{m}_h(x) = \frac{N^{-1} \sum_{j=1}^N K_h(x - X_j) Y_j}{N^{-1} \sum_{j=1}^N K_h(x - X_j)}. \quad (18a)$$

Analog folgt für den Histogrammschätzer [6, 4]:

$$\hat{m}_h(x) = \frac{\sum_{j,l=1}^{N,n} I(x \in B_l) I(X_j \in B_l) Y_j}{\sum_{j,l=1}^{N,n} I(x \in B_l) I(X_j \in B_l)} \quad (19)$$

2.3.2 Lokal Polynomiale Kernel-Regression

Die oben beschriebene kernelbasierte Regression mit dem Nadaraya-Watson-Schätzer (18) ist als Spezialfall eines allgemeineren Konzepts der Kernel-Regression zu verstehen. Gemeint ist die *lokal polynomiale Regression*, die basierend auf der Methode der kleinsten Quadrate (*least squares fit*), in einer Umgebung um x ein Polynom vom Grad p mit einer Kernelgewichtung schätzt. Gesucht sind also die Koeffizienten $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$, die

$$\sum_{j=1}^N [Y_j - \beta_0 - \beta_1(X_j - x) - \dots - \beta_p(X_j - x)^p]^2 K_h(x - X_j) \quad (20)$$

minimieren.

Ausgedrückt in Matrizen lässt sich das Problem wie folgt formulieren:

$$\hat{\beta}(x) = \begin{pmatrix} S_0(x) & S_1(x) & \dots & S_p(x) \\ S_1(x) & S_2(x) & \dots & S_{p+1}(x) \\ \vdots & \vdots & \ddots & \vdots \\ S_p(x) & S_{p+1}(x) & \dots & S_{2p}(x) \end{pmatrix}^{-1} \begin{pmatrix} T_0(x) \\ T_1(x) \\ \vdots \\ T_p(x) \end{pmatrix} \quad (21a)$$

mit

$$S_i(x) = \sum_{j=1}^N K_h(x - X_j)(X_j - x)^i \quad (21b)$$

$$T_i(x) = \sum_{j=1}^N K_h(x - X_j)(X_j - x)^i Y_j. \quad (21c)$$

Der gesuchte bedingte Erwartungswert $\hat{m}_h(x)$ entspricht dabei $\hat{\beta}_0(x)$. Im bisher betrachteten *lokal konstanten* Fall mit $p = 0$ reduziert sich $\hat{\beta}$ zu $\hat{\beta}_0$ und man kann schreiben:

$$\hat{m}_h(x) = \frac{T_0(x)}{S_0(x)} \quad (22)$$

was mit (18) übereinstimmt.

Betrachtet man den Fall $p = 1$, erhält man aus der üblichen Formel für das Inverse einer 2×2 -Matrix den *lokal linearen* Schätzer

$$\hat{m}_h(x) = \frac{T_0(x)S_2(x) - T_1(x)S_1(x)}{S_0(x)S_2(x) - S_1(x)^2}. \quad (23)$$

Ein großer Vorteil der lokal linearen Regression ist laut [11] die verbesserte Schätzung an den Rändern der Datenwerte. Bei der Verwendung des Nadaraya-Watson-Schätzers ergeben sich durch die einseitige Umgebung an den Rändern große Fehler. Dem soll mit der Schätzung eines Polynoms von höherem Grad entgegen gewirkt werden. Zudem lässt sich die Ableitung des bedingten Erwartungswerts direkt aus

$$\hat{m}_h^{(\nu)}(x) = \nu! \hat{\beta}_\nu(x) \quad (24)$$

bestimmen. So werden Fehler bei der numerischen Berechnung von $\hat{m}_h'(x)$ aus der geschätzten Funktion $\hat{m}_h(x)$ vermieden.

2.3.3 Multivariate Regression

Die bisher behandelten Regressionsmethoden beziehen sich lediglich auf den eindimensionalen Fall mit der abhängigen Variablen $Y = m(X) + \epsilon$. Dieses Problem lässt sich natürlich auf Observablen erweitern, die von mehreren Variablen abhängen. In dem Fall können die Datenpunkte $\mathbf{X}_j = (X_{j1}, \dots, X_{jd})^T$ als d -dimensionaler Vektor aufgefasst werden.

Für die histogrammbasierte Regression wird lediglich die Dimension der Bins entsprechend erweitert. Für $d = 2$ schreibt man beispielsweise:

$$B_{lm} = [x_{01} + (l-1)h_1; x_{01} + lh_1] \times [x_{02} + (l-1)h_2; x_{02} + lh_2], \quad l, m \in \mathbb{Z} \quad (25)$$

wobei x_{0i} und h_i der Startpunkt bzw. die Binbreite in der jeweiligen Dimensionsrichtung sind.

Bei der Kernel-Regression bietet es sich an, als Kernfunktion eine d -dimensionale Funktion zu wählen. Für zwei Dimensionen wählt man beispielsweise eine radial-symmetrische Kernfunktion. Dabei gilt:

$$\tilde{K}_{\mathbf{h}}(\mathbf{u}) = \frac{1}{h_1 h_2} \cdot \tilde{K}\left(\frac{u_1}{h_1}, \frac{u_2}{h_2}\right) \quad (26a)$$

$$\tilde{K}(\mathbf{u}) \propto K(\|\mathbf{u}\|). \quad (26b)$$

Dies bedeutet, die Kernfunktion kann als eindimensionale Funktion der Norm des Vektors $\mathbf{u} = (u_1, u_2)^T$ aufgefasst werden, und zwar für alle Punkte, die innerhalb eines Radius liegen, der hauptsächlich durch die Bandbreiten $\mathbf{h} = (h_1, h_2)^T$ bestimmt ist. Für die Wahl der Bandbreiten lässt sich (13) auf den d -dimensionalen Fall erweitern zu

$$\tilde{h}_j = \left(\frac{4}{d+2}\right)^{1/(d+4)} N^{-1/(d+4)} \sigma_j. \quad (27)$$

Damit folgt schließlich analog zu (18) für den bedingten Erwartungswert:

$$\hat{m}_{\mathbf{h}}(\mathbf{x}) = \frac{\sum_{j=1}^N \tilde{K}_{\mathbf{h}}(\mathbf{X}_j - \mathbf{x}) Y_j}{\sum_{j=1}^N \tilde{K}_{\mathbf{h}}(\mathbf{X}_j - \mathbf{x})} \quad (28)$$

bzw. das Minimierungsproblem für die lokal lineare Regression

$$\sum_{j=1}^N [Y_j - \beta_0 - \beta_1^T (\mathbf{X}_j - \mathbf{x})]^2 \tilde{K}_{\mathbf{h}}(\mathbf{x} - \mathbf{X}_j). \quad (29)$$

Letzteres lässt sich wieder analog zu (21) in Matrizen ausdrücken. Für den 2-dimensionalen Fall ergibt sich somit:

$$\hat{\beta} = \begin{pmatrix} S_{00} & S_{10} & S_{01} \\ S_{10} & S_{20} & S_{11} \\ S_{01} & S_{11} & S_{02} \end{pmatrix}^{-1} \begin{pmatrix} T_{00} \\ T_{10} \\ T_{01} \end{pmatrix} \quad (30a)$$

mit

$$S_{pq}(\mathbf{x}) = \sum_{j=1}^N \tilde{K}_{\mathbf{h}}(\mathbf{x} - \mathbf{X}_j) (X_{j1} - x_1)^p (X_{j2} - x_2)^q \quad (30b)$$

$$T_{pq}(\mathbf{x}) = \sum_{j=1}^N \tilde{K}_{\mathbf{h}}(\mathbf{x} - \mathbf{X}_j) (X_{j1} - x_1)^p (X_{j2} - x_2)^q Y_j. \quad (30c)$$

Durch lösen der Gleichung erhält man aus $\hat{\beta}_0(\mathbf{x})$ schließlich wieder den bedingten Erwartungswert $\hat{m}_h(\mathbf{x})$, während die Einträge von $\hat{\beta}_1$ die jeweiligen Ableitungen in entsprechender Raumrichtung liefern.

2.4 Elektroenzephalografie (EEG)

In dieser Arbeit werden die gewonnenen Kenntnisse aus der Analyse von synthetisch erzeugten Datensätzen beispielhaft auf EEG Daten angewandt. Bei der Elektroenzephalografie werden elektrische Felder, erzeugt durch neuronale Aktivität, über Silber- oder Silberchloridelektroden, die auf dem Kopf aufgesetzt werden, gemessen. Die erhaltenen Daten werden vor allem zu klinischen und wissenschaftlichen Zwecken verwendet.

Im Folgenden soll der Prozess genauer erläutert werden. Für detaillierte Informationen zur Elektroenzephalografie sei auf [12, 13, 14] hingewiesen.

2.4.1 Neurophysiologischer Hintergrund

Entlang der Zellwände von Neuronen herrscht ständig ein elektrisches Potenzial; ein Ruhepotenzial von etwa 75 Millivolt [12]. Dieses wird hervorgerufen durch die ungleiche Ionenverteilung innerhalb und außerhalb der Zelle. Bei neuronaler Aktivität werden spannungsempfindliche Kanäle in der Zellmembran geöffnet, sodass Ionen passieren können. Dadurch kommt es zu einer Depolarisation des Membranpotenzials in positiver Richtung und löst beim Überschreiten eines Schwellwerts ein sogenanntes *Aktionspotenzial* aus, welches etwa eine Millisekunde andauert. Dieses entsteht nahe des Zellkerns und wird entlang des *Axons*, ein wenige Mikrometer dicker Nervenzellfortsatz, weitergeleitet.

Erreicht ein Aktionspotenzial eine Synapse, löst dies die Freisetzung von chemischen Neurotransmittern aus. Auf der anderen Seite der Synapse befinden sich *Dendriten*, die mit Rezeptoren für die Neurotransmitter ausgestattet sind. Ähnlich wie das Axon sind die Dendriten mit dem Zellkern des Neurons verbunden. Potenzialänderungen, ausgelöst durch die Neurotransmitter, werden so an den Zellkern des *post-synaptischen Neurons* weitergeleitet, wo ein neues Aktionspotenzial ausgelöst werden kann. Die post-synaptischen Potenziale (PSPs) an den Dendriten können bis zu mehrere hundert Millisekunden andauern und sowohl in positiver als auch negativer Richtung verschoben sein (auch exzitatorisches bzw. inhibitorisches Potenzial).

Bei der EEG Messung wird ausgenutzt, dass PSPs einen elektrischen Dipol erzeugen. Fließen zum Beispiel positiv geladene Natrium Ionen durch die Membran in die Zelle hinein, bleibt außerhalb im Bereich der Dendriten eine (relativ zum anderen Ende der Zelle) negative Ladung zurück. Diese Dipolausbildung ist vor allem bei sogenannten *Pyramidenzellen* zu beobachten, da ihre Dendriten weit vom Zellkern entfernt lokalisiert sind. Natürlich ist das elektrische Feld des Dipols, erzeugt

durch eine einzelne Zelle, viel zu schwach, um außerhalb der Kopfhaut gemessen zu werden. Vielmehr sorgen große Populationen von tausenden bis hundert tausend synchron aktiven Neuronen für messbare Signale. Dabei summieren sich gleich ausgerichtete Feldlinien, während sich entgegengesetzte Dipole auslöschen. Es ist zu beachten, dass die gemessene Amplitude dann maximal ist, wenn der erzeugte Dipol senkrecht zur Kopfoberfläche ausgerichtet ist, und sich bei paralleler Ausrichtung zu null reduziert. Die für die EEG Signale hauptverantwortlichen Pyramidenzellen sind trotz der gewundenen Struktur des Gehirns im Kortex überwiegend senkrecht zur Kopfoberfläche ausgerichtet [12].

2.4.2 Aufnahme von EEG Daten

Um die elektrischen Dipolfelder, die von aktiven Neuronenpopulationen erzeugt werden, zu messen, wird dem Patienten eine netzartige Haube mit mehreren Elektroden aufgesetzt. Um einen elektrischen Kontakt zwischen Kopfhaut und Elektrode zu ermöglichen wird zusätzlich ein leitendes Elektrolyt-Gel verwendet. Die Signale, die bei den Elektroden (meist aus Silber mit Silber-Chlorid-Mantel [13]) ankommen, befinden sich im Bereich von einigen μV . Daher werden sie zunächst durch einen Verstärker um einen Faktor zwischen 10^3 und 10^5 verstärkt [13]. Dabei kommt ein Differenzverstärker zum Einsatz, um das starke Rauschen (verursacht zum Beispiel durch den elektrischen Schaltkreis selbst oder externe Störsignale von anderen elektronischen Geräten in der Umgebung) zu reduzieren. Das eigentliche EEG Signal und ein Referenzsignal werden beide gegen eine gemeinsame Masseelektrode gemessen. Durch die Verstärkung der Differenz dieser beiden Potenziale kann somit das Rauschen (Referenzsignal) größtenteils eliminiert werden. Anschließend werden die EEG Signale über einen Analog-Digital Wandler für die Weiterverarbeitung am Computer in binäre Signale umgewandelt. Hier können beispielsweise biologische Störsignale, wie spontane Änderungen der Hautleitfähigkeit oder elektrische Impulse durch Muskelkontraktionen und Augenbewegungen, durch Filter reduziert werden.

Gewöhnliche EEG Signale weisen eine oszillatorische Charakteristik auf. Diese lassen sich abhängig von der Frequenz grob in bestimmte Bewusstseinszustände des Gehirns einteilen. So ist zum Beispiel das β -Band (13-30 Hz) typisch für leichte kognitive Aktivitäten, während das niederfrequente δ -Band (< 4 Hz) meist nur in Tiefschlafphasen anzutreffen ist [13]. Invarianzen dieser Oszillationen weisen meist auf pathologische Störungen hin, so auch Epilepsie. Es wird nach Möglichkeiten geforscht, diese Invarianzen bereits vor Eintreten eines epileptischen Anfalls innerhalb der Dynamik des Signals festzustellen.

3 Analyse synthetischer Daten

Um die Qualität der vorgestellten Regressionsmethoden zu überprüfen und zu vergleichen, werden diese zunächst beispielhaft auf synthetisch erzeugte Daten angewandt. Dabei wird untersucht, wie sich die Regression mit dem Histogrammschätzer und die lokal konstante sowie lokal lineare Kernel-Regression bei unterschiedlicher Datenmenge verhalten. Außerdem soll überprüft werden, ob die direkte Ableitung der geschätzten Funktion aus der lokal linearen Regression bessere Ergebnisse liefert, als die numerische Ableitung. Dies wird sowohl für den eindimensionalen als auch zweidimensionalen Fall durchgeführt.

Anschließend wird die Regression eines Langevin-Prozess mit zeitabhängiger Driftfunktion $D^{(1)}(x, t)$ untersucht. Dazu wird die Datenreihe zunächst in mehrere Zeitfenster mit bestimmter Datenpunktzahl eingeteilt, in denen der Prozess dann als stationär angesehen wird. Anschließend werden die Regressionen jeweils auf die Daten innerhalb der Zeitfenster angewandt. Zusätzlich wird die Zeit als zweite Dimension aufgefasst und somit auch die zu analysierenden Zeitfenster mit einer Kernelfunktion gewichtet. Es soll überprüft werden, wie die Breite der Zeitfenster im Zusammenhang mit der Stärke der zeitlichen Änderung Einfluss auf die Regression nimmt.

3.1 Erzeugung synthetischer Daten

Nach dem Euler-Verfahren erhält man für die Langevin-Gleichung

$$\dot{x}(t) = D^{(1)}(x, t) + \sqrt{2D^{(2)}(x, t)}\epsilon(t) \quad (31)$$

folgende Lösung [8]:

$$x(t + \Delta t) = x(t) + D^{(1)}(x, t) \cdot \Delta t + \sqrt{2D^{(2)}(x, t)}\epsilon(t) \cdot \sqrt{\Delta t}. \quad (32)$$

wobei $\sqrt{\Delta t}$ auf der stochastischen Integration nach Itô beruht.

Diese wurde benutzt, um einen Datensatz zu erzeugen, der einen Langevin-Prozess simuliert. Dabei wurden N Datenpunkte $X_j = X(t_j)$ mit $j = 1, \dots, N$ erzeugt, wobei $t_j = j \cdot \Delta t$ über den Zeitschritt $\Delta t = 0.01$ mit einer "realen" Zeitskala verknüpft ist. Für $D^{(2)}(x, t) = c$ wird dabei stets eine konstante Funktion $c = 1$ angenommen.

3.2 Vergleich der 1-dimensionalen Regressionsmethoden

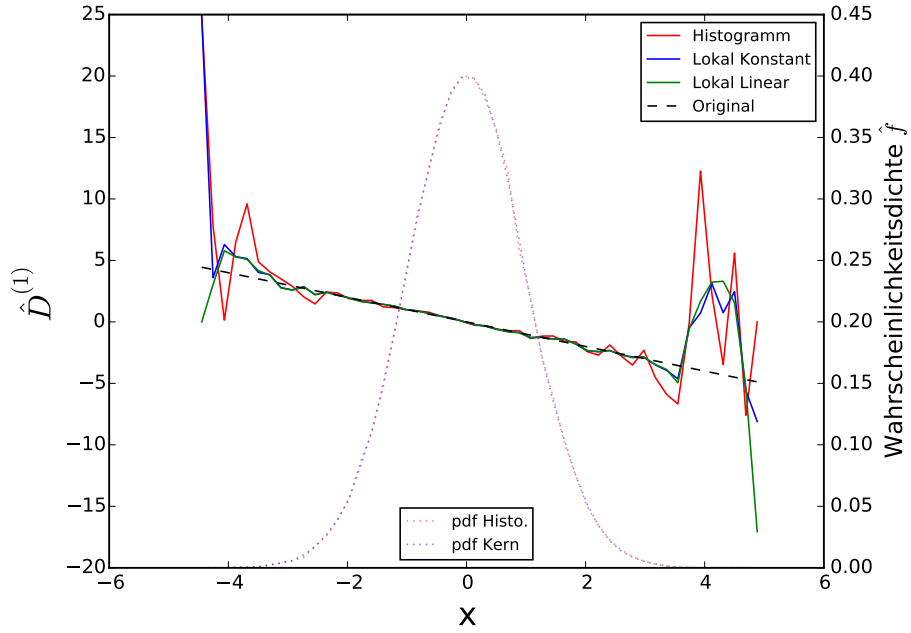
Zum Vergleich der Regression mit dem Histogrammschätzer und der lokal konstanten bzw. linearen Kernel-Regression wurde folgende stationäre Driftfunktion zur

Simulation des Langevinprozesses verwendet:

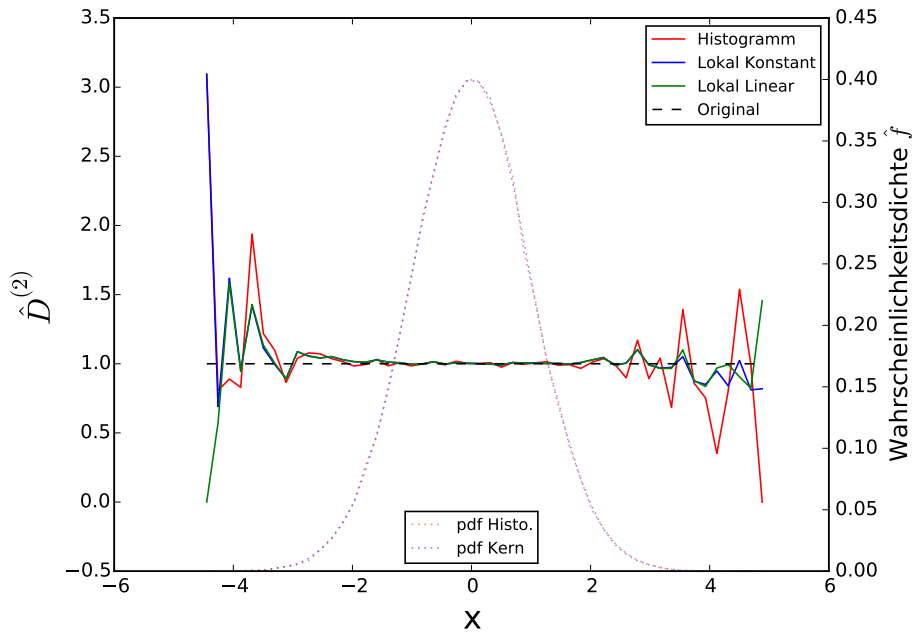
$$D^{(1)}(x) = \alpha \cdot x$$

mit $\alpha = -1$.

Es wurden 20 Datensätze mit jeweils $N = 10^6$ Datenpunkten erzeugt. Die Regressionsverfahren wurden dann jeweils auf eine unterschiedliche Datenmenge n aus diesen Sätzen angewandt. Zur Regression wurden jeweils $n_s = 50$ Samplepunkte verwendet. Abbildung 2 zeigt das Ergebnis der Regression für einen der Datensätze bei $n = 10^6$ Datenpunkten. Man sieht, dass die Schätzung am Rand deutlich schwankt. Zur Verdeutlichung ist in Abbildung 3 nochmal die Differenz zwischen der geschätzten und originalen Funktion dargestellt. Die Schwankungen sind auch zu erwarten, da die Datendichte an den Rändern immer weiter abnimmt, somit weniger Information für eine gelungene Regression zur Verfügung steht. Bereits hier ist zu erkennen, dass der Histogrammschätzer stärker abweicht, als die relativ gleichförmigen Kernschätzer.



(a) Driftfunktion $D^{(1)}(x) = -x$



(b) Diffusionsfunktion $D^{(2)} = 1$

Abbildung 2: Geschätzte Drift- (a) und Diffusionsfunktion (b) eines stationären Langevin-Prozesses aus der Regression mit Histogrammschätzer und lokal konstantem bzw. linearen Kernschätzer. Ebenfalls abgebildet sind die jeweils geschätzten pdf.

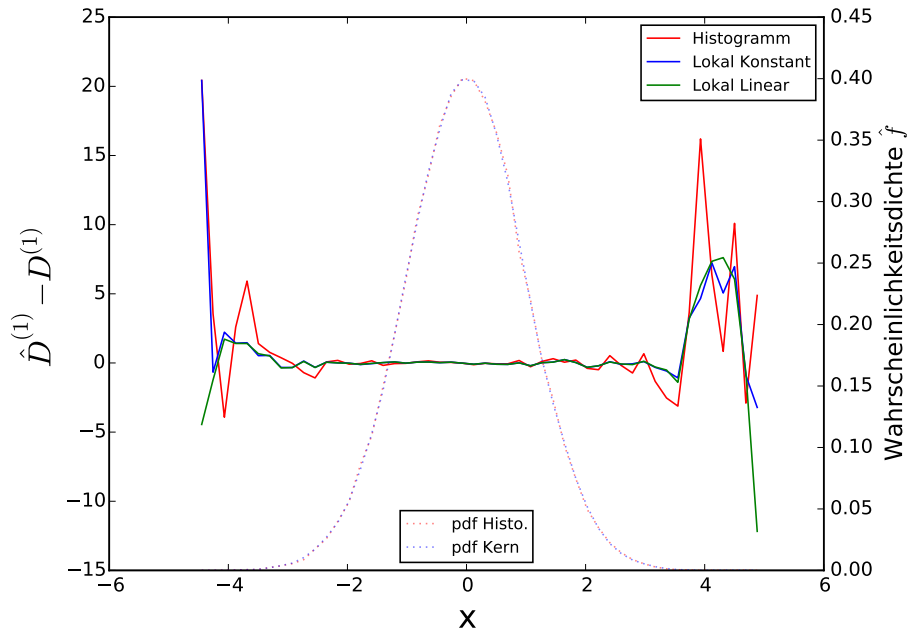


Abbildung 3: Differenz der geschätzten und originalen Driftfunktion.

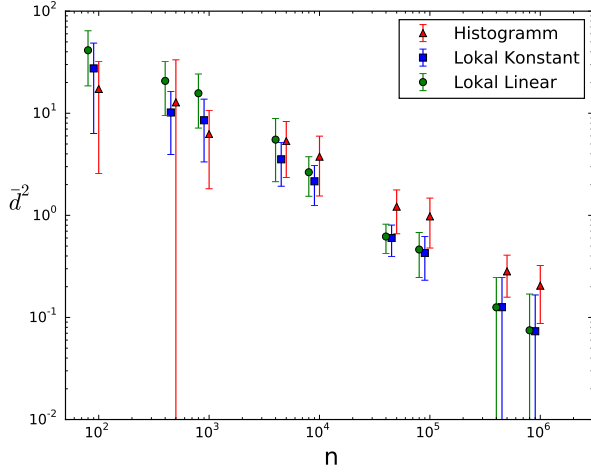
Um eine bessere Aussage über die Qualität der einzelnen Schätzmethoden zu treffen, wird als Maß die mittlere quadratische Abweichung d^2 der geschätzten Funktion $\hat{D}^{(1)}(x)$ von der originalen Funktion $D^{(1)}(x) = -x$ ermittelt:

$$d^2 = \frac{1}{n_s} \sum_{s=1}^{n_s} [D^{(1)}(x_s) - \hat{D}^{(1)}(x_s)]^2 \quad (33)$$

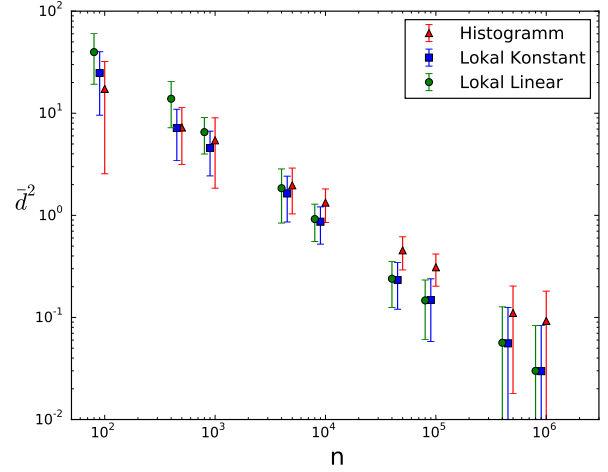
wobei x_s die Samplestellen sind. Für größere statistische Genauigkeit wird dies anschließend über alle Datensätze gemittelt. Dieses Verfahren wurde jeweils für die gewählten Datenmengen n durchgeführt. Somit kann die Qualität der Regressionsmethoden in Abhängigkeit der Datenmengen verglichen werden. Abbildung 4 zeigt die über die Datensätze gemittelte mittlere quadratische Abweichung $\bar{d}^2$ der Driftfunktion. Dabei wurden für die Bestimmung von $\bar{d}^2$ Samplepunkte ausgelassen, die in einem Bereich lagen, wo die Wahrscheinlichkeitsdichte $\hat{f}(x)$ einen bestimmten Prozentsatz p des Maximums $\hat{f}(x_{max})$ nicht überschreitet. Es wird davon ausgegangen, dass jene Bereiche eine zu geringe Informationsmenge bieten und daher keine zuverlässigen Ergebnisse liefern.

Wie erwartet nimmt die Qualität bei einer größeren Datenmenge als Grundlage zu. Dabei konvergieren beide kernelbasierten Regressionsmethoden gleichermaßen, während der Histogrammschätzer etwas schlechter abschneidet. Im Bereich kleinerer Datenmengen laufen alle Regressionsmethoden zusammen und unterliegen deutlich größeren Schwankungen. Es lässt sich also keine eindeutige Aussage darüber treffen, welche Regression hier am besten zu wählen ist.

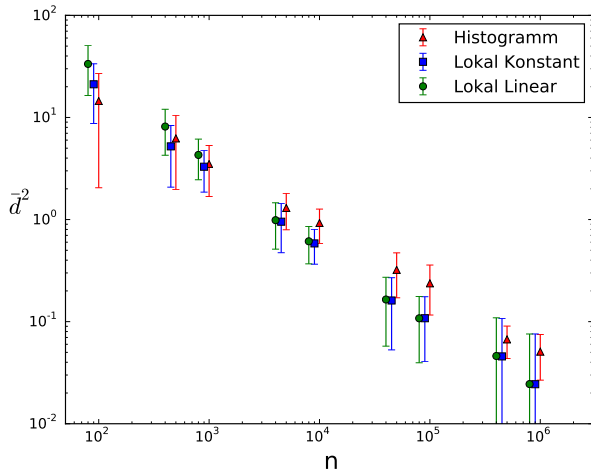
Es zeigt sich dennoch, dass die Benutzung eines gewichteten Schätzers allgemein



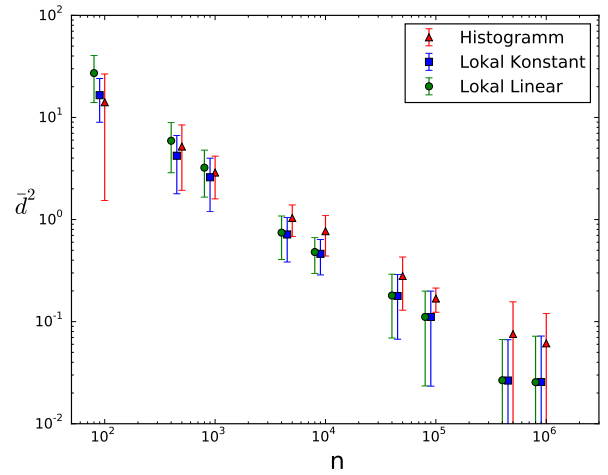
(a) $p = 0.01$



(b) $p = 0.05$



(c) $p = 0.10$



(d) $p = 0.15$

Abbildung 4: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der geschätzten Driftfunktion vom analytischen Wert, gemittelt über 20 Datensätze. Die Fehlerbalken entsprechen dabei der Standardabweichung vom Mittelwert. Die Regressionen wurden für verschiedene Datenmengen n durchgeführt. Für bessere Übersichtlichkeit sind die Punkte der Kernschätzer mit einem Offset horizontal verschoben. Es wurden nur die Samplestellen x_s mit $\hat{f}(x_s) > p \cdot \hat{f}(x_{max})$ berücksichtigt.

bessere Ergebnisse liefert. Auch Lamouroux und Lehnertz konnten in [6] zeigen, dass für die Schätzung eines stochastischen Oszillators die kernelbasierte Regression bei allen verwendeten Datenmengen besser abschneidet als die histogrammbasierte. Zudem waren weniger als die Hälfte der Daten nötig, um mit dem Kernschätzer die gleichen Ergebnisse wie mit dem Histogrammschätzer zu erzielen. Eine Verbesserung durch die Benutzung der lokal linearen Regression gegenüber der konstanten lässt sich hier für den 1-dimensionalen Fall nicht feststellen.

Durch das Abschneiden der ermittelten Werte bei geringen Wahrscheinlichkeitsdichten lässt sich zudem der Fehler allgemein reduzieren. Dies scheint jedoch keinen direkten Einfluss auf die einzelnen Schätzmethoden zu haben.

Weiterhin wurde nach dem gleichen Verfahren die Güte der Ableitung aus der lokal linearen Kernel-Regression (24) untersucht. Wie zuvor wurde die mittlere quadratische Abweichung $\bar{d}^2$ der Ableitung vom analytischen Wert ermittelt. Hierbei wurden die Ableitungen für den Histogrammschätzer und die lokal konstante Kernelregression numerisch mithilfe des zentralen Differenzenquotienten bestimmt. Die für die lokal lineare Kernelregression ist dagegen direkt aus den vorhandenen Daten geschätzt worden. In Abbildung 5 ist der Vergleich für $p = 0.1$ zu sehen.

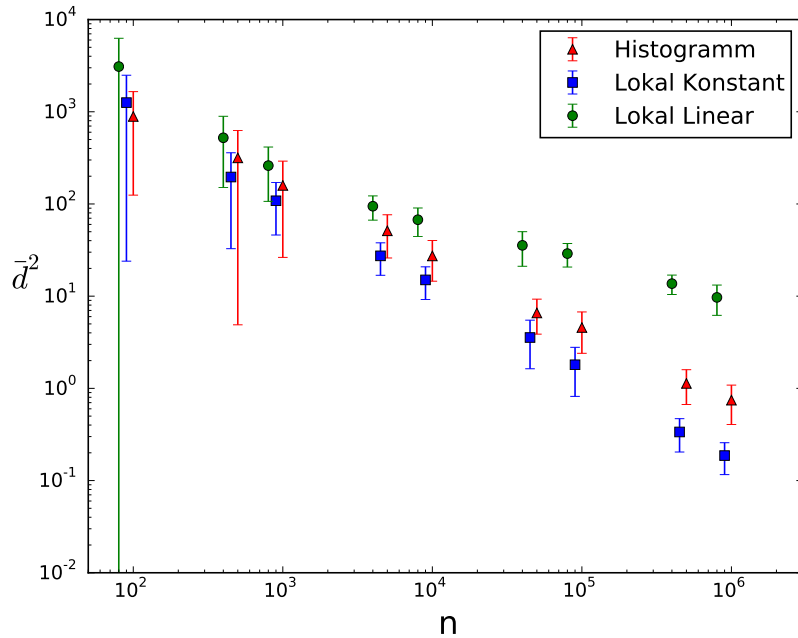


Abbildung 5: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der Ableitung der geschätzten Driftfunktionen für den Histogrammschätzer und der lokal konstanten Kernel-Regression, sowie die direkt geschätzte Ableitung aus der lokal linearen Regression.

Trotz der theoretischen Überlegung, dass es eine qualitative Verbesserung durch die direkte Schätzung der Ableitung gegenüber der fehlerbehafteten numerischen

Berechnung aus der Ausgangsfunktion geben müsse, liefert die Analyse ein deutlich gegensätzliches Ergebnis. Bei größeren Datenmengen ($n \approx 10^6$) ist die mittlere quadratische Abweichung der lokal linear geschätzten Ableitung um 2 Ordnungen größer als die aus der lokal konstanten Kernel-Regression berechneten. Selbst die Methode mit dem Histogrammschätzer liefert ähnlich gute Ergebnisse. Aber auch hier treten im Bereich kleinerer Datenmengen wieder erhebliche Schwankungen auf.

Das Verfahren wurde ebenfalls auf einen Langevin-Prozess mit dem Driftterm $D^{(1)}(x) = x - x^3$, also einer Funktion höheren Grades, durchgeführt. Allerdings liefert die Analyse ähnliche Ergebnisse (siehe Abbildung 6; in Abbildung 7 ist zusätzlich die Schätzung zu sehen). Man sieht aber, dass sich im Bereich größerer Datenmengen die Qualität beim Histogramm- und lokal konstanten Kernschätzer deutlich stärker verschlechtert (eine ganze Größenordnung) im Gegensatz zur direkten Schätzung aus der lokal linearen Regression.

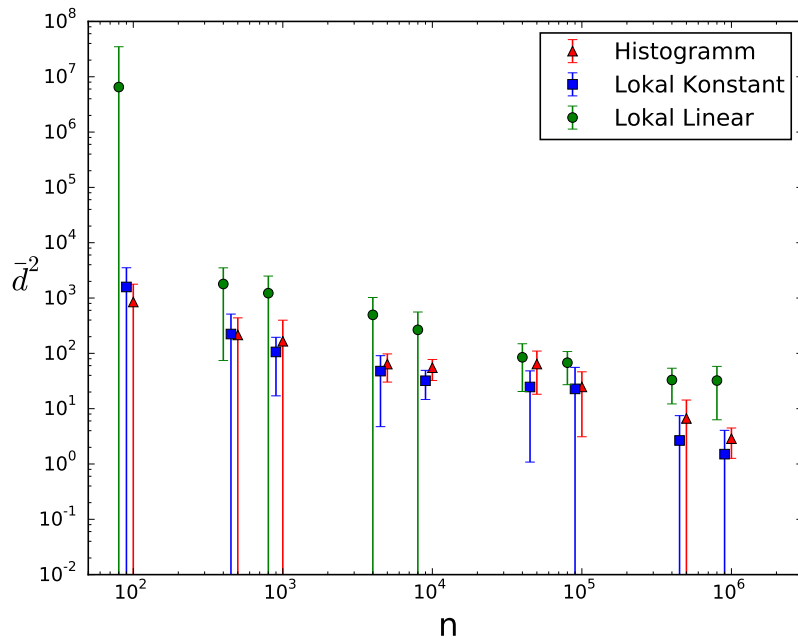
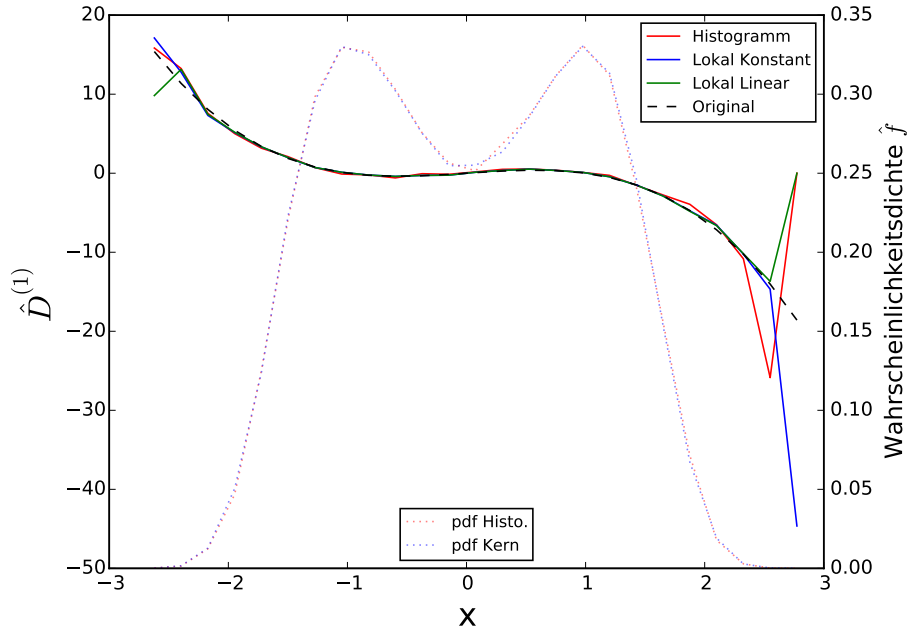
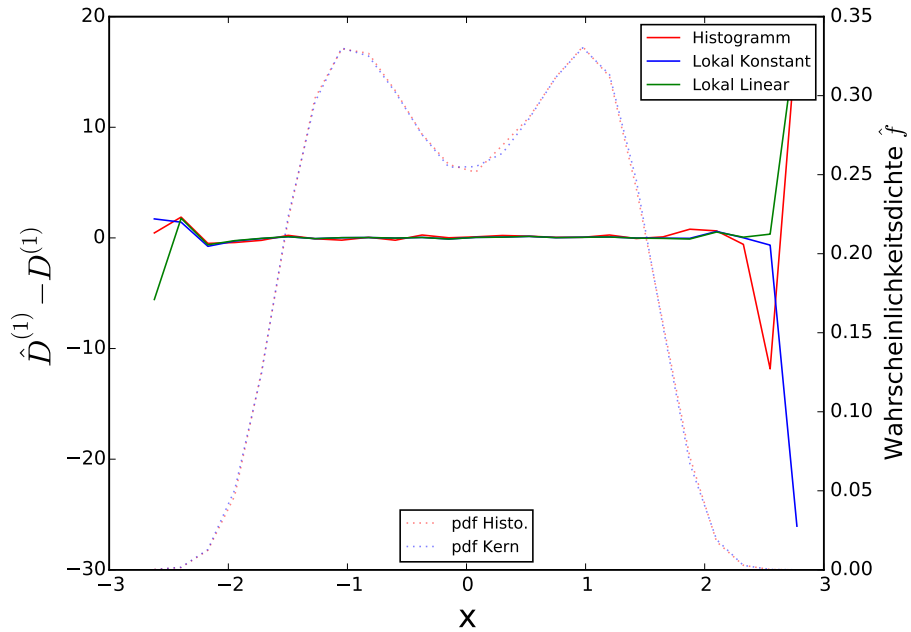


Abbildung 6: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der Ableitung der geschätzten Driftfunktionen für den Histogrammschätzer und der lokal konstanten Kernel-Regression, sowie die direkt geschätzte Ableitung aus der lokal linearen Regression.



(a) Schätzung von $D^{(1)}(x) = x - x^3$



(b) Differenz von Schätzung und Original.

Abbildung 7: Geschätzte Driftfunktion (a) und Differenz zur originalen Funktion dritten Grades (b) eines stationären Langevin-Prozesses aus der Regression mit Histogrammschätzer und lokal konstantem bzw. linearen Kernschätzer. Ebenfalls abgebildet sind die jeweils geschätzten pdf.

3.3 Vergleich der 2-dimensionalen Regressionsmethoden

Die Regression einer Zeitreihenmessung soll von einer 1-dimensionalen Analyse für verschiedene Zeitfenster in 2 Dimensionen überführt werden. Somit stellt die Zeit eine zusätzliche Dimension dar, welche in der Regression direkt mitberücksichtigt wird und im Fall der Kernel-Regression als gewichtete Zeitfenster auftreten (siehe Abschnitt 3.4).

Zunächst sollen die verschiedenen Regressionsmethoden für den 2-dimensionalen Fall wieder miteinander verglichen werden. Der Anschaulichkeit halber werden Datensätze erzeugt, denen folgende Regressionsfunktion zugrunde liegt:

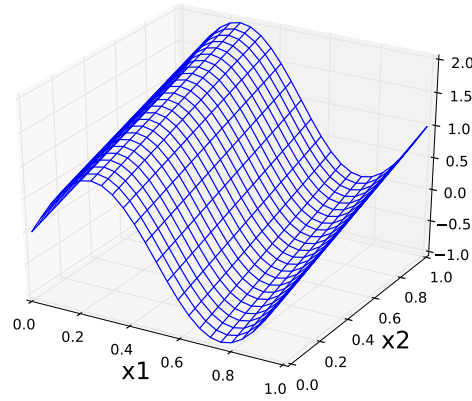
$$m(x_1, x_2) = \sin(2\pi x_1) + x_2.$$

Dabei sind die Datenpunkte mit einem Gaußschen Rauschen ϵ behaftet. Für $Y(X_1, X_2)$ folgt damit:

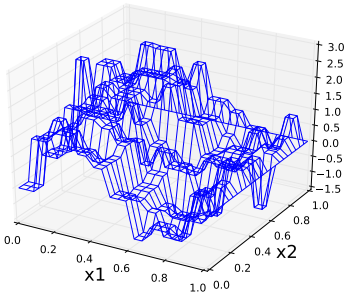
$$Y(X_1, X_2) = m(X_1, X_2) + \epsilon$$

Die Punkte $\mathbf{X}_j$ sind als 25×25 -Anordnung äquidistant in einer $[0, 1] \times [0, 1]$ Ebene verteilt. Es wurden wieder 20 verschiedene Datensätze generiert und mit unterschiedlicher Datenmenge n analysiert. Für die Regression mit dem Histogrammschätzer wurde die Binbreite $h_H = 0.1$ gewählt, für beide Kernel-Regressionen $h_K = 0.3$.

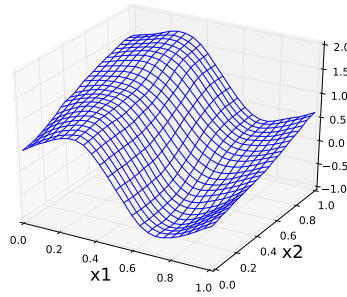
Die Ergebnisse der Regressionen bei $n = 625$ sind in Abbildung 8 gegenübergestellt. Man sieht, dass die lokal lineare Kernel-Regression an den Rändern wesentlich besser die tatsächliche Form der Regressionsfunktion wiedergibt. Dies hängt damit zusammen, dass im Gegensatz zu den anderen Regressionsmethoden bei der lokal linearen trotz geringerer Datenmenge am Rand immerhin eine Funktion ersten Grades, statt eines konstanten Werts, angenähert wird. Somit reagiert die lokal lineare Kernel-Regression besser auf Änderungen (vgl. hierzu auch [11]).



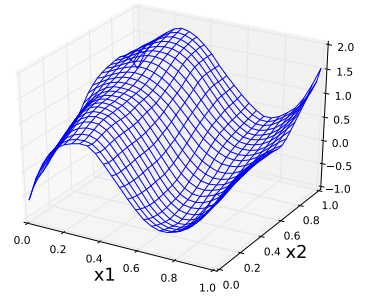
(a) Original: $m(x_1, x_2) = \sin(2\pi x_1) + x_2$



(b) Histogramm



(c) Lokal Konstant



(d) Lokal Linear

Abbildung 8: Vergleich der geschätzten 2-dimensionalen Regressionsfunktionen durch den Histogrammschätzer und lokal konstanten sowie linearen Kernschätzer.

Wie auch bei der 1-dimensionalen Analyse (Abschnitt 3.2) wurde jeweils für alle drei Regressionsmethoden die mittlere quadratische Abweichung von der originalen Funktion berechnet und über alle 20 Datensätze gemittelt. Dies wurde sowohl für die Regressionsfunktion (Abbildung 9) als auch deren Ableitung in x_1 -Richtung (Abbildung 10) durchgeführt.

Es wird deutlich, dass die Kernel-Regressionen insgesamt besser abschneiden als die histogrammbasierte. Im Bereich kleinerer Datenmengen ($n \approx 100$) liefert die lokal lineare Kernel-Regression die besten Ergebnisse, sowohl für die Regressionsfunktion, als auch für deren Ableitung. Allerdings unterscheidet sich bei letzteren die direkte Schätzung der Ableitung nur kaum von der aus der geschätzten Regressionsfunktion numerisch berechneten. Der Unterschied wird womöglich deutlicher, wenn ein polynomialer Schätzer höheren Grades verwendet wird. Trotzdem erweist sich diese Regressionsmethode für einen 2-dimensionalen Fall als besser geeignet.

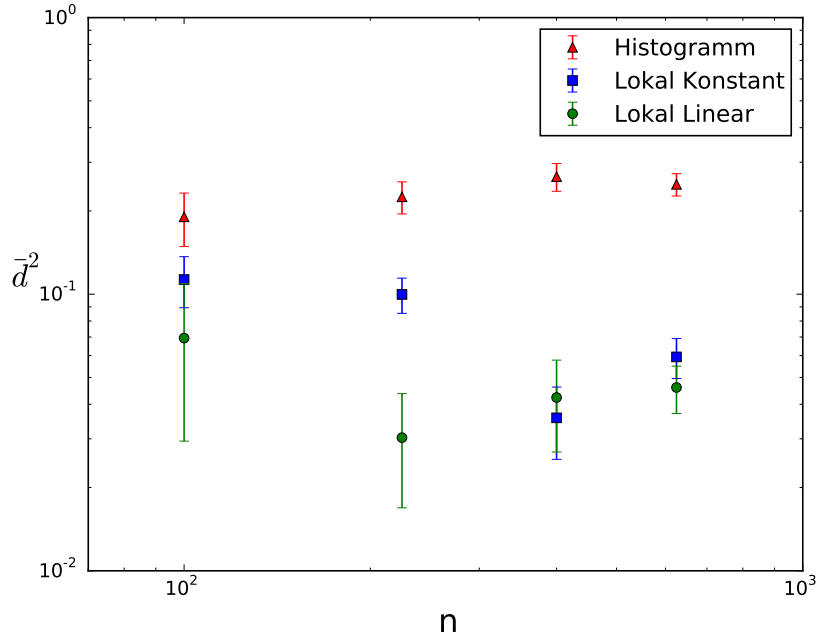


Abbildung 9: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der geschätzten 2-dimensionalen Funktion vom analytischen Wert, gemittelt über 20 Datensätze. Die Fehlerbalken entsprechen dabei der Standardabweichung. Die Regressionen wurden für verschiedene Datenmengen n durchgeführt.

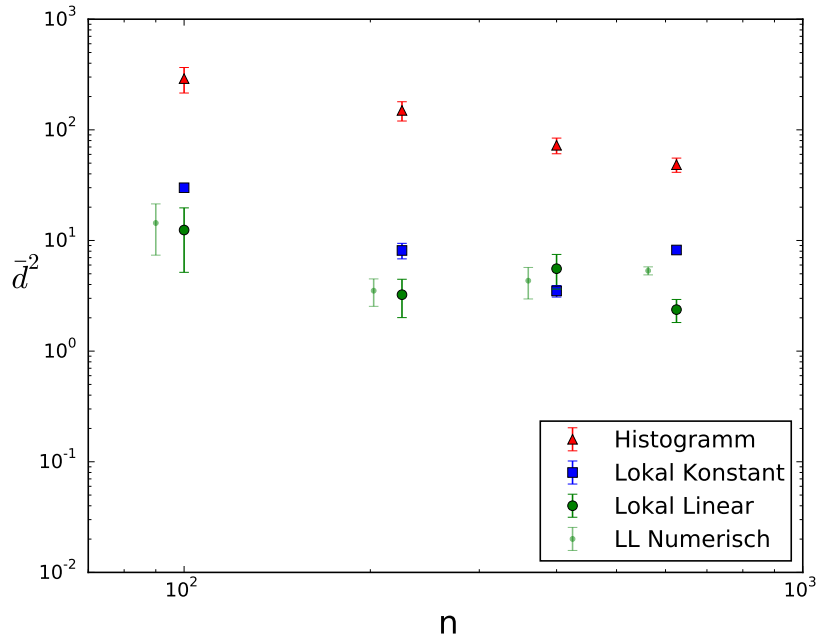


Abbildung 10: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der Ableitung in x_1 -Richtung der geschätzten Funktion für den Histogrammschätzer und der lokal konstanten Kernel-Regression, sowie die direkt geschätzte Ableitung aus der lokal linearen Regression. Zum Vergleich wurde die Ableitung zusätzlich numerisch über den zentralen Differenzenquotienten aus der lokal linear geschätzten Funktion berechnet.

3.4 Nichtstationärer Langevin-Prozess

Als nächstes wurde wieder ein Langevin-Prozess untersucht; diesmal mit einer zeitabhängigen Driftfunktion

$$D^{(1)}(x, t) = \tanh[\alpha \cdot (t - T_0)] \cdot x - x^3$$

wobei T_0 als Mittelpunkt der Zeitreihe festgelegt wurde. Mit dem Parameter α lässt sich sehr einfach die Geschwindigkeit der zeitlichen Veränderung des Drifts beeinflussen. Zur Veranschaulichung ist der zeitabhängige Teil in Abbildung 11 für $\alpha = 0.01$ und $\alpha = 0.001$ dargestellt.

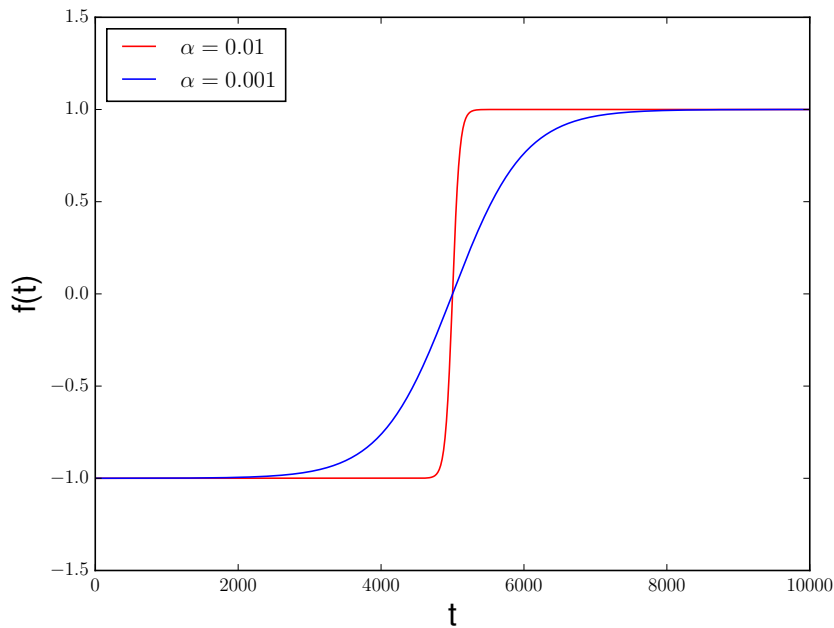


Abbildung 11: $f(t) = \tanh[\alpha \cdot (t - 5000)]$ für $\alpha = 0.01$ und $\alpha = 0.001$

In [6] wurde die Zeitreihenanalyse innerhalb fester Zeitfenster durchgeführt. Dort galt der Prozess jeweils als stationär und es konnte wie im 1-dimensionalen Fall die Regression durchgeführt werden. Hier soll nun die Regression in zwei Dimensionen stattfinden: in räumlicher x - und zeitlicher t -Richtung werden gewichtende Kernfunktionen verwendet. Die zusätzliche Gewichtung in zeitlicher Dimension soll helfen, die nichtstationäre Dynamik des Prozesses besser aufzulösen. Der Vorteil ist, dass es kein festes Zeitfenster gibt, in dem alle Datenpunkte gleichermaßen beitragen, sondern die Daten um die Samplestelle herum, abhängig von der Entfernung, unterschiedlich ausgewertet werden. Das bedeutet, für einen Zeitpunkt relevantere Datenpunkte werden stärker berücksichtigt als diejenigen, die weiter zurück bzw. weiter in der Zukunft liegen.

Es gilt nun die Frage zu klären, wie die Bandbreite h_t in zeitlicher Richtung gewählt werden muss, damit genug Daten für eine zuverlässige Regression erfasst

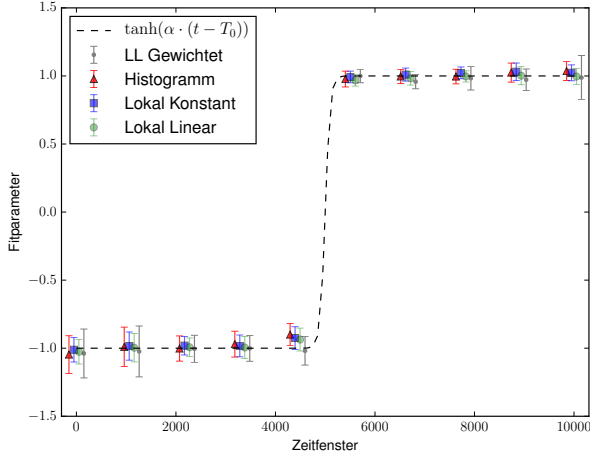
werden, aber auch die zeitliche Auflösung der Dynamik erhalten bleibt. Dazu wurde bei oben genanntem Prozess mit $N = 10^6$ Datenpunkten die verschiedenen Regressionsmethoden angewandt. Dabei wurde jeweils die Zeitfensterbreite variiert ($n = 10^5$ und $n = 10^4$) und die Bandbreite h_t für die gewichteten Zeitfenster entsprechend angepasst. Verglichen werden die einzelnen Schätzungen für einen sich relativ schnell ändernden ($\alpha = 0.01$) und sich relativ langsam ändernden Prozess ($\alpha = 0.001$). Dazu werden die geschätzten Funktionen pro Fenster mit einer Funktion der Form

$$f(x) = c \cdot x - x^3$$

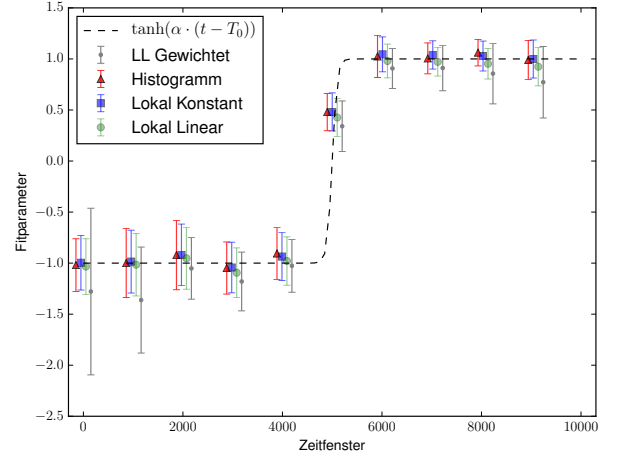
gefittet und dann der Fitparameter c gegen die Zeit aufgetragen. Die Ergebnisse sind in Abbildung 12 zu sehen. Zudem wurde für jedes Zeitfenster wie auch zuvor die mittlere quadratische Abweichung gemittelt über 20 Datensätze berechnet und in Abbildung 13 aufgetragen.

Es fällt auf, dass die Schätzungen bei dem größeren Zeitfenster mit $n = 10^5$ Datenpunkten generell zuverlässigere Ergebnisse liefern. Dies sieht man zum Einen an den geringeren Schwankungen in Abbildung 12 gegenüber den Schätzungen bei $n = 10^4$. Außerdem sinkt die mittlere quadratische Abweichung um etwa eine Größenordnung (vgl. Abbildung 13). Man erkennt hier auch, dass der lokal konstante und lokal lineare Kernschätzer sehr ähnliche Ergebnisse liefern, während die Regression mit dem Histogrammschätzer stets etwas schlechter abschneidet. Die lokal lineare Schätzung mit den gewichteten Zeitfenstern liefert an den Rändern der Zeitreihe, also wo die zeitliche Änderung der Driftfunktion relativ konstant ist, weniger gute Ergebnisse. Allerdings nimmt die Qualität um den Mittelpunkt T_0 , wo die stärkste zeitliche Änderung stattfindet, zu. Scheinbar werden für den konstanten Verlauf durch die Gewichtung zu wenige Daten ausgewertet, die dann zuverlässige Ergebnisse liefern, während dies bei stärkeren zeitlichen Änderungen der Schätzung zugute kommt, da hier nur für den Zeitpunkt relevante Zeitpunkte berücksichtigt werden.

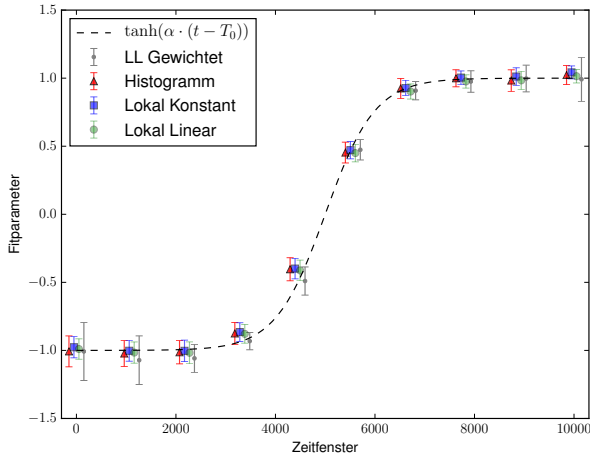
Zwischen dem sich schnell ändernden (oben) und langsam ändernden Prozess (unten) sind kaum Unterschiede festzustellen. Möglicherweise sind die Geschwindigkeiten der zeitlichen Änderung zu ähnlich, um signifikante Merkmale festzustellen.



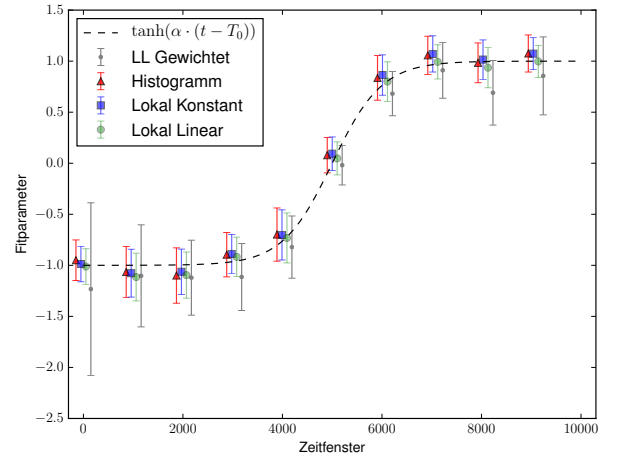
(a) $\alpha = 0.01, n = 10^5$



(b) $\alpha = 0.01, n = 10^4$

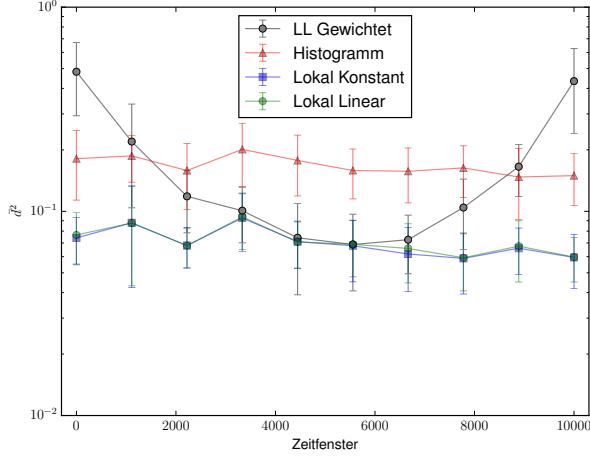


(c) $\alpha = 0.001, n = 10^5$

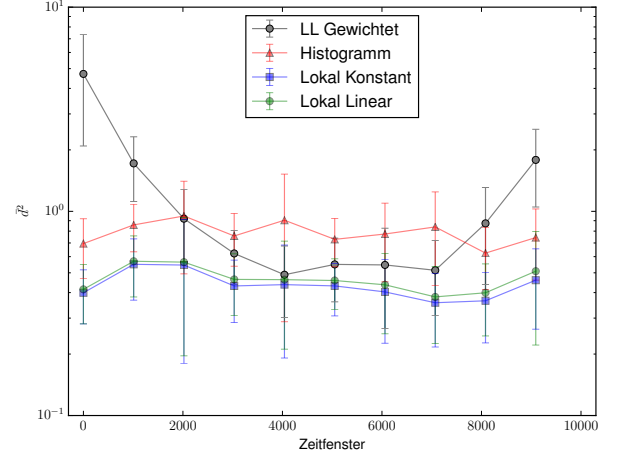


(d) $\alpha = 0.001, n = 10^4$

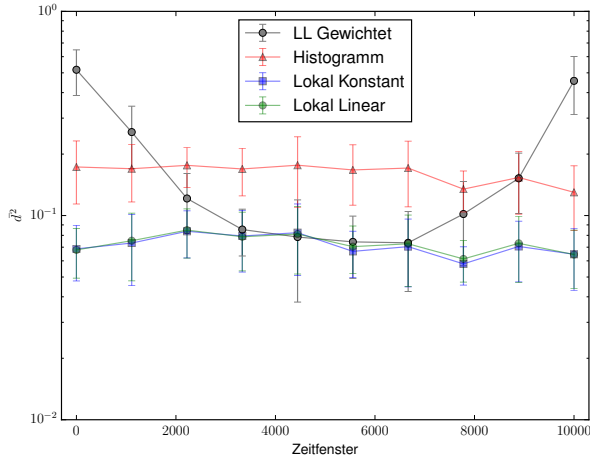
Abbildung 12: Rekonstruktion des zeitabhängigen Teils der Driftfunktion. Aufgetragen sind die Fitparameter c pro Zeitfenster gemittelt über 20 Datensätze. Die Fehlerbalken entsprechen dabei der Standardabweichung. Betrachtet werden zwei verschiedene Langevin-Prozesse mit $D^{(1)}(x, t) = \tanh[\alpha \cdot (t - T_0)] \cdot x - x^3$. Die Regressionen wurden für verschiedene Zeitfenster mit Datenmengen n durchgeführt. Für bessere Übersichtlichkeit sind die Punkte mit einem Offset horizontal verschoben. Es wurden nur die Samplestellen x_s mit $\hat{f}(x_s) > 0.1 \cdot \hat{f}(x_{max})$ berücksichtigt.



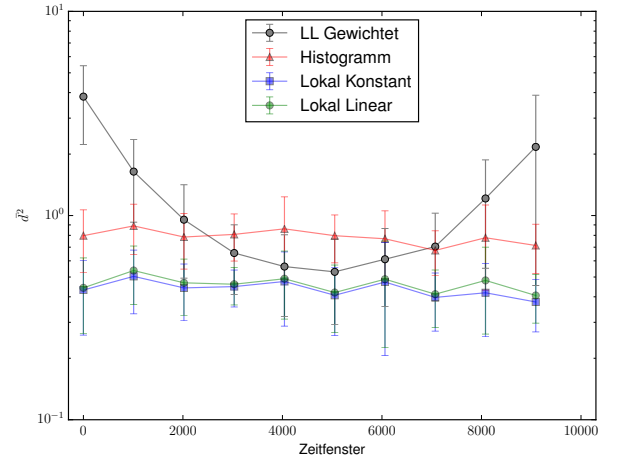
(a) $\alpha = 0.01, n = 10^5$



(b) $\alpha = 0.01, n = 10^4$



(c) $\alpha = 0.001, n = 10^5$



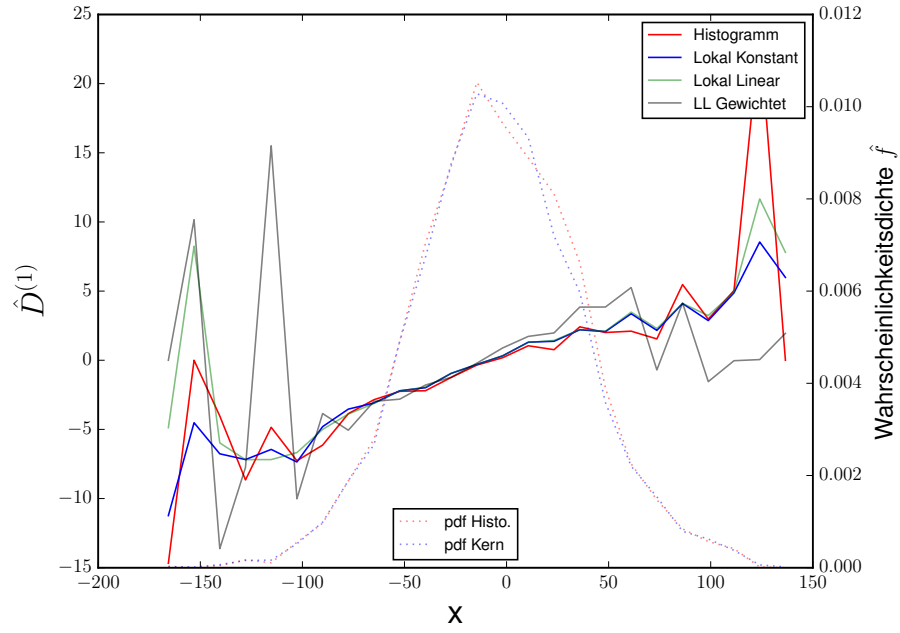
(d) $\alpha = 0.001, n = 10^4$

Abbildung 13: Vergleich der mittleren quadratischen Abweichung $\bar{d}^2$ der geschätzten Driftfunktion vom analytischen Wert, gemittelt über 20 Datensätze. Die Fehlerbalken entsprechen dabei der Standardabweichung. Betrachtet werden zwei verschiedene Langevin-Prozesse mit $D^{(1)}(x, t) = \tanh[\alpha \cdot (t - T_0)] \cdot x - x^3$. Die Regressionen wurden für verschiedene Zeitfenster mit Datenmengen n durchgeführt. Es wurden nur die Samplestellen x_s mit $\hat{f}(x_s) > 0.1 \cdot \hat{f}(x_{max})$ berücksichtigt.

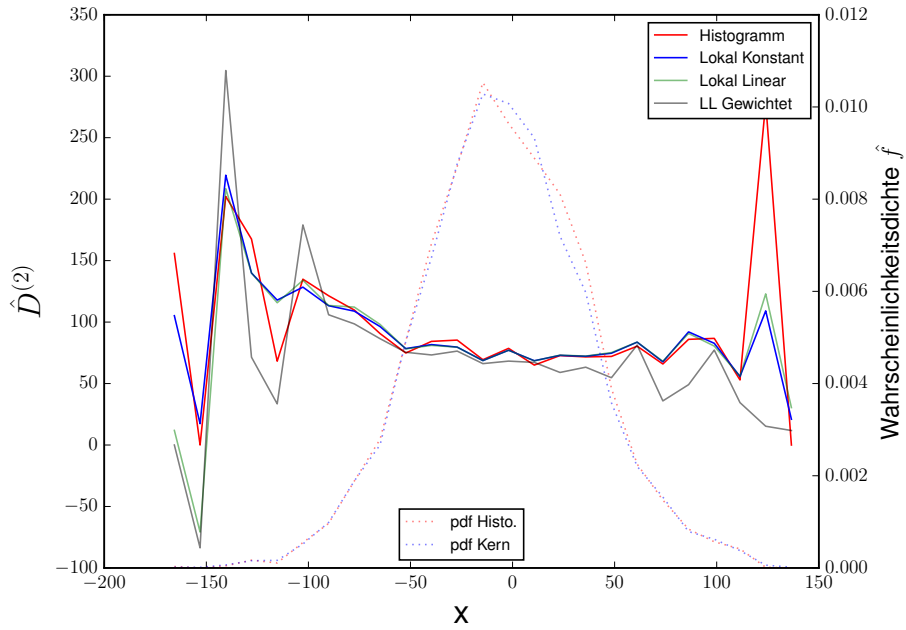
4 Anwendung auf EEG Daten

Die zuvor vorgestellten Regressionsmethoden, das Histogramm und die kernelbasierte Regression mit lokal konstantem und lokal linearem Kernschätzer, wurden probeweise auf Daten einer realistischen Zeitreihe angewandt. Dabei handelt es sich um EEG Daten von 4 verschiedenen Stellen, die etwa 12 Stunden lang mit einer Frequenz von 200 Hz aufgenommen wurden. Sie enthalten zudem 2 Zeitpunkte, wo epileptische Anfälle verzeichnet worden sind.

Die Regressionen wurden bei Zeitfenstern mit $n = 10^4$ Datenpunkten durchgeführt; das entspricht 50 Sekunden. Die Bandbreite h_t für die zeitlich gewichtete Schätzung wurde dabei entsprechend angepasst. In Abbildung 14 sind die Schätzungen des Drifts und der Diffusion für eines der Zeitfenster dargestellt. Für diesen Fall lässt sich vermutlich ein linearer Funktionsverlauf für den Drift bzw. ein konstanter für die Diffusion annähern.



(a) Drift



(b) Diffusion

Abbildung 14: Geschätzte Drift- (a) und Diffusionsfunktion (b) aus EEG Daten innerhalb eines Zeitfensters mit $n = 10^4$ Datenpunkten. Die Regression wurde mit Histogrammschätzer und lokal konstantem bzw. linearen Kernschätzer sowie mit gewichtetem Zeitfenster durchgeführt. Ebenfalls abgebildet sind die jeweils geschätzten pdf.

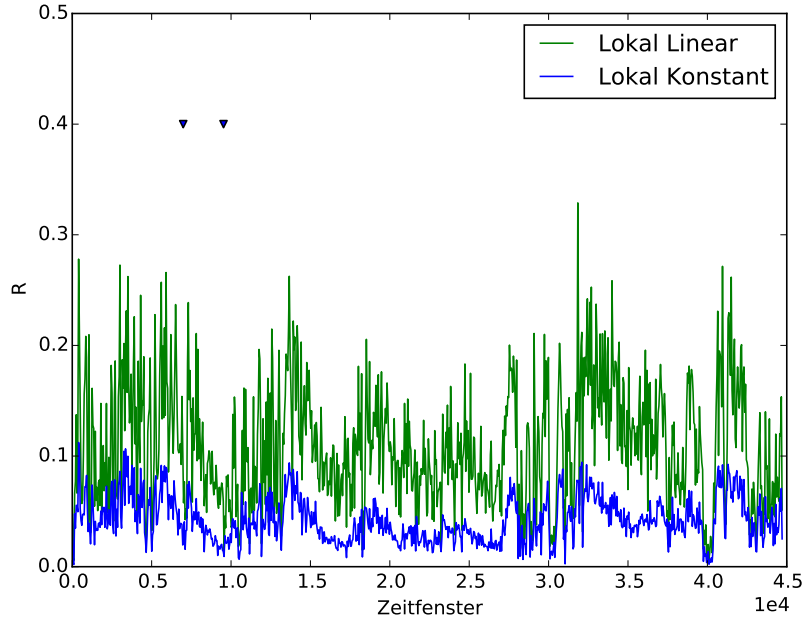
[6, 4] folgend wird die Stärke der Schwankungen $R^{(i)}$ über den zeitlichen Verlauf hinweg bestimmt mit:

$$R^{(i)} = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} dx |\partial_x D^{(i)}(x)|, \quad i = 1, 2 \quad (34)$$

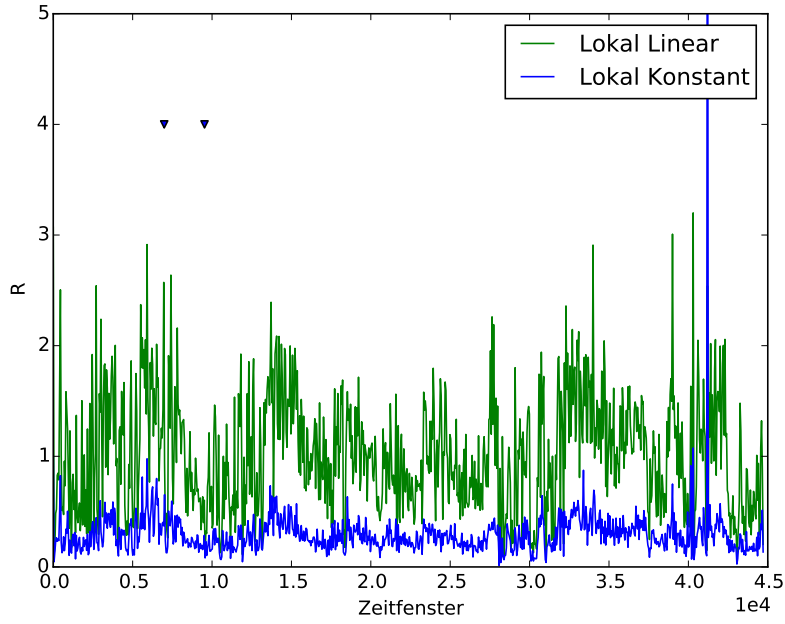
Dabei sind x_1 und x_2 so gewählt, dass der Bereich, in dem die pdf unterhalb eines Prozentsatzes (hier $p = 0.05$) des Maximalwerts liegt, bei der Integration ausgelassen wird.

In Abbildung 15 und 16 ist $R^{(i)}$ für die Kernel-Regressionen ohne Gewichtung der Zeitfenster bzw. mit Gewichtung zu sehen. Zwischen der lokal konstanten und linearen Kernel-Regression ist ein quantitativer Unterschied zu erkennen, der darauf basiert, dass zur Berechnung von $\partial_x D^{(i)}$ die direkt geschätzte Ableitung aus der lokal linearen Regression verwendet wurde. Diese zeigte bereits in Abschnitt 3.2 deutliche Unterschiede zur numerisch berechneten. Ob dies nun Auswirkungen auf den Zusammenhang zwischen den Schwankungen und dem Einsetzen der epileptischen Anfälle hat, lässt sich hier nicht sagen. Generell ist kein eindeutiger Zusammenhang zu erkennen. Zwar sieht man, dass die Schwankungen kurz vor dem ersten Anfall größer werden, aber sogleich sich stark reduzieren bevor der nächste eintritt. Dieser Verlauf ist aber ebenso bei anderen Zeitpunkten zu beobachten, wo Anfälle ausblieben.

Die Gewichtung der Zeitfenster bringt kaum Veränderungen gegenüber dem lokal linearen Kernschätzer ohne Gewichtung mit sich. Grund hierfür kann die Wahl der Dimension sein. Signifikante Ergebnisse zeigen sich möglicherweise erst bei der Analyse der Zusammenhänge zwischen mehreren Zeitreihen, die an verschiedenen Stellen aufgenommen wurden. Die Dynamik der epileptischen Anfälle lassen sich womöglich durch die Änderung der Wechselwirkung einzelner Neuronengruppen charakterisieren.

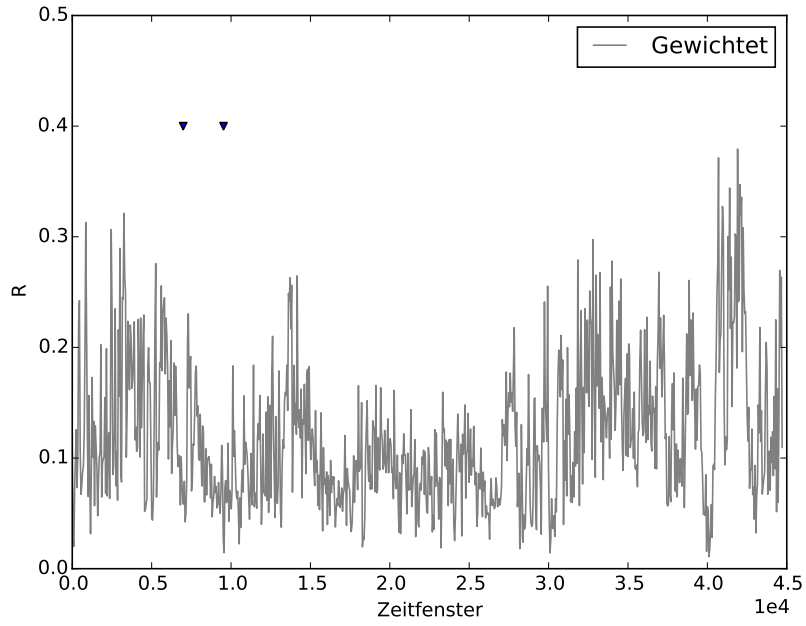


(a) Drift

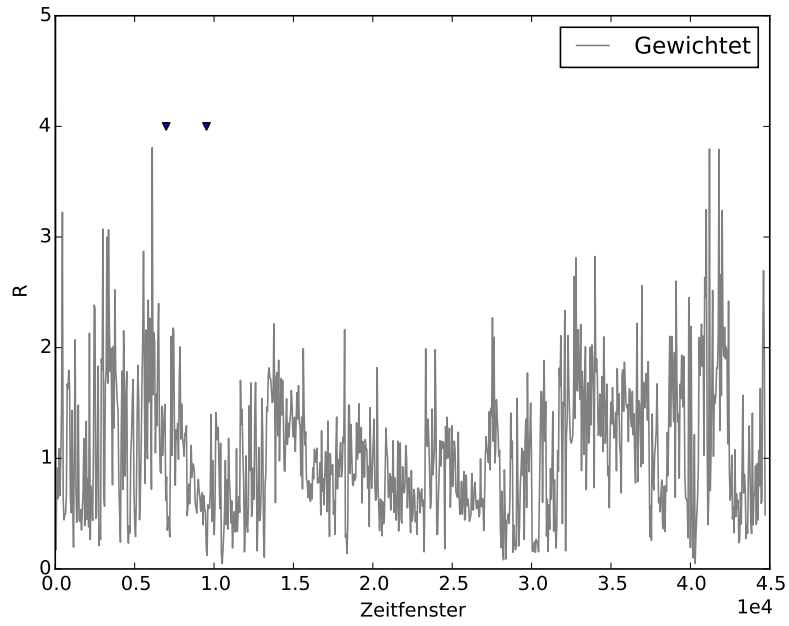


(b) Diffusion

Abbildung 15: Verlauf der Schwankung $R^{(i)}$ der Drift- (a) und Diffusionsfunktion (b) aus EEG Daten. Die Regression wurde mit lokal konstantem bzw. linearen Kernschätzer und ungewichteten Zeitfenstern mit jeweils 10^4 Datenpunkten durchgeführt. Ebenfalls angezeigt sind die Zeitpunkte der verzeichneten Anfälle.



(a) Drift



(b) Diffusion

Abbildung 16: Verlauf der Schwankung $R^{(i)}$ der Drift- (a) und Diffusionsfunktion (b) aus EEG Daten. Die Regression wurde mit dem lokal linearen Kernschätzer und gewichteten Zeitfenstern durchgeführt. Ebenfalls angezeigt sind die Zeitpunkte der verzeichneten Anfälle.

5 Zusammenfassung und Diskussion

Mithilfe der Analyse von synthetischen Datensätzen konnte gezeigt werden, dass die kernelbasierte Regression zuverlässigere Ergebnisse liefert als die histogrammbasierte. Dabei konvergiert ersteres mit Zunahme der Datenmengen schneller. Allerdings ließ sich für geringe Datenmengen ($n \approx 100$) keine genaue Aussage treffen, da die statistischen Schwankungen viel zu groß waren.

Eine Verbesserung durch die lokal lineare Kernel-Regression im Gegensatz zur lokal konstanten konnte für den 1-dimensionalen Fall nicht festgestellt werden. Hier wäre zu untersuchen, ob sich signifikante Verbesserungen bei einer lokal polynomialen Kernschätzung höheren Grades einstellen. Außerdem könnte die Komplexität der zugrunde liegenden Driftfunktion ebenfalls Einfluss haben, da die lokal polynomiale Regression besonders an den Rändern der Datensätze besser auf Änderungen reagiert [11].

Desweiteren wurde untersucht, ob die direkte Schätzung der Ableitung mittels lokal linearer Schätzung Vorteile aufweist. Hier zeigt sich allerdings entgegen der Erwartungen eine Verschlechterung. Möglicherweise sind auch hier Polynome höherer Ordnung notwendig, um Verbesserungen zu zeigen.

