
Extremstellenverlust als Gütekriterium in der nichtparametrischen Regression

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund



Der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von
Henrike Weinert

Dortmund 2009

1. Gutachter: Prof. Dr. Ursula Gather

2. Gutachter: PD Dr. Sonja Kuhnt

Tag der mündlichen Prüfung: 30. September 2009

INHALTSVERZEICHNIS

1	Einleitung	1
2	Nichtparametrische Regression	5
2.1	Regressionsmodell und Testbed-Funktionen	5
2.2	Methoden der Nichtparametrischen Regression	9
2.2.1	Waveletmethoden (WH und WN6)	9
2.2.2	Unbalanced Haar-Methode (UH)	11
2.2.3	Minimum Description Length Denoising (MDL)	13
2.2.4	Lokaler Plug-in Kernschätzer (LPI)	15
2.2.5	Adaptive Weights Smoothing (AWSk und AWSg)	16
2.2.6	Taut String (TS)	17
2.2.7	Glatte Taut String-Methode (STS)	19
2.2.8	Potts-Schätzer (POT)	20
2.3	Gütekriterien	21
3	Extremstellenverlust	23
3.1	Definitionen und Notation	25
3.2	Herleitung des Extremstellenverlustes	29
3.2.1	Gewünschte Eigenschaften eines Extremstellenverlustes	30
3.2.2	Herleitung verschiedener Verlustfunktionen	31

3.2.3	Vergleich und Auswahl der neuen Verlustfunktion	43
4	Toleranzen für Extremstellen	49
5	Extremwertverlust in Y -Richtung	61
6	Alternative Herangehensweise: Der Softloss	73
7	Simulationsstudie	81
7.1	Andere Gütemaße	81
7.2	Ergebnisse der Simulationsstudie	84
8	Zusammenfassung und Ausblick	125
	Abkürzungsverzeichnis	128
	Literaturverzeichnis	129

KAPITEL 1

EINLEITUNG

Die nichtparametrische Regression ist ein bedeutendes Teilgebiet der Statistik, das vielfältige Verfahren für verschiedene Anwendungsbereiche liefert, in denen keine parametrischen Annahmen getroffen werden können. Aufgabe der Regression ist es, eine Funktion aus verrauschten Daten zu schätzen, die den Zusammenhang zwischen einer Einflussvariablen und einer Zielvariablen wiedergibt. Die geschätzte Funktion kann für verschiedene Anwendungen genutzt werden. Beispielsweise kann das Interesse in Vorhersagen oder in bestimmten Merkmalen der Regressionsfunktion, wie Sprungstellen oder Extremwerten, liegen (vgl. Härdle 1990, Marron und Tsybakov 1995, Chaudhuri und Marron, 2000, Kovac 2007, Boysen et al. 2009).

Die Qualität einer nichtparametrischen Regressionsmethode muss sowohl in der Entwicklung als auch in der späteren Validierung anhand von Gütekriterien beurteilt werden. Meist dienen L_p -Normen oder daraus abgeleitete Kriterien wie zum Beispiel der Mittlere Quadratische Fehler als Gütemaße. Neben theoretischen Betrachtungen wie unter anderem zum asymptotischen Verhalten einer Methode, werden häufig Simulationsstudien durchgeführt. Insbesondere zum Vergleich verschiedener Regressionsmethoden ist die Anwendung auf simulierte Datensätze sinnvoll. Dafür wird eine bekannte Signalfunktion mit Rauschen überlagert. Auf die so simulierte Daten wird die zu untersuchende Schätzmethode angewendet und das Ergebnis mit der wahren Funktion verglichen. Über die einzelnen Ergebnisse aus den Simulationsläufen wird wiