

Ein optimiertes Glättungsverfahren motiviert durch eine technische Fragestellung

Frank Klinker and Günter Skoruppa

Preprint 2012-05

April 2012

EIN OPTIMIERTES GLÄTTUNGSVERFAHREN MOTIVIERT DURCH EINE TECHNISCHE FRAGESTELLUNG

FRANK KLINKER AND GÜNTER SKORUPPA

ZUSAMMENFASSUNG. Ausgehend von einer konkreten technischen Fragestellung diskutieren wir in dieser Notiz die Anwendung verschiedener Glättungsverfahren auf Datensätze mit vorgegebener Struktur. Wir stellen die Verfahren im Detail vor und besprechen die Vor- und Nachteile. Insbesondere stellen wir hier die symmetrisierte exponentielle Glättung vor, die ein sehr gutes Glättungsverhalten mit einem hohen Maß an Symmetrieerhaltung kombiniert.

1. AUSGANGSSITUATION UND FRAGESTELLUNG

Wir präsentieren in dieser Notiz eine Lösung zur folgenden technischen Problemstellung. Gegeben ist das Höhenprofil einer Seite einer Keilsicherungsscheibe (DIN 25201). Diese Scheibe ist beidseitig gezahnt mit jeweils einer grob und einer fein gezahnten Seite, siehe Abbildung 1. Die Ermittlung des Profils erfolgt mit Hilfe eines

ABBILDUNG 1. Keilsicherungsscheibe DIN 25201



2D/3D-Lasersensors. Dieser liefert über $N \leq 750$ äquidistanten Stützstellen einer Sekante der Scheibe eine Reihe von Messpunkten, wobei jeder einzelne Messpunkt der Höhe des Objekts über der jeweiligen Stützstelle entspricht. Zum schematischen Aufbau der Messapparatur vergleiche Abbildung 2. Das Ziel ist, die Zahl der Zähne der ausgemessenen Scheibenseite zu ermitteln. Dies geschieht durch Auszählen der

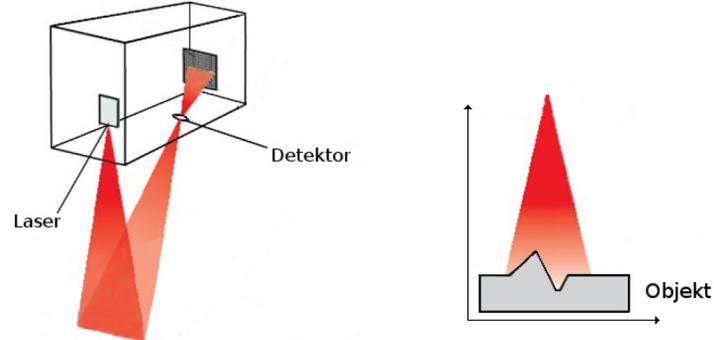
✉: Fakultät für Mathematik, TU Dortmund, 44221 Dortmund, Germany.

✉: frank.klinker@math.tu-dortmund.de .

✉: guenter.skoruppa@math.tu-dortmund.de .

Math. Semesterber. **59** (2012), no.1, 29-55. DOI [10.1007/s00591-012-0098-1](https://doi.org/10.1007/s00591-012-0098-1).

ABBILDUNG 2. Schematischer Aufbau der Messapparatur

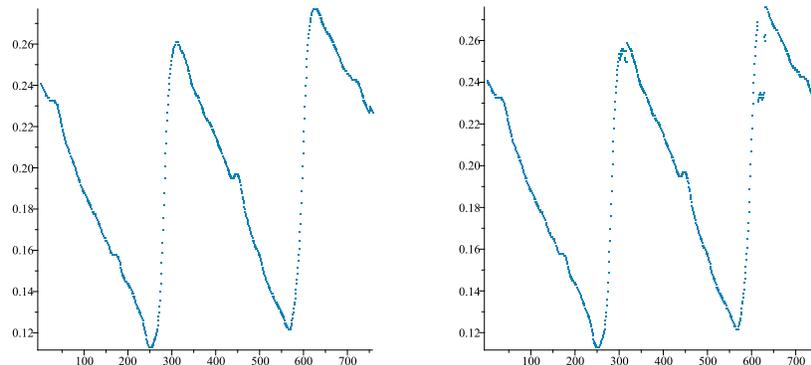


Maxima des Datensatzes. Dieses Auszählen und die daraus resultierende Entscheidung, ob sich die Scheibe in der richtigen Lage befindet oder gedreht werden muss, soll in einem automatisierten Prozess geschehen.

Dabei tritt das Problem auf, dass die Messwerte durch Verunreinigungen der Scheibe vom Soll abweichen können, zum Beispiel durch Verschmutzung während des Arbeitsprozesses oder auch durch Farbaufträge oder eine nicht ganz fehlerfreie Produktion des Werkstücks. Geringe Abweichungen, insbesondere solche durch Verunreinigungen, sollen toleriert werden und die Zahnzählung nicht behindern.

Die Grafiken in Abbildung 3 zeigen eine typische und eine hieraus erzeugte pathologische Messreihe. Gut erkennbar sind in Abbildung 3 die Imperfektionen an den nach rechts absteigenden Flanken. Diese werden als zusätzliche Zähne die Zählung verfälschen. Der rechte Teil der Abbildung 3 wurde aus dem linken durch manuell eingefügte, lokale Störungen in den Messwerten erzeugt. Solche Störungen treten auch in realen Beispielen auf, wie die Messreihe **I** zeigt.

ABBILDUNG 3. Typische und pathologische Messreihe



Fragestellung. Lassen sich unwesentliche Störungen derart beseitigen, dass das bereinigte Zahnprofil durch Auszählung der Spitzen unmittelbar die Entscheidung der technischen Fragestellung ermöglicht? Dabei darf die ursprüngliche Messreihe

verändert werden mit der Einschränkung, dass die Zahnzahl erhalten bleiben soll. Die Modifikation der Daten wird dabei umso besser sein, je mehr der „Charakter“ der ursprünglichen Daten erhalten bleibt.

2. LÖSUNGSANSÄTZE.

Glättungsoperationen, wie sie etwa in der Statistik eingesetzt werden, sind ein geeignetes Mittel zur Störungsbeseitigung in den Daten. Wir werden uns hier zunächst auf die Vorstellung zweier gängiger Verfahren beschränken und ihre Vor- und Nachteile diskutieren. Im Einzelnen sind das der *gleitende arithmetische Durchschnitt* (*moving arithmetic mean*, MA), und die *exponentielle Glättung* (*exponential average*, EA). Zur Diskussion allgemeiner Durchschnitte und ihrer gleitenden Varianten verweisen wir gerne auf [5]. Anschliessend bieten wir ein Glättungsverfahren an, das jeweils die Vorteile beider vereint. Dieses Verfahren nennen wir *symmetrisierte exponentielle Glättung*, (*symmetrized exponential average*, SEA).

2.1. Der gleitende arithmetische Durchschnitt (MA). Gegeben sei ein Datensatz¹ $(y_1, y_2, \dots, y_N, \dots)$. Die Idee des gleitenden arithmetischen Mittels ist es nun, statt des Wertes y_i ein arithmetisches Mittel \bar{y}_i eines Teils des Datensatzes zu ersetzen. Die Berechnung des gleitenden arithmetischen Durchschnitts benötigt zwei Parameter. Zum einen die *Gleitlänge* $n + 1 \leq N$ zum anderen das *Gewicht* $\ell \in \{0, \dots, n\}$. Die Spezialfälle $\ell = 0$ bzw. $\ell = n$ nennt man *rechts-* bzw. *linkseitigen gleitenden arithmetischen Durchschnitt der Länge $n + 1$* und den Spezialfall n gerade und $\ell = \frac{n}{2}$ nennt man *zentrierten gleitenden arithmetischen Durchschnitt der Länge $n + 1$* . Das oben angesprochene arithmetische Mittel wird dann aus den $n + 1$ Werten der Menge $\{y_{i-\ell}, \dots, y_{i+n-\ell}\}$ berechnet, also

$$\begin{aligned} \bar{y}_i &:= \frac{1}{n+1} (y_{i-\ell} + y_{i-\ell+1} + \dots + y_i + \dots + y_{i+n-\ell-1} + y_{i+n-\ell}) \\ &= \frac{1}{n+1} \sum_{k=-\ell}^{n-\ell} y_{i+k}, \end{aligned} \tag{2.1}$$

für $i = \ell + 1, \dots, N - n + \ell$. Der neue Datensatz ist insbesondere um n Werte kürzer als der alte, und zwar entfallen am Anfang ℓ und am Ende $n - \ell$ Werte. Ist der Datensatz nicht endlich, so entfallen selbstverständlich nur die Werte am Anfang.

Wie man die Parameter wählt, hängt von der Anwendung ab. Da in unserer Anwendung alle Werte der Messreihe gleichberechtigt sind, haben wir uns für die zentrierte Variante entschieden und MA bezeichnet im Folgenden genau diese.² Hier ist der Verlust an Daten am Anfang und am Ende des Datensatzes gleich. In diesem Fall

¹Ist der Datensatz endlich, so sei im Folgenden immer N die Anzahl der Elemente.

²Die Beschränkung auf die zentrierte Variante ist insofern nicht zwingend notwendig, da einerseits die Wahl von $\ell \in \{0, \dots, n\}$ unerheblich für das Ergebnis der Glättung ist, und andererseits in dem hier vorliegenden Fall $n \ll N$ auch die Symmetrie der Originalmessreihe im Wesentlichen erhalten bleibt.

ist dann $n = 2\ell$ und (2.1) wird zu

$$\bar{y}_i = \frac{1}{2\ell + 1} \sum_{k=-\ell}^{\ell} y_{i+k}. \quad (2.2)$$

Zur Berechnung des MA muss man nicht in jedem Schritt die gesamte Summe (2.1) bzw. (2.2) berechnen. Die Folge der Durchschnitte \bar{y}_i erfüllt vielmehr die folgende rekursive Gleichung:

$$\bar{y}_i = \bar{y}_{i-1} + \frac{1}{2\ell + 1} (y_{i+\ell} - y_{i-1-\ell}). \quad (2.3)$$

Daher sieht ein den Aufwand minimierender Berechnungsalgorithmus für den zentrierten MA wie folgt aus.

- (1) Wähle einen positiven ganzzahligen Parameter ℓ , so dass $2\ell + 1 \leq N$.
- (2) Berechne den Startwert

$$\bar{y}_{\ell+1} := \frac{1}{2\ell + 1} (y_1 + y_2 + \dots + y_{2\ell+1}).$$

- (3) Berechne für $i = \ell + 2, \dots, N - \ell$ rekursiv

$$\bar{y}_i := \bar{y}_{i-1} + \frac{1}{2\ell + 1} (y_{i+\ell} - y_{i-1-\ell}).$$

Bemerkungen zum MA.

- Je größer die Gleitlänge $n + 1 \in \mathbb{N}$ gewählt wird, desto stärker wird die Glättung ausfallen. Zu große Gleitlängen $n + 1$ mitteln die Werte so stark, dass gegebenenfalls wichtige charakteristische Eigenschaften der Originalmessreihe verloren gehen. Mit anderen Worten, der Einfluss des einzelnen Messwertes y_i geht verloren, weil alle Werte mit dem gleichen Gewicht $\frac{1}{n+1}$ in den Durchschnitt eingehen.
- Neben dem eventuellen Verlust an charakteristischen Eigenschaften hat der MA insbesondere bei großen Gleitlängen den Nachteil, dass sie den Datensatz stark verkürzen.
- Für den Grenzfall $2\ell + 1 = N$ erhält man genau einen Wert $\bar{y}_{\ell+1}$, der dem gewöhnlichen arithmetischen Mittel aller Messpunkte entspricht, und für $\ell = 0$ bekommt man als Ergebnis die ursprüngliche Messreihe zurück.
- Sieht man einmal von der einmaligen Startwertberechnung ab, benötigt die Berechnung eines neuen Datenwertes 3 Rechenoperationen, vgl. (2.3).
- Der MA berücksichtigt die Symmetrie eines Systems: Hat der beschränkte Ausgangsdatsatz die Symmetrie $y_i = y_{N-i}$, so hat der geglättete Datensatz diese ebenfalls.

2.2. Die exponentielle Glättung (EA). Bei dem in Abschnitt 2.1 vorgestellten Verfahren der Glättung wird nur ein Ausschnitt des gesamten Datensatzes um einen ausgezeichneten Wert y_i betrachtet und alle Werte dieses Abschnitts tragen mit dem gleichen Gewicht $\frac{1}{n+1}$ zur Glättung, also zum Wert \bar{y}_i , bei, siehe (2.1). Bei der exponentiellen Glättung werden im Gegensatz dazu bei der Berechnung alle Werte der Testreihe bis zur betrachteten Stelle y_i mit einbezogen. Dabei tragen allerdings die "älteren Werte" - das sind die Werte im Datensatz, die sich vor der betrachteten

Stelle y_i befinden – mit fallendem Gewicht zur Berechnung der Glättung bei. Die Definition des EA erfolgt rekursiv, analog zur Berechnung des MA gemäß (2.3).

Gegeben sei ein endlicher oder unendlicher Datensatz $(y_1, y_2, \dots, y_N, \dots)$. Der EA dieses Datensatzes ist gegeben durch

$$\begin{aligned}\hat{y}_1 &:= y_1, \\ \hat{y}_i &:= (1 - \alpha)\hat{y}_{i-1} + \alpha y_i, \quad \text{für } i = 2, \dots, N.\end{aligned}\tag{2.4}$$

Die Berechnung des EA hängt hier von einem Parameter $0 \leq \alpha \leq 1$ ab. Für $\alpha \neq 0$ heißt das Inverse dieses Parameters, $\frac{1}{\alpha}$, die *Gleitlänge* der exponentiellen Glättung. Wie auch im Fall des MA gibt es hier eine geschlossene Berechnungsformel. In dieser sieht man auch das Fallen der Gewichte, mit dem die älteren Datenwerte zum Durchschnittswert beitragen. Es gilt

$$\begin{aligned}\hat{y}_i &= (1 - \alpha)^{i-1}y_1 + \alpha(1 - \alpha)^{i-2}y_2 + \dots + \alpha(1 - \alpha)y_{i-1} + \alpha y_i \\ &= (1 - \alpha)^{i-1}y_1 + \alpha \sum_{r=2}^i (1 - \alpha)^{i-r}y_r.\end{aligned}\tag{2.5}$$

Die algorithmische Beschreibung der Berechnung des EA erfolgt wieder über die rekursive Formel (2.4).

- (1) Wähle Parameter $\alpha \in [0, 1]$ und setze $\hat{y}_1 := y_1$.
- (2) Berechne rekursiv für³ $i = 2, 3, \dots, N$:

$$\hat{y}_i = (1 - \alpha)\hat{y}_{i-1} + \alpha y_i.$$

Bemerkungen zum EA.

- Anders als der zentrierte MA ist das Verfahren einseitig, d.h. es betrachtet von y_i aus gesehen nur frühere Daten. Dies macht EA besonders interessant für Prognosen, etwa bei Börsenkursen. Für eine Diskussion dieser Anwendung siehe zum Beispiel [5].
- Dass die exponentielle Glättung einseitig arbeitet und die komplette Historie reflektiert, kann bei nahezu periodischen oder anderen Symmetrien aufweisenden Datensätzen von Nachteil sein, da die Einseitigkeit diese Symmetrie zerstört. Dieses Phänomen wird insbesondere am Beispiel in Abschnitt 3.4 deutlich werden.
- Die Einseitigkeit des EA hat auch einen negativen Einfluss auf die Glättung, insbesondere wenn ein Bereich, der eine starke Glättung erfordert, recht früh im Datensatz zu finden ist. Vergleiche dazu Abbildung 9 und beachte, dass dort der rEA gemäß (2.6) angewendet wurde und somit der frühe Bereich rechts in den Kurven zu finden ist.

2.3. Die symmetrisierte exponentielle Glättung (SEA). In diesem Abschnitt liefern wir eine Variante der exponentiellen Glättung, die den Nachteil der Einseitigkeit des EA kompensiert. Dieses Glättungsverfahren kombiniert die guten Glättungseigenschaften des EA und dessen Betonung des Wertes y_i bei der Berechnung der Glättung \hat{y}_i , einerseits, mit der Symmetrieerhaltung des MA, andererseits.

³Ist der Datensatz unendlich, so entfällt selbstverständlich die obere Beschränkung hier, genauso wie schon in (2.4).

Zur Konstruktion benötigen wir einen endlichen Datensatz und wir schauen uns zur Berechnung der Glättung den gesamten Datensatz an und nicht – wie beim EA – nur den „linken Teil“ oder – wie beim MA – nur einen Ausschnitt.

Sei also der endliche Datensatz (y_1, y_2, \dots, y_N) gegeben. Zu diesem konstruieren wir durch Auslesen des Datensatzes von rechts nach links einen neuen Datensatz (z_1, \dots, z_N) mit $z_j := y_{N+1-j}$. Wir definieren dann die *exponentielle Glättung von rechts* (rEA) des Datensatzes (y_1, \dots, y_N) als die exponentielle Glättung des Datensatzes (z_1, \dots, z_N) . Diesen bezeichnen wir mit $\check{y}_i := \hat{z}_{N-i+1}$. Der rEA berechnet sich dann analog zu (2.5) gemäß

$$\begin{aligned} \check{y}_i &= (1 - \alpha)^{N-i} y_N + \alpha(1 - \alpha)^{N-i-1} y_{N-1} + \dots + \alpha(1 - \alpha) y_{i+1} + \alpha y_i \\ &= (1 - \alpha)^{N-i} y_N + \alpha \sum_{r=i}^{N-1} (1 - \alpha)^{r-i} y_r. \end{aligned} \tag{2.6}$$

Die rekursive Variante berechnet sich dann analog zu (2.4) zu

$$\begin{aligned} \check{y}_N &:= y_N, \\ \check{y}_i &:= (1 - \alpha)\check{y}_{i+1} + \alpha y_i \quad \text{für } i = N - 1, \dots, 1. \end{aligned} \tag{2.7}$$

Definition: SEA. Die *symmetrisierte exponentielle Glättung* (SEA) des Datensatzes ist dann definiert als das arithmetische Mittel aus EA und rEA:

$$\bar{y}_i := \frac{1}{2} (\hat{y}_i + \check{y}_i). \tag{2.8}$$

Die entsprechende algorithmische Beschreibung des SEA lautet

- (1) Wähle Parameter $\alpha \in [0, 1]$ und setze $\hat{y}_1 := y_1$ und $\check{y}_N = y_N$.
- (2) Berechne rekursiv für $i = 2, \dots, N$ den EA

$$\hat{y}_i = (1 - \alpha)\hat{y}_{i-1} + \alpha y_i,$$

und für $j = N - 1, \dots, 1$ den rEA

$$\check{y}_j = (1 - \alpha)\check{y}_{j+1} + \alpha y_j.$$

- (3) Berechne für $i = 1, \dots, N$ den SEA

$$\bar{y}_i = \frac{1}{2} (\hat{y}_i + \check{y}_i).$$

Bemerkungen zum EA, rEA und SEA.

- Bei der Berechnung des EA, rEA und des SEA wird der Datensatz im Gegensatz zum MA nicht verkürzt. Das ist insbesondere von Vorteil, wenn zur Glättung hohe Gleitlängen benötigt werden.
- Die Berechnung eines neuen Datenwertes benötigt für EA und rEA wie beim MA ebenfalls 3 Operationen, vergleiche (2.4). Zur Berechnung des SEA benötigt man insgesamt 8 Operationen.
- Zur Berechnung eines Wertes des SEA werden alle Werte der Ursprungsreihe mit herangezogen und nicht, wie beim MA und EA, nur ein Teil der Werte.

- Je kleiner α , desto stärker wird die Glättung ausfallen. Das bedeutet aber gleichzeitig, dass mehr Charakteristiken des Datensatzes verloren gehen. Die Grenzfälle liefern für $\alpha = 1$ keine Veränderung der Messreihe und für $\alpha = 0$ eine konstante Folge, nämlich das arithmetische Mittel aus y_1 und y_N .
- Bei den Berechnungen des EA und des SEA geht der Wert y_i mit dem Gewicht α ein. Bei der Berechnung des MA in (2.2) geht y_i mit dem Gewicht $\frac{1}{n+1}$ ein. Das und auch die Diskussion der Grenzfälle motiviert die Interpretation und die Definition des Parameters $\frac{1}{\alpha}$ als Gleitlänge.
- Im Gegensatz zu MA geht der Wert y_i bei der Berechnung des SEA mit dem größten Gewicht ein. Dies gilt schon für den EA und den rEA, vgl. (2.5) und (2.6).

3. BEISPIELE

In den Abschnitten 3.1 und 3.2 widmen wir uns den vorgestellten Beispielen realer Datensätze gemäss Abbildung 3. Daneben betrachten wir noch weitere Beispiele um die vorgeschlagenen Glättungsverfahren zu testen und ihre Eigenschaften zu verdeutlichen. Dies sind im Einzelnen in Abschnitt 3.3 eine Überlagerung zweier Schwingung und in Abschnitt 3.4 ein fourierentwickeltes Rechtecksignal.

Die in den Abschnitten 3.1 und 5 verwendeten Datensätze wurden uns freundlicherweise von der Firma [KLANZ SOFTWARE ENGINEERING](#) zur Verfügung gestellt.

3.1. Laserausmessung eines Zahnscheibenprofils. Zunächst untersuchen wir eine typische Messreihe. Bei dieser ist an einer absteigenden Flanke ein zusätzliches Maximum zu erkennen, das bei der Zählung der Maxima allerdings vernachlässigt werden soll.

ABBILDUNG 4. Originaldaten einer Laserausmessung und Glättung mittels MA ($n + 1 = 20$)

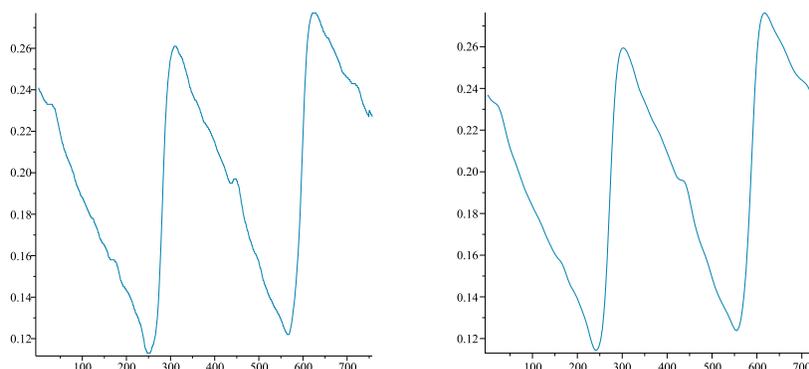
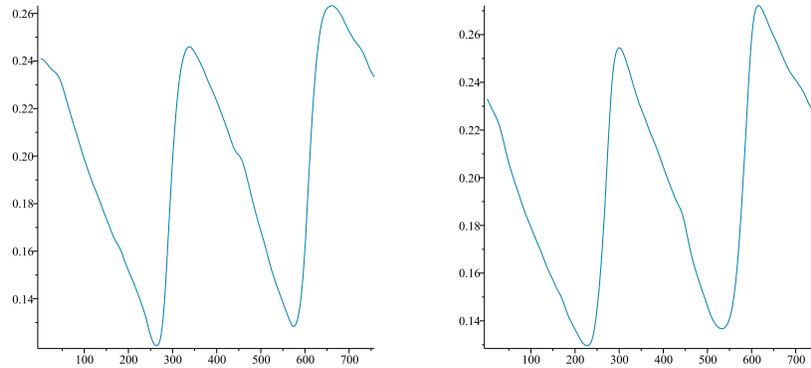
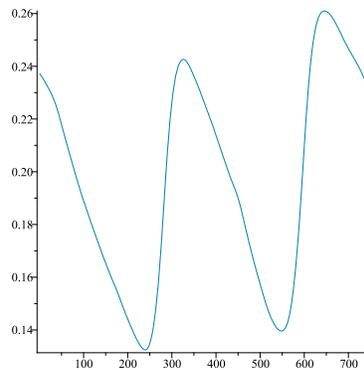


ABBILDUNG 5. Glättung mittels EA bzw. rEA ($\alpha^{-1} = 20$)ABBILDUNG 6. Glättung mittels SEA ($\alpha^{-1} = 20$)

Alle Ansätze führen schon bei niedriger Gleitlänge (hier 20) zu guten bis sehr guten Ergebnissen und beseitigen die Störungen auf den rechten Flanken. Diese Glättung ist bei SEA am besten ausgeprägt.

Mit Beseitigung der Störung ist an dieser Stelle eine Glättung zu verstehen, die im Sinne der Fragestellung keine weiteren als die gewünschten Extrema aufweist. Insofern, liefert auch die Glättung mittels MA in Abbildung 4 trotz der auffälligen Störung an der Flanke ein zufriedenstellendes Ergebnis.

In Abschnitt 6 werden wir kurz auf die Wahl der Gleitlänge hinsichtlich der Ausgangsfragestellung eingehen.

3.2. Daten mit stark destrukturierten Extrema. Durch eine starke Destrukturierung des Datensatzes insbesondere in den Extrema kommt es zu einer Vielzahl unerwünschter Maxima. Das Ziel ist auch hier die Ausglättung dieser.

ABBILDUNG 7. Originaldaten einer Laserausmessung mit Artefakten

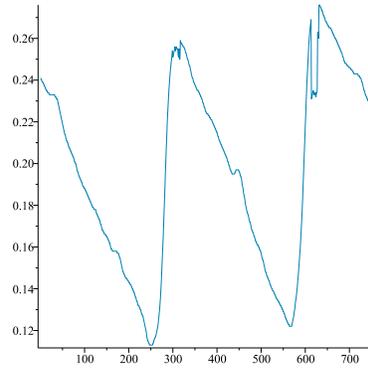
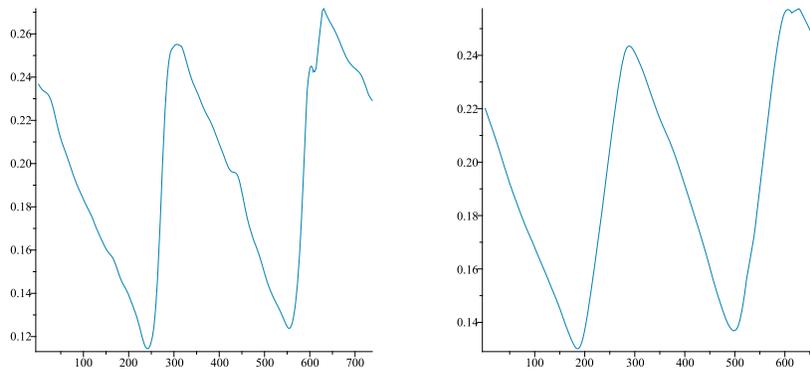
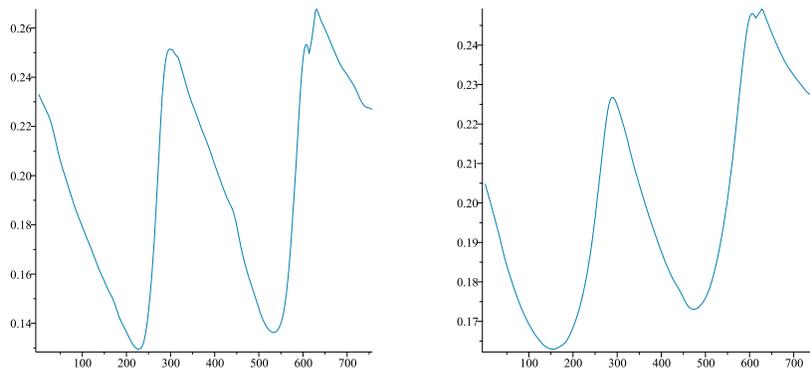
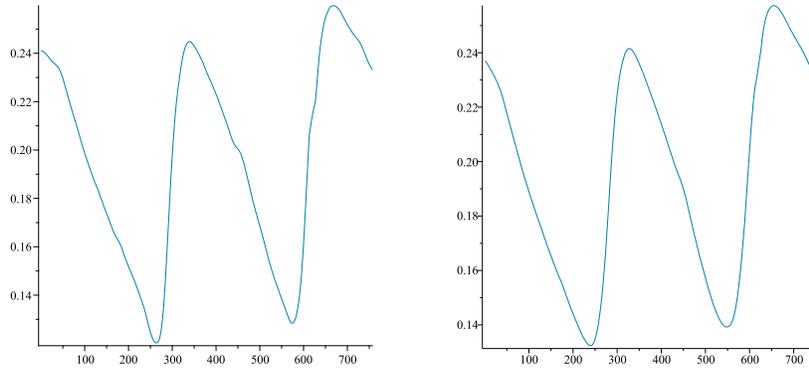
ABBILDUNG 8. Glättung mittels MA ($n + 1 = 20$ und $n + 1 = 90$)ABBILDUNG 9. Glättung mittels rEA ($\alpha^{-1} = 20$ und $\alpha^{-1} = 90$)

ABBILDUNG 10. Glättung mittels EA bzw. SEA ($\alpha^{-1} = 20$)

In den Abbildungen 7 und 9 sieht man, dass MA und rEA es nicht schaffen, die starken Artefakte rechts in der Grafik zu glätten. Das Ergebnis kann auch mit steigender Gleitlänge nicht verbessert werden. Beim rEA liegt das an der Tatsache, dass die Störungen sehr früh auftreten und daher nur wenig Werte zur Berechnung der Glättung herangezogen werden. Dieser Effekt wird dadurch, dass die Störungen recht stark ausfallen, noch verstärkt. Beim MA liegt es an der Tatsache, dass alle Werte mit dem gleichen Gewicht auftreten und deshalb im Allgemeinen sehr starke Schwankungen in der Messreihe erst bei hoher Gleitlänge aufgefangen werden können.

EA und insbesondere SEA liefern hier auch bei kleiner Gleitlänge gute bis sehr gute Ergebnisse. Bei EA liegt es daran, dass die großen Artefakte sehr spät auftreten und deshalb sehr viele Werte in die Berechnung mit einfließen (hier ca 80%). Bei der Berechnung des SEA spielt die Stelle des Auftretens der Artefakte keine Rolle, da der Mangel des rEA durch die gleichzeitige Verwendung des EA aufgehoben wird.

3.3. Überlagerung zweier periodischer Schwingungen. Als erstes künstliches Beispiel zum Testen der Glättungsverfahren betrachten wir eine Überlagerung zweier periodischer Schwingungen mit unterschiedlichen Frequenzen. Dabei ist die Frequenzdifferenz hoch gewählt und die Amplitude der hochfrequenten Störung klein im Vergleich zur Amplitude der niederfrequenten Grundschwingung.

ABBILDUNG 11. Originaldaten der Überlagerung

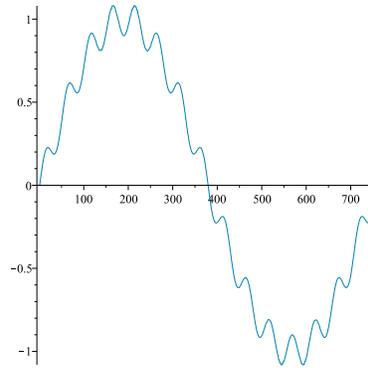
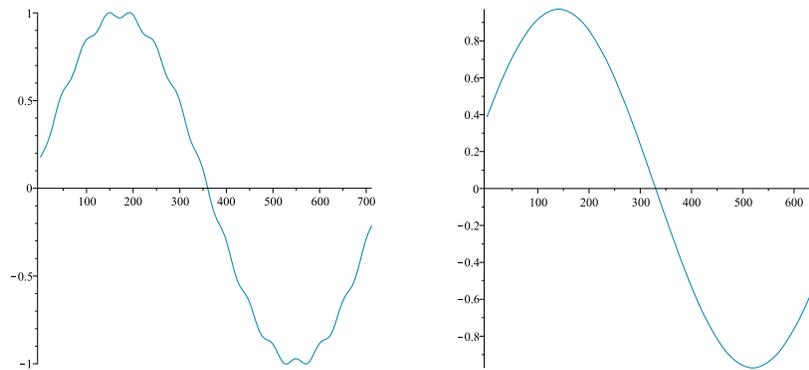
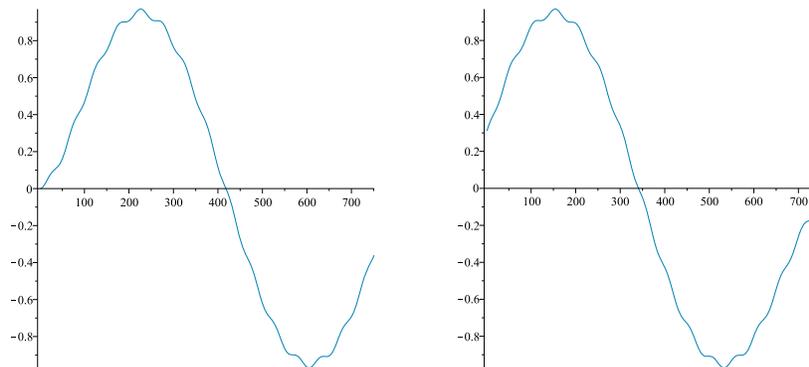
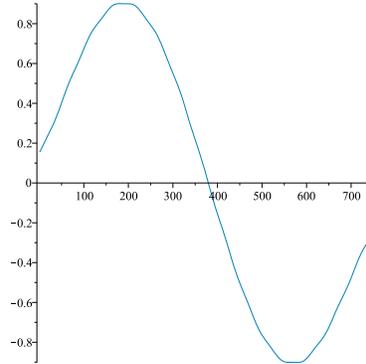
ABBILDUNG 12. Glättung mittels MA ($n + 1 = 40$ und $n + 1 = 100$)ABBILDUNG 13. Glättung mittels EA bzw. rEA ($\alpha^{-1} = 40$)

ABBILDUNG 14. Glättung mittels SEA ($\alpha^{-1} = 40$)

Die Glättungsverfahren MA, EA und rEA haben in dem Beispiel der überlagerten Schwingung Probleme. MA schafft es mit steigender Gleitlänge jedoch eher – wenn auch nicht viel eher – als rEA und EA die Oberschwingungen auszuglätten. In diesem Beispiel ist die gleiche Gewichtung aller Beiträge zur Glättung ein Vorteil des MA gegenüber dem EA. Wieder liefert auch in diesem Beispiel der SEA das beste Ergebnis, denn schon bei vergleichbarer mittlerer Gleitlänge von 40 werden die Oberschwingungen beseitigt.

Wir möchten an dieser Stelle betonen, dass der Charakter der Kurve als Überlagerung von Schwingungen (Abb. 11) beim Übergang zur Glättung (Abb. 14) selbstverständlich verändert wird. Im Hinblick auf unsere Fragestellung wird der Charakter der Kurve jedoch allein durch die Grundschwingung bestimmt, und die Oberschwingung wird als Störung angesehen. Insofern erhält auch hier die Glättung den Charakter der Ausgangskurve.

Diese Veränderung des Schwingungscharakters, also das Ausblenden der Oberschwingungen, bedeutet, dass die Glättungsverfahren im Fall hoher Gleitlängen als Tiefpassfilter wirken. Zur Diskussion des MA und des EA als Tiefpassfilter und zu Eigenschaften der zugehörigen Filterfunktionen siehe zum Beispiel [6]. Ein Tiefpassfilter ruft in der Regel eine Phasenverschiebung der Grundschwingung hervor. Dies kann man hier gut durch das Verschieben der Extrema nach rechts bzw. links bei der Glättung mittels EA bzw. rEA (Abb. 13) erkennen. Vergleicht man die Abbildungen 11 und 14 mit 13 so sieht man dass die Symmetrisierung bei der Anwendung des SEA die zwei Phasenverschiebungen kompensiert und somit keine nennenswerte Verschiebung der Extrema auftritt.

3.4. Rechtecksignal mit endlicher Fourierentwicklung. Wir betrachten als weiteres Beispiel die Fourierentwicklung eines Rechtecksignals bzw. genauer die Teilsumme zehnter Ordnung der Fourierentwicklung. Deutlich erkennt man in Abbildung 15 das so genannte Gibbs'sche Phänomen, das durch die hochfrequente Überschwingungen an den Sprüngen des Signals charakterisiert ist. Dieses Phänomen der Überschwingung tritt unabhängig davon auf, wie spät die Fourierreihe abgebrochen wird, und ihre Amplitude hängt von der Amplitude des Rechtecksignals

ab. Das Phänomen tritt nicht nur bei Rechecksignalen auf, sondern ist grundsätzlich zu finden, wenn die Ausgangsfunktion einen Sprung aufweist. Dabei nimmt die Amplitude der Überschwingung zwar mit steigender Ordnung der Fourierentwicklung ab, unterschreitet jedoch niemals ca. $\sim 9\%$ der Sprunghöhe. Vergleiche dazu den schönen Übersichtsartikel [3] bzw. den klassischen Artikel [1] oder die Ausführungen in [4].

ABBILDUNG 15. Originaldaten der Fourierentwicklung

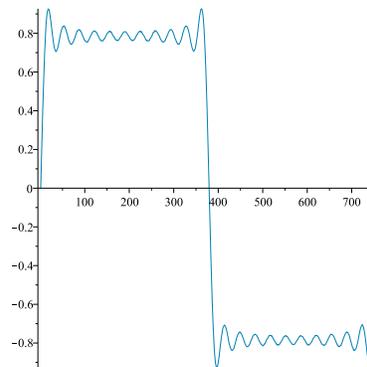
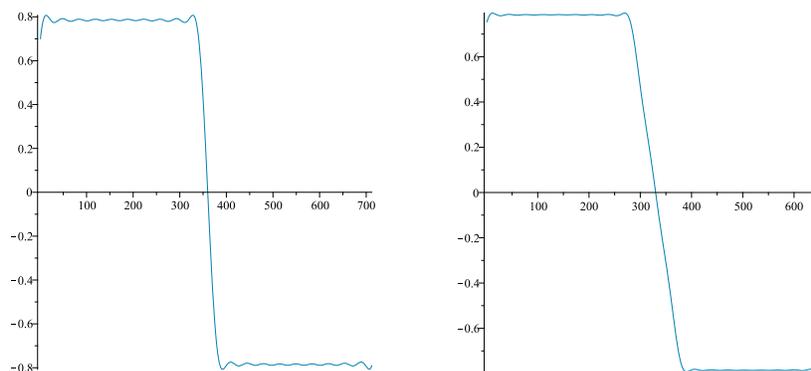
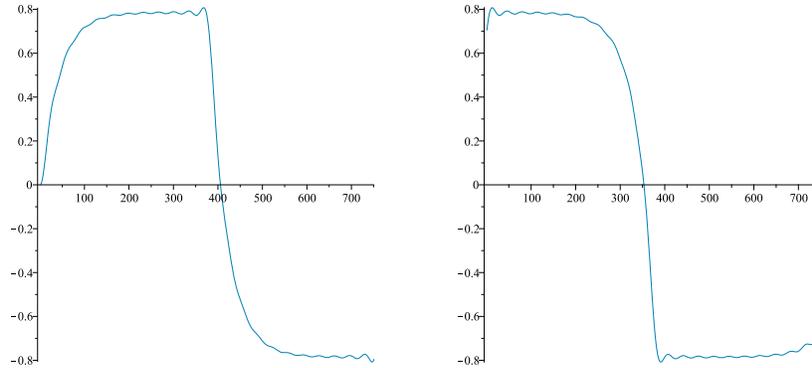
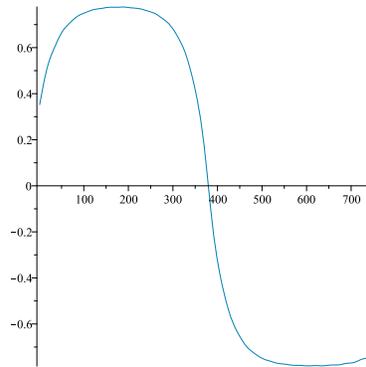
ABBILDUNG 16. Glättung mittels MA ($n + 1 = 40$ und $n + 1 = 100$)

ABBILDUNG 17. Glättung mittels EA bzw. rEA ($\alpha^{-1} = 40$)ABBILDUNG 18. Glättung mittels SEA ($\alpha^{-1} = 40$)

Die Glättungsverfahren MA, EA und rEA liefern auch hier kein ausreichendes Ergebnis. Das wird auch mit steigender Gleitlänge nicht besser. Zwar werden die Amplituden der Schwingungen auf den Plateaus des Rechtecksignals verringert, verschwinden jedoch auch bei Erhöhung der Gleitlänge nicht in einem befriedigendem Maße. Auffällig ist jedoch insbesondere, dass die Gibbschen Überschwinger von keinem der drei Verfahren aufgefangen werden. An diesem Beispiel sieht man auch sehr gut, dass der EA die Symmetrie der Ausgangsdaten nicht widerspiegelt.

Ein optimales Ergebnis liefert auch in diesem Beispiel der SEA sowohl hinsichtlich der Glättung schon bei mittlerer Gleitlänge (hier 40) als auch hinsichtlich der Symmetrieeerhaltung.

4. DIE BEHANDLUNG VON AUSREISSERN

4.1. Der Median und der gleitende Median. Gegeben sei die Menge von $2\ell+1$ Werten, $\{z_1, \dots, z_{2\ell+1}\}$. Zur Bildung des *Medians* dieser Menge geht man wie folgt vor. Man sortiere die Werte der Größe nach und erhält $\{z_{i_1}, z_{i_2}, \dots, z_{i_\ell}, \dots, z_{i_{2\ell+1}}\}$ mit $z_{i_1} \leq z_{i_2} \leq \dots \leq z_{i_{2\ell+1}}$. Der Median ist dann der Wert in der Mitte: $m := z_{i_{\ell+1}}$.

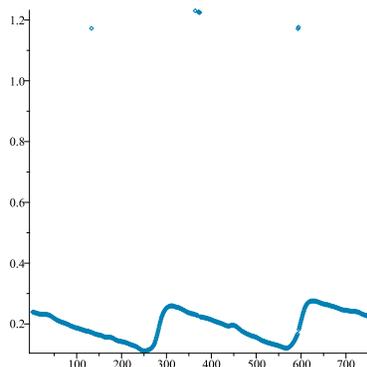
Es sei nun wie in Abschnitt 2 der Datensatz $(y_1, y_2, \dots, y_N, \dots)$ als eine Reihe von Messpunkten gegeben. Der *gleitende Median mit Gleitlänge* $2\ell + 1$ dieser Messreihe ist eine Reihe von Daten, m_i , die durch die folgende Vorschrift definiert sind:

$$m_i := \text{Median der Werte } \{y_{i-\ell}, \dots, y_i, \dots, y_{i+\ell}\} \quad (4.1)$$

für $i = \ell + 1, \dots, N - \ell$. Auch hier entfällt natürlich die obere Grenze, wenn der Ausgangsdatsatz endlich ist. Der Algorithmus zur Berechnung des Medians benötigt ein Sortierverfahren, für welches man etwa in [2] Beispiele mit zugehörigen Algorithmen finden kann.

4.2. Die Glättung von Messreihen mit Ausreißern. Ein Problem, das in der Praxis auftreten kann, sind Ausreißer im Datensatz. Im Gegensatz zu den Artefakten aus Abschnitt 3.2 sind Ausreißer dadurch charakterisiert, dass sie eine wesentliche größere Höhe relativ zu den Extrema des Datensatzes haben, siehe⁴ Abbildung 19. Solche extremen Werte würden bei den hier vorgestellten Verfahren immer Einfluss auf die Glättung haben und weiterhin als zusätzliche Extrema zu sehen sein, siehe Abbildungen 20 und 21. Um diese Ausreißer auszusortieren kann man zunächst – bevor der eigentliche Glättungsprozess mit den hier vorgestellten Methoden erfolgt – den Datensatz mit Hilfe eines gleitenden Medians bearbeiten. Das Ergebnis dieser Medianbildung findet man in Abbildung 22 mit minimaler Gleitlänge (hier 9) sowie mit Gleitlänge vergleichbar mit denen der verwendeten Glättungsverfahren. Man sieht neben der Eigenschaft die Ausreißer abzufangen, dass der Median glättende Eigenschaften hat. Hierzu verweisen wir auf den nächsten Abschnitt.

ABBILDUNG 19. Originaldaten einer Messung mit Ausreißern



⁴Hier haben wir in die Originalmessreihe aus Abbildung 3 bzw. Abschnitt 3.1 künstlich Ausreißer eingefügt.

ABBILDUNG 20. Glättung mittels EA bzw. rEA ($\alpha^{-1} = 40$)

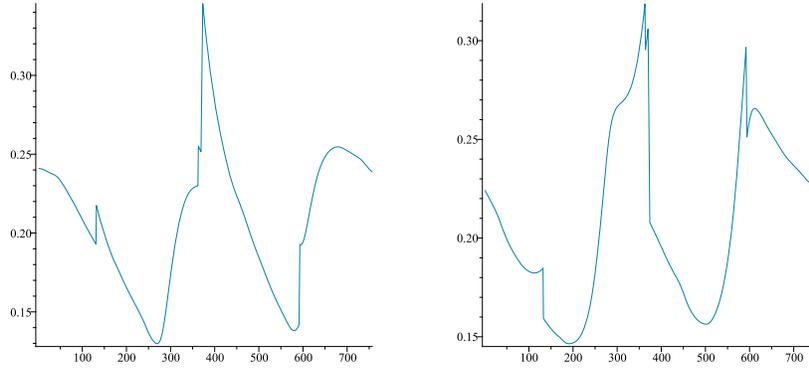


ABBILDUNG 21. Glättung mittels MA ($n + 1 = 40$) bzw. mittels SEA ($\alpha^{-1} = 40$)

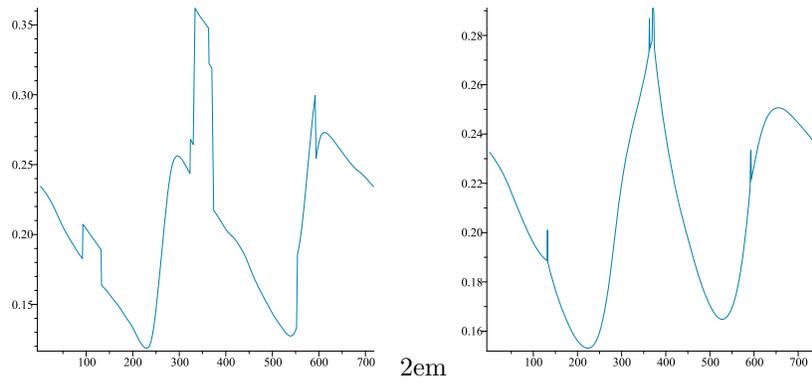
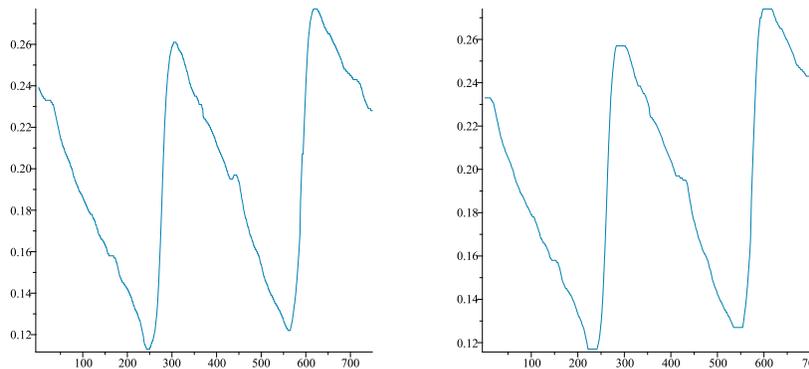


ABBILDUNG 22. Medianbildung mit Gleitlänge 9 und 40



Damit die Ausreißer aus dem Datensatz wirklich eliminiert werden, muss die Gleitlänge des gleitenden Medians geeignet gewählt werden. Die Wahl der *minimalen Gleitlänge* zur Eliminierung aller Ausreißer geschieht wie folgt:

Wähle die Gleitlänge $2\ell + 1 \leq N$ des Medians derart, dass sich für jede $(2\ell + 1)$ -elementige Teilmenge $\{y_k, \dots, y_{k+2\ell}\}$ des Ausgangsdatsatzes höchstens ℓ Ausreißer in dieser Teilmenge befinden. Insbesondere reicht bei einzelnen Ausreißern, die nicht gehäuft auftreten – das heißt, der Abstand zwischen einzelnen Ausreißer ist mindestens zwei – eine Gleitlänge von 3. In unserem Beispiel gemäß Abbildung 19 gibt es vier aufeinanderfolgende Ausreißer, so dass eine Gleitlänge von 9 notwendig ist.

4.3. Der gleitende Median als Glättungsverfahren. Mit Blick auf Abbildung 22 kann man geneigt sein, den gleitenden Median selbst als Glättungsverfahren in Betracht zu ziehen. Aufgrund der bisherigen Diskussion und der folgenden Betrachtung ist davon allerdings abzuraten. Ähnlich wie beim MA muss man auch beim gleitenden Median eine Verkürzung des Datensatzes in Kauf nehmen, wenn man ihn zur Glättung benutzt. In den folgenden Abbildungen sind zum Vergleich der MA (links) und der gleitende Median (rechts) mit Gleitlänge 20 bzw. 40 zu sehen (Abbildung 23 bzw. 24). Als Grundlage haben wir hier die Originaldaten der Laserausmessung aus Abschnitt 3.1 gewählt. Wie man schon an diesem Beispiel erkennt, besitzt der MA im Gegensatz zum gleitenden Median die besseren Glättungseigenschaften. Darüberhinaus neigt der Datensatz nach Anwendung des gleitenden Medians bei steigender Gleitlänge zur Plateaubildung in den Extrema.

ABBILDUNG 23. MA und Medianbildung mit Gleitlänge 20

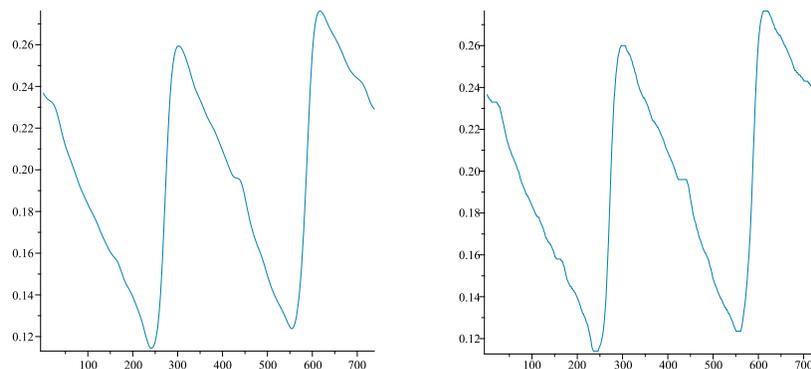
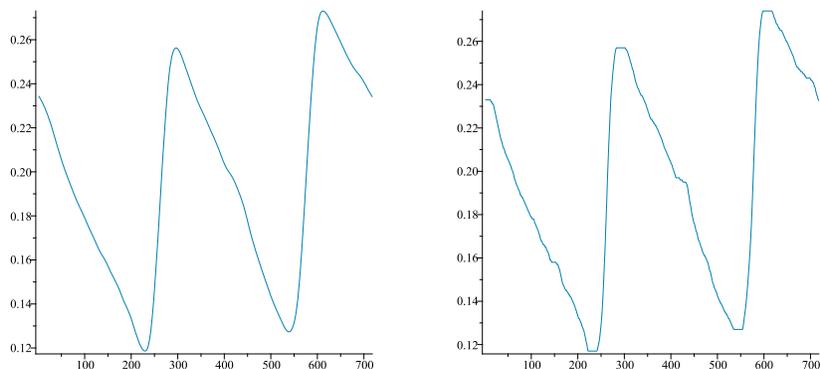


ABBILDUNG 24. MA und Medianbildung mit Gleitlänge 40



5. PRAXISDATEN IM VERGLEICH

5.1. Vorbereitung der Praxisdaten. In diesem Abschnitt stellen wir anhand von vier Messreihen zu der in Abschnitt 1 beschriebenen technischen Fragestellung unsere diskutierten Verfahren vor und begründen anhand der Ergebnisse unser Fazit. Desweiteren werden wir eine notwendige Vorbereitung der Datensätze aus der Praxis begründen. Wir betrachten hier jeweils zwei Messungen für die grobe Seite und für die feine Seite der Zahnscheibe.

Die Messreihen sind derart, dass vor und nach den eigentlichen Messwerten ein Nulllauf auftritt. Die Länge dieser Nullläufe ist für die vorliegenden Messreihen wie folgt:

Messung	I	II	III	IV
Nullvorlauf	40	150	80	30
Nullnachlauf	110	110	40	80

Da der Nulllauf verhältnismäßig lang ist, kann man ihn nicht mehr als Menge von Ausreißern behandeln, vergleiche Abschnitt 4.1. Eine Medianbildung ist also zur Beseitigung nicht angezeigt. Wie unsere Diskussion zeigen wird, ist eine Abtrennung des Nulllaufs als Vorbereitung des Datensatzes sinnvoll und sollte der Behandlung des Nulllaufs als Teil des Datensatzes vorgezogen werden.

In den Abbildungen 25-29 und 31-45 ist für die Messreihe jeweils links die Messreihe inklusive Nulllauf sowie ihre Glättungen zu sehen. Dabei haben wir bei den geglätteten Reihen zur besseren Vergleichbarkeit auf den relevanten Teil der Grafik gezoomt. Rechts findet man dann jeweils die um den Nulllauf bereinigte Messreihe sowie ihre Glättungen.

Die ergänzende Abbildung 30 zur Messreihe I wird in der folgenden Bemerkung näher erläutert.

Wir wählen eine kleine Gleitlänge von 10 zum Vergleich der Glättungsverfahren, und sehen, dass das in den meisten Fällen genügt.

Bemerkungen zu den Abbildungen in Abschnitt 5.2

- Zunächst erkennt man, dass das Abtrennen der Nullläufe von der Messreihe nur bedingt Auswirkung auf die Glättung des relevanten Messbereiches hat.
- Die im vorigen Punkt gemachte Einschränkung ist aus folgenden Gründen notwendig.
 - Bei einem hohen Nulllauf mit im Vergleich dazu schmalen Extrema werden die Randextrema gegebenenfalls weggeglättet, siehe Messreihe **IV**, Abbildungen 36-40. Da aber die Differenz zwischen der Anzahl der Zähne auf der groben Seite der Scheibe einerseits und der feinen Seite andererseits relativ gross ist, wäre der Verlust der Randextrema für die hier betrachtete praktische Fragestellung nicht relevant.
 - Ausschlaggebend für die Entscheidung zur Abtrennung des Nulllaufs ist jedoch der folgende Punkt: In den Praxisdaten ist der Nulllauf nicht konstant, sondern er weist eine hochfrequente Grundschwingung auf, die auf eine Vibration des gesamten Versuchsaufbaus zurückzuführen sein könnte, vergleiche Abbildungen 25, 31, 36 und 41. Diese Grundschwingung kann erst durch eine Erhöhung der Gleitlänge der Glättung aufgefangen werden, und kann somit die Zählung beeinflussen. Vergleiche dazu auch die Diskussion in Abschnitt 3.3.
- Die symmetrisierte exponentielle Glättung (SEA) gemäß (2.8) ist mit Abstand die effektivste Glättungsmethode. Sie liefert schon bei kleiner Gleitlänge sehr gute Ergebnisse und auch bei einer hochfrequenten Störschwingung reicht eine moderate Gleitlänge zur Glättung aus.
- An Messreihe **I** lassen sich die im Text beschriebenen Probleme der Glättungsverfahren im Zusammenhang mit stark zerklüfteten Extrema erkennen, siehe Abbildungen 25-29 und dort speziell das vierte Maximum.
- Der Lösungsvorschlag der Vorglättung mittels eines gleitenden Medians liefert hier die Lösung. Vergleiche dazu in Abbildung 30 die gezoomte Originalmessreihe (links) mit der Medianbildung (rechts), die dann als Ausgangsreihe herangezogen wird.

5.2. Abbildungen zu den Praxisdaten.

ABBILDUNG 25. Messreihe **I**, Originaldaten bzw. bereinigte Originaldaten

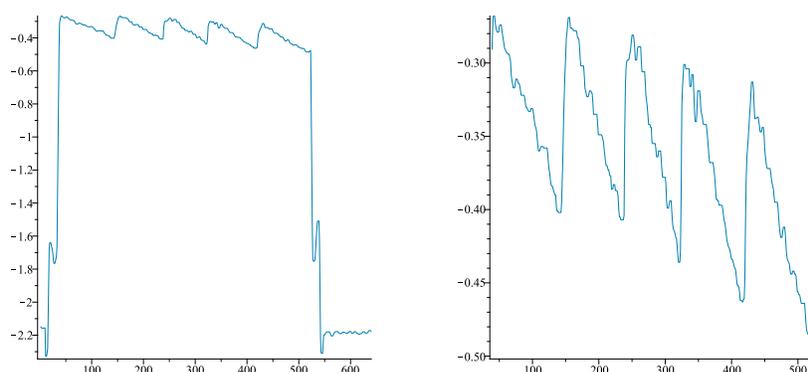


ABBILDUNG 26. Messreihe I, Glättung mit MA, Gleitlänge 10.

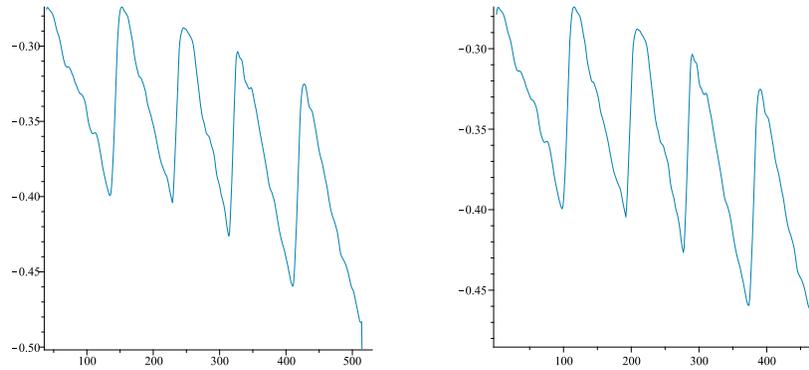


ABBILDUNG 27. Messreihe I, Glättung mit EA, Gleitlänge 10.

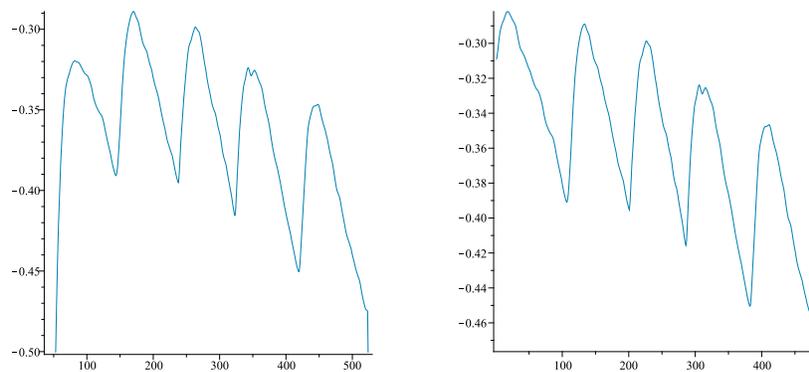


ABBILDUNG 28. Messreihe I, Glättung mit rEA, Gleitlänge 10.

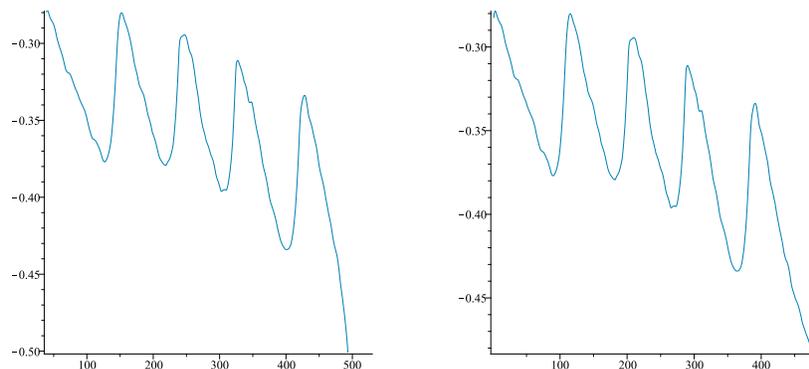


ABBILDUNG 29. **Messreihe I**, Glättung mit SEA, Gleitlänge 10.

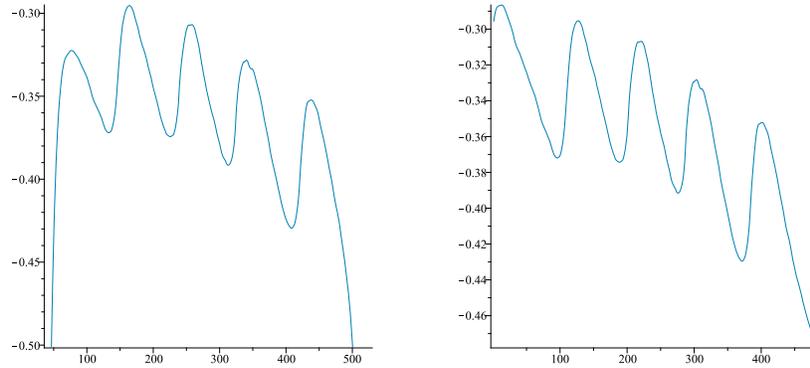


ABBILDUNG 30. **Messreihe I**, Originaldaten und Medianbildung mit Gleitlänge 9.

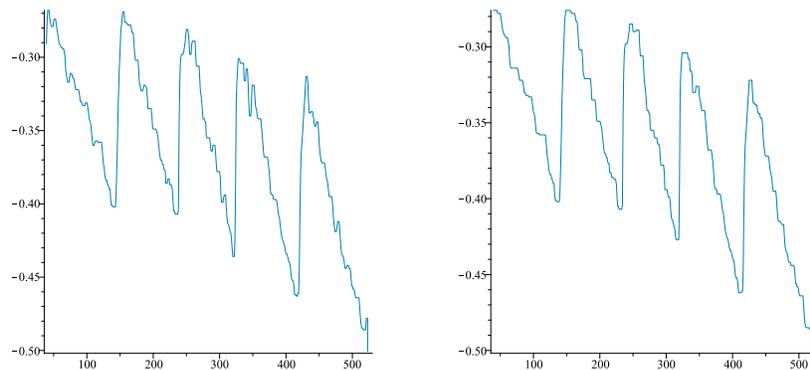


ABBILDUNG 31. **Messreihe II**, Originaldaten bzw. bereinigte Originaldaten

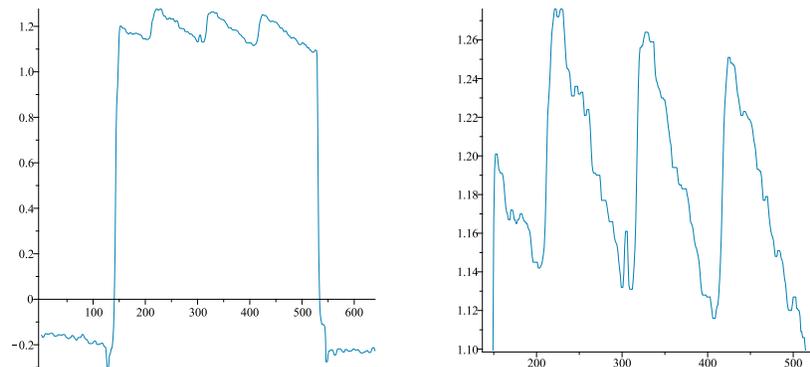


ABBILDUNG 32. Messreihe II, Glättung mit MA, Gleitlänge 10.

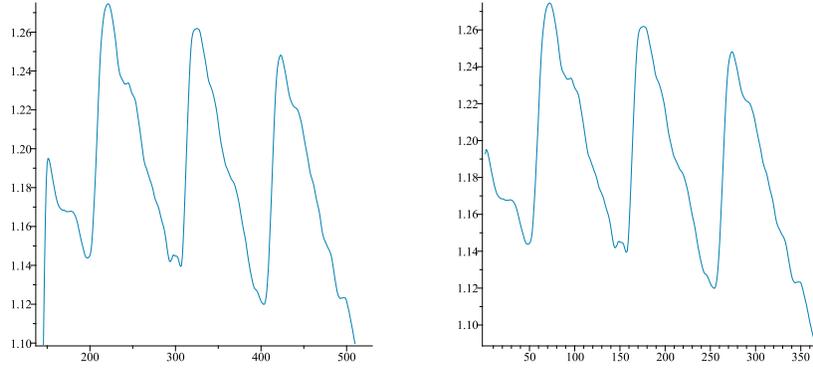


ABBILDUNG 33. Messreihe II, Glättung mit EA, Gleitlänge 10.

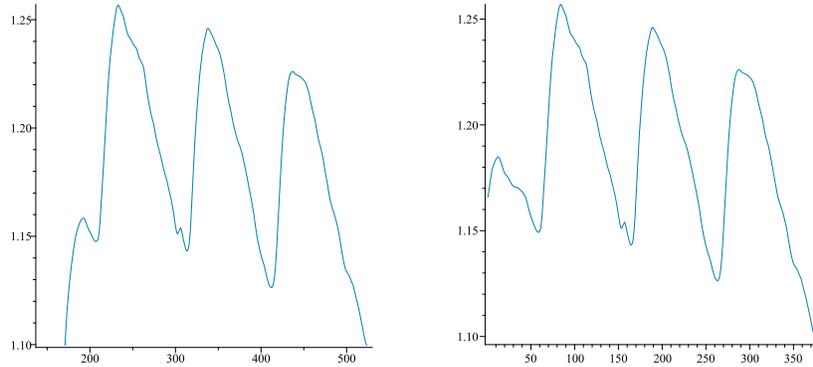


ABBILDUNG 34. Messreihe II, Glättung mit rEA, Gleitlänge 10.

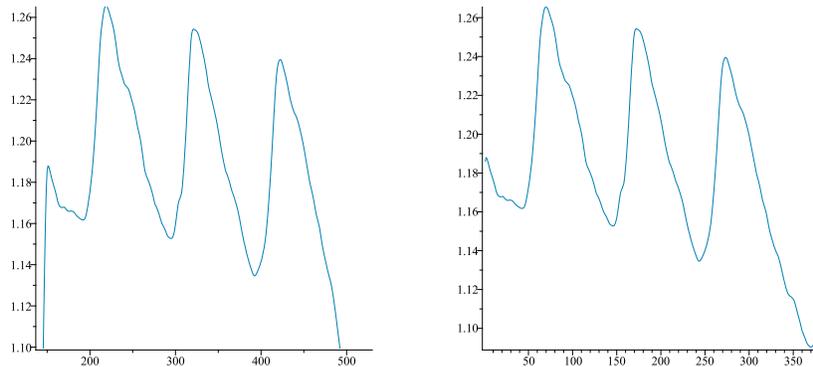


ABBILDUNG 35. Messreihe II, Glättung mit SEA, Gleitlänge 10.

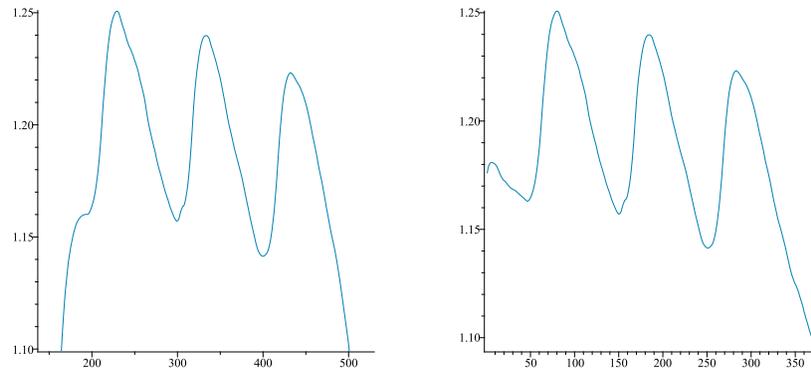


ABBILDUNG 36. Messreihe III, Originaldaten bzw. bereinigte Originaldaten

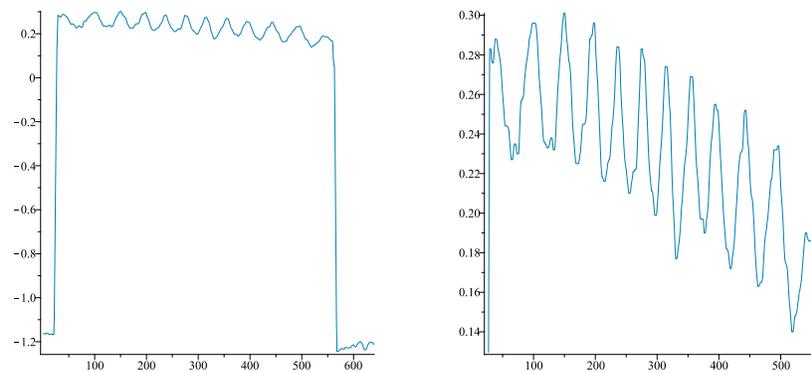


ABBILDUNG 37. Messreihe III, Glättung mit MA, Gleitlänge 10.

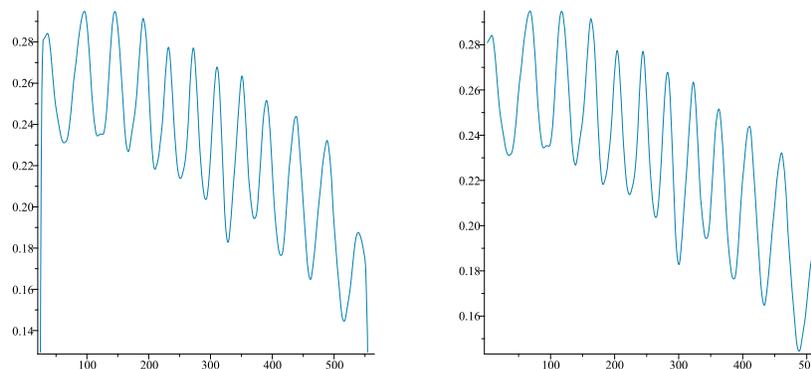


ABBILDUNG 38. Messreihe III, Glättung mit EA, Gleitlänge 10.

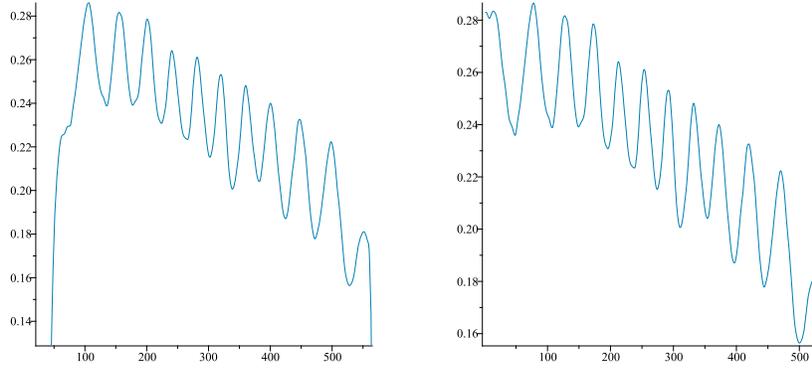


ABBILDUNG 39. Messreihe III, Glättung mit rEA, Gleitlänge 10.

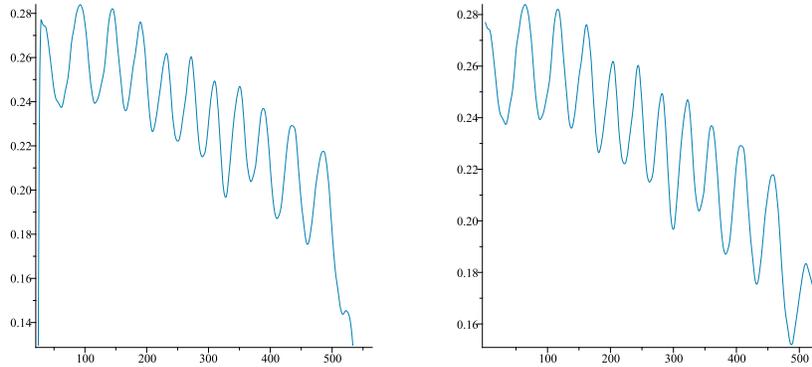


ABBILDUNG 40. Messreihe III, Glättung mit SEA, Gleitlänge 10.

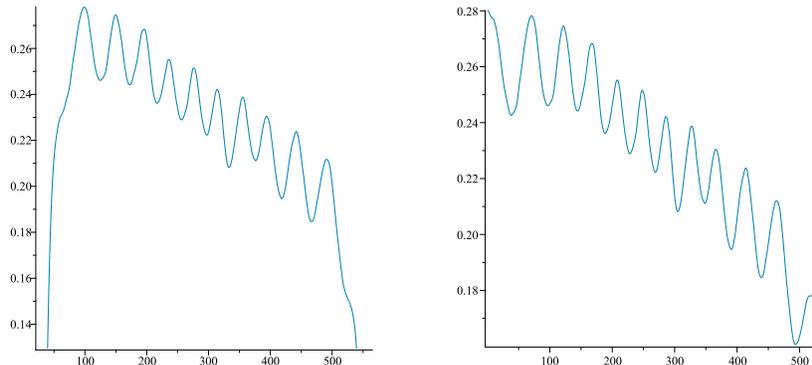


ABBILDUNG 41. Messreihe IV, Originaldaten bzw. bereinigte Originaldaten

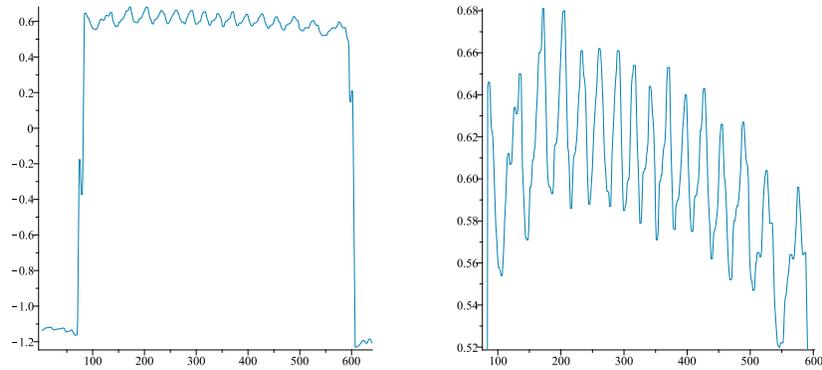


ABBILDUNG 42. Messreihe IV, Glättung mit MA, Gleitlänge 10.

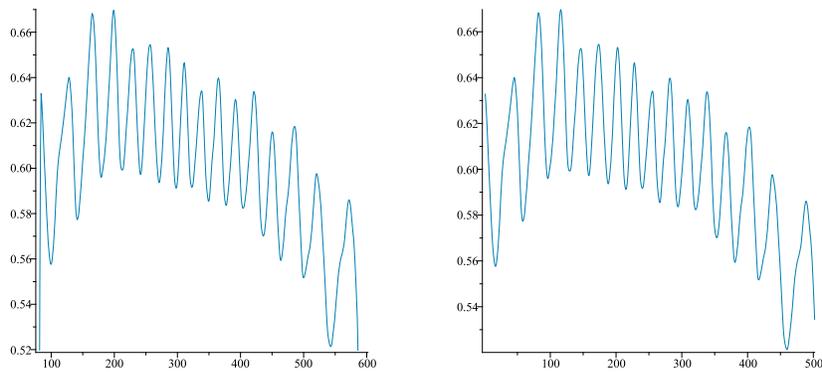


ABBILDUNG 43. Messreihe IV, Glättung mit EA, Gleitlänge 10.

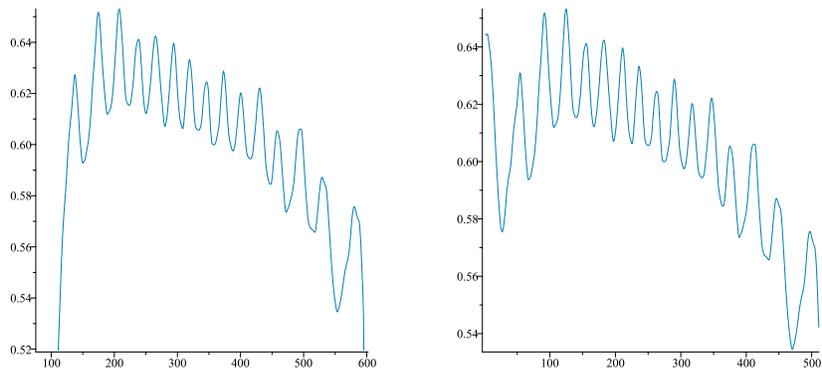


ABBILDUNG 44. Messreihe IV, Glättung mit rEA, Gleitlänge 10.

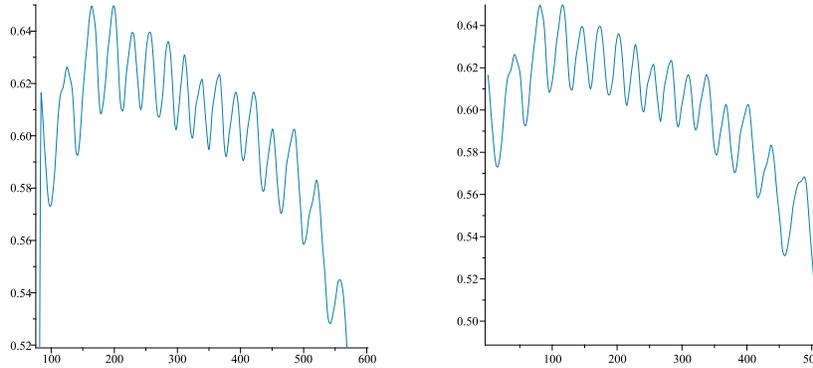
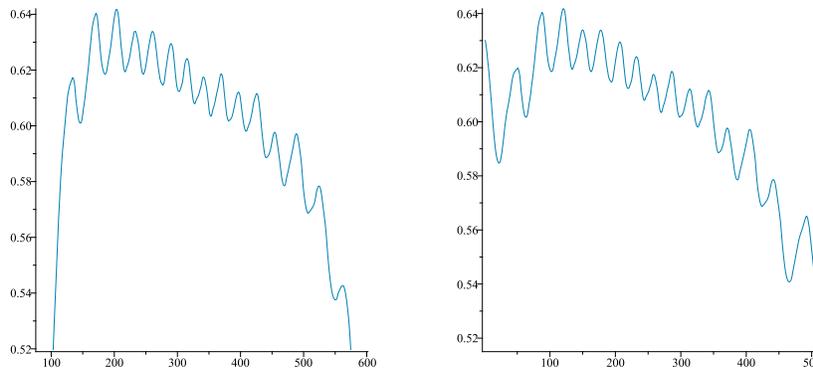


ABBILDUNG 45. Messreihe IV, Glättung mit SEA, Gleitlänge 10.



6. FAZIT

Zur Lösung der in der Einleitung beschriebenen technischen Fragestellung schlagen wir als Glättungsverfahren die in Abschnitt 2.3 definierte symmetrisierte exponentielle Glättung (SEA) vor. Um Ausreißer abzufangen, sollte die Glättung in Verbindung mit einer vorher durchgeführten Anwendung des gleitenden Medians durchgeführt werden. Für beide sollte eine moderate Gleitlänge gewählt werden.

Eine a priori Bestimmung der Gleitlänge ist aufgrund der technischen Umsetzung und der damit willkürlichen Schwankungen unterliegenden Messreihen nicht möglich. Vielmehr beruht seine Wahl auf Erfahrungswerten aus der Auswertung einer hinreichend großen Anzahl von Messreihen. Eine praktische Anwendung dieser empirischen a priori Festlegung der Gleitlänge kann etwa die Aussortierung defekter Bauteile sein. Die Güte der verwendeten Bauteile kann dann wiederum mit Hilfe einer Anpassung der Gleitlänge beeinflusst werden.

Einerseits weist der SEA schon bei geringen Gleitlängen ein sehr gutes Glättungsverhalten auf und andererseits respektiert er die Symmetrie des Datensatzes. Der

reine EA ist aus Gründen der Unsymmetrie nicht geeignet, da zu frühe Artefakte nicht hinreichend geglättet werden können, siehe Abschnitt 3.2, Abbildung 9. Die sehr guten Glättungseigenschaften des EA kommen erst nach einer hinreichend langen Anfangsphase zum Tragen. Trotz der guten Symmetrieeigenschaften ist der MA nicht in dem gleichen Maße geeignet wie der SEA, da man für vergleichbare Glättung eine höhere Gleitlänge benötigt und somit eine starke Verkürzung des Datensatzes in Kauf nehmen muss.

Das Vorschalten des gleitenden Medians fängt zum einen Ausreißer ein, zum anderen schafft er eine Vorabglättung eventuell vorhandener zerklüfteter Extrema. Mit dieser Form der Entartung haben alle Glättungsverfahren – wenn auch in unterschiedlichem Maße – Probleme bei niedrigen Gleitlängen, vergleiche Abschnitt 3.2 und ebenfalls die Messreihe II, Abbildungen 25-30.

Zur Vorbereitung der Praxisdaten ist eine Abtrennung eines gegebenenfalls vorhandenen Nulllaufs angezeigt, da dieser die zur Glättung notwendige Gleitlänge unnötig erhöhen kann, vergleiche Abschnitt 5.1.

LITERATUR

- [1] Gronwall, Thomas Hakon: Über die Gibbssche Erscheinung und die trigonometrischen Summen $\sum_{k=1}^n \frac{1}{k} \sin kx$. *Math. Ann.* **72** (1912), no. 2, 228-243.
- [2] Gumm, H.-P. und Sommer, M.: Einführung in die Informatik. Oldenbourg Verlag, 4. Aufl., 2000.
- [3] Hewitt Edwin und Hewitt, Robert E: The Gibbs-Wilbraham Phenomenon: An Episode in Fourier Analysis. *Arch. Hist. Exact Sci.* **21** (1979-1980), no. 2, 129-160.
- [4] Howell, Kenneth B.: *Principles of Fourier Analysis*. Chapman & Hall/CRC, 2001.
- [5] Klinker, Frank: Exponential Moving Average versus Moving Exponential Average. *Math. Semesterber.* **58** (2011), no. 1, 97-107.
- [6] Kreiß, Jens-Peter und Neuhaus, Georg: *Einführung in die Zeitreihenanalyse*. Springer, 2006.

Preprints ab 2010/03

- 2012-05 **Frank Klinker and Günter Skoruppa**
Ein optimiertes Glättungsverfahren motiviert durch eine technische Fragestellung
- 2012-04 **Patrick Henning, Mario Ohlberger, and Ben Schweizer**
Homogenization of the degenerate two-phase flow equations
- 2012-03 **Andreas Rätz**
A new diffuse-interface model for step flow in epitaxial growth
- 2012-02 **Andreas Rätz and Ben Schweizer**
Hysteresis models and gravity fingering in porous media
- 2012-01 **Wilfried Hazod**
Intrinsic topologies on H-contraction groups with applications to semistability
- 2011-14 **Guy Bouchitté and Ben Schweizer**
Plasmonic waves allow perfect transmission through sub-wavelength metallic gratings
- 2011-13 **Waldemar Grundmann**
Moment functions and Central Limit Theorem for Jacobi hypergroups on $[0, \infty[$
- 2011-12 **J. Koch, A. Rätz, and B. Schweizer**
Two-phase flow equations with a dynamic capillary pressure
- 2011-11 **Michael Voit**
Central limit theorems for hyperbolic spaces and Jacobi processes on $[0, \infty[$
- 2011-10 **Ben Schweizer**
The Richards equation with hysteresis and degenerate capillary pressure
- 2011-09 **Andreas Rätz and Matthias Röger**
Turing instabilities in a mathematical model for signaling networks
- 2011-08 **Matthias Röger and Reiner Schätzle**
Control of the isoperimetric deficit by the Willmore deficit
- 2011-07 **Frank Klinker**
Generalized duality for k-forms
- 2011-06 **Sebastian Aland, Andreas Rätz, Matthias Röger, and Axel Voigt**
Buckling instability of viral capsides - a continuum approach
- 2011-05 **Wilfried Hazod**
The concentration function problem for locally compact groups revisited: Non-dissipating space-time random walks, τ -decomposable laws and their continuous time analogues
- 2011-04 **Wilfried Hazod, Katrin Kosfeld**
Multiple decomposability of probabilities on contractible locally compact groups
- 2011-03 **Alexandra Monzner* and Frol Zapolsky†**
A comparison of symplectic homogenization and Calabi quasi-states

- 2011-02 **Stefan Jäschke, Karl Friedrich Siburg and Pavel A. Stoimenov**
Modelling dependence of extreme events in energy markets using tail copulas
- 2011-01 **Ben Schweizer and Marco Veneroni**
The needle problem approach to non-periodic homogenization
- 2010-16 **Sebastian Engelke and Jeannette H.C. Woerner**
A unifying approach to fractional Lévy processes
- 2010-15 **Alexander Schnurr and Jeannette H.C. Woerner**
Well-balanced Lévy Driven Ornstein-Uhlenbeck Processes
- 2010-14 **Lorenz J. Schwachhöfer**
On the Solvability of the Transvection group of Extrinsic Symplectic Symmetric Spaces
- 2010-13 **Marco Veneroni**
Stochastic homogenization of subdifferential inclusions via scale integration
- 2010-12 **Agnes Lamacz, Andreas Rätz, and Ben Schweizer**
A well-posed hysteresis model for flows in porous media and applications to fingering effects
- 2010-11 **Luca Lussardi and Annibale Magni**
 Γ -limits of convolution functionals
- 2010-10 **Patrick W. Dondl, Luca Mugnai, and Matthias Röger**
Confined elastic curves
- 2010-09 **Matthias Röger and Hendrik Weber**
Tightness for a stochastic Allen–Cahn equation
- 2010-08 **Michael Voit**
Multidimensional Heisenberg convolutions and product formulas for multivariate Laguerre polynomials
- 2010-07 **Ben Schweizer**
Instability of gravity wetting fronts for Richards equations with hysteresis
- 2010-06 **Lorenz J. Schwachhöfer**
Holonomy Groups and Algebras
- 2010-05 **Agnes Lamacz**
Dispersive effective models for waves in heterogeneous media
- 2010-04 **Ben Schweizer and Marco Veneroni**
Periodic homogenization of Prandtl-Reuss plasticity equations in arbitrary dimension
- 2010-03 **Holger Dette and Karl Friedrich Siburg and Pavel A. Stoimenov**
A copula-based nonparametric measure of regression dependence