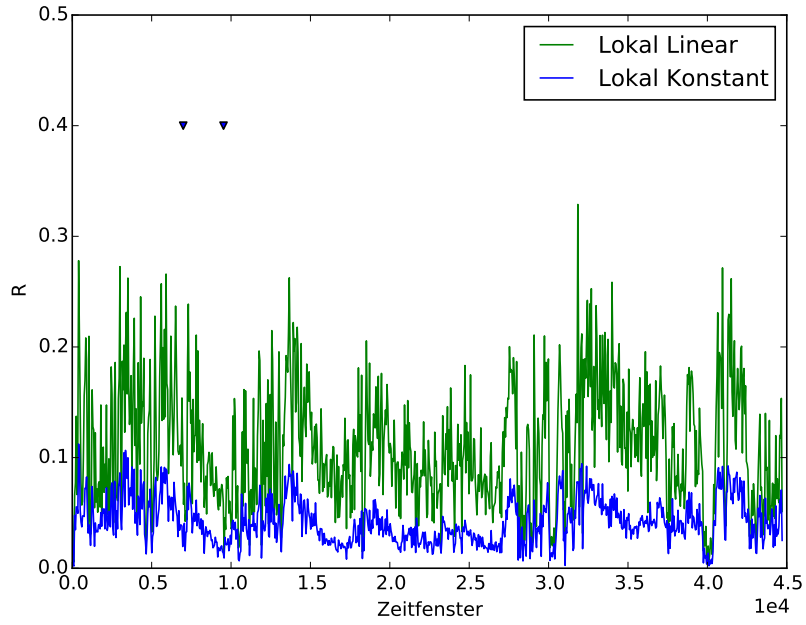
Dagegen erweist sich die Methode des lokal linearen Kernschätzers im Falle einer 2-dimensionalen Regressionsfunktion als vorteilhaft, besonders bei geringeren Datenmengen. Hier zeigte sich, dass die Schätzung vor allem am Rand, wo die Datendichte abnimmt, durch die lineare Annäherung die tatsächliche Regressionsfunktion wesentlich besser approximiert. Auch die direkt geschätzte Ableitung aus dem lokal linearen Kernschätzer liefert eine qualitative Verbesserung, allerdings nur mit geringem Unterschied zur numerisch berechneten. Der Grad des polynomialen Schätzers könnte auch hier einen Einfluss haben.

Dennoch zeigt sich, dass die lokal lineare Kernel-Regression im Falle höher dimensionaler Datensätze eine brauchbare Methode darstellt. Es bleibt zu untersuchen, inwiefern die Anzahl der Dimensionen einen Einfluss nehmen. Dadurch wäre es möglich, die Dynamik von komplexen Systemen auf verschiedene Parameter hin zu untersuchen.

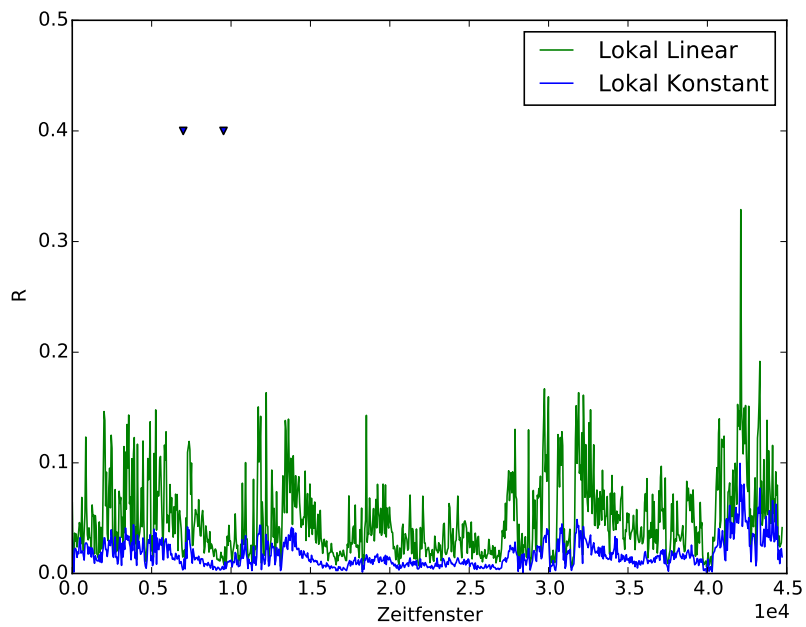
Anschließend wurde versucht, durch die zusätzliche Kernel-Gewichtung in zeitlicher Richtung eine Verbesserung der Auflösung von sich zeitlich ändernden Dynamiken zu erzielen. Der Vergleich mit den vorherigen Schätzmethoden zeigt, dass die Qualität bei gewichteten Zeitfenster sehr stark mit der zeitlichen Änderung zusammenhängt; im Bereich stärkerer Änderungen nahm auch die Qualität deutlich zu. Allerdings ließ sich keine Verbesserung gegenüber der Kernel-Regression mit ungewichteten Zeitfenstern feststellen. Dennoch zeigt die Analyse, dass die gewichteten

Zeitfenster wesentlich stärker auf zeitliche Änderungen reagieren. Es müssten weitere verschiedene nichtstationäre Prozesse im Zusammenhang mit der gewählten Bandbreite untersucht werden, um eindeutige Aussagen über den Zusammenhang treffen zu können.

Die Anwendung der untersuchten Regressionsmethoden auf realistische EEG Daten ließ keine eindeutige Aussage über einen Zusammenhang zwischen der Dynamik und den verzeichneten epileptischen Anfällen zu. Dennoch ist anzumerken, dass sich durch eine mehrdimensionale Analyse über mehrere Aufnahmestellen möglicherweise ein gewinnbringender Zusammenhang heraus stellt. Dies wird klar, wenn man den ähnlichen Verlauf der Schwankungen von zwei verschiedenen Zeitreihen, aufgenommen von unterschiedlichen Elektroden, vergleicht (siehe Abbildung 17). Die Analyse von synthetischen Daten zeigte, dass sich durch die Verwendung eines lokal linearen Kernschätzers im 2-dimensionalen Fall qualitative Verbesserungen gegenüber den anderen Regressionsmethoden im Zusammenhang mit der zugrunde liegenden Datenmenge einstellen. Es bleibt also zu untersuchen inwiefern die Dimension und der Grad des verwendeten polynomialen Schätzers eine Rolle spielen und welche Auswirkungen dies auf die Charakterisierung der Dynamik von Felddaten haben könnte.



(a) Zeitreihe 1



(b) Zeitreihe 2

Abbildung 17: Verlauf der Schwankung $R^{(i)}$ der Driftfunktionen aus EEG Daten aufgenommen von zwei verschiedenen Stellen.

Abbildungsverzeichnis

1	Beispiel einer generierten Zeitreihe mit zeitabhängigem Drift.	4
2	Schätzung der Drift- und Diffusionsfunktion eines stationären Langevin-Prozesses.	15
3	Differenz der geschätzten und originalen Driftfunktion.	16
4	Vergleich der mittleren quadratischen Abweichung der geschätzten Driftfunktion eines stationären Langevin-Prozesses.	17
5	Vergleich der mittleren quadratischen Abweichung der Ableitung der geschätzten Driftfunktion eines stationären Langevin-Prozesses. . . .	18
6	Vergleich der mittleren quadratischen Abweichung der Ableitung der geschätzten Driftfunktion eines stationären Langevin-Prozesses höheren Grades.	19
7	Schätzung der Driftfunktion eines stationären Langevin-Prozesses höheren Grades.	20
8	Vergleich der Schätzung einer 2-dimensionalen Regressionsfunktion mit verschiedenen Schätzmethoden.	22
9	Vergleich der mittleren quadratischen Abweichung der geschätzten 2-dimensionalen Regressionsfunktion.	23
10	Vergleich der mittleren quadratischen Abweichung der Ableitung einer geschätzten 2-dimensionalen Regressionsfunktion.	23
11	Veranschaulichung einer sich parameterabhängig, zeitlich ändernden Funktion.	24
12	Rekonstruktion des zeitabhängigen Teils der Driftfunktion nichtstationärer Langevin-Prozesse.	26
13	Vergleich der mittleren quadratischen Abweichung der geschätzten Driftfunktion nichtstationärer Langevin-Prozesse.	27
14	Schätzung der Drift- und Diffusionsfunktion aus EEG Daten innerhalb eines Zeitfensters.	29
15	Verlauf der Schwankung der Drift- und Diffusionsfunktion aus EEG Daten ermittelt mit ungewichteten Zeitfenstern.	31
16	Verlauf der Schwankung der Drift- und Diffusionsfunktion aus EEG Daten ermittelt mit gewichteten Zeitfenstern.	32
17	Verlauf der Schwankung der Driftfunktionen aus EEG Daten aufgenommen von zwei verschiedenen Stellen.	35

Literatur

- [1] A. Fuchs. *Nonlinear Dynamics in Complex Systems*. Springer-Verlag Berlin Heidelberg, 2013.
- [2] H. Haken. *Synergetics*. Springer Series in Synergetics. Springer-Verlag Berlin Heidelberg, 2004.
- [3] C. Honisch. Stochastische Analyse kollektiver Variabler in komplexen Systemen. Master's thesis, WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER, 2010.
- [4] D. Lamouroux. Kernel-basierte Schätzung der Fokker-Planck-Gleichung zur verbesserten Auflösbarkeit zeitlicher Veränderungen komplexer Dynamiken. Master's thesis, RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN, 2009.
- [5] D. Kleinhans. ESTIMATION OF DRIFT AND DIFFUSION FUNCTIONS FROM TIME SERIES DATA: A MAXIMUM LIKELIHOOD FRAMEWORK. *PHYSICAL REVIEW E*, 85(026705), 2012.
- [6] D. Lamouroux and K. Lehnertz. KERNEL-BASED REGRESSION OF DRIFT AND DIFFUSION COEFFICIENTS OF STOCHASTIC PROCESSES. *PHYSICS LETTERS A*, 373:3507–3512, 2009.
- [7] J. Prusseit and K. Lehnertz. STOCHASTIC QUALIFIERS OF EPILEPTIC BRAIN DYNAMICS. *PHYSICS REVIEW LETTERS*, 98:138103, 2007.
- [8] H. Risken. *The Fokker-Planck Equation*. Springer Series in Synergetics. Springer-Verlag Berlin Heidelberg New York, 2. edition, 1989.
- [9] N. G. van Kampen. ITÔ VERSUS STRATONOVICH. *JOURNAL OF STATISTICAL PHYSICS*, 24(1), 1981.
- [10] P. Hänggi and H. Thomas. STOCHASTIC PROCESSES: TIME EVOLUTION, SYMMETRIES AND LINEAR RESPONSE. *PHYSICS REPORTS*, 88(4):207–319, 1982.
- [11] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer Series in Statistics. Springer-Verlag Berlin Heidelberg New York, 2004.
- [12] C. L. Dickter. *EEG METHODS FOR THE PSYCHOLOGICAL SCIENCES*. SAGE Publications Inc., 2014.

- [13] S. J. Luck. *AN INTRODUCTION TO THE EVENT-RELATED POTENTIAL TECHNIQUE*. Massachusetts Institute of Technology, 2. edition, 2014.
- [14] D. L. Schomer and F. H. Lopes da Silva, editors. *NIEDERMEYER'S ELECTROENCEPHALOGRAPHY*. Lippincott Williams and Wilkins, 6. edition, 2011.

Plagiatserklärung

Hiermit versichere ich, dass die vorliegende Arbeit über die *Analyse nichtstationärer stochastischer Prozesse am Beispiel von EEG Daten* selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

(Datum, Unterschrift)

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in eine Datenbank einverstanden.

(Datum, Unterschrift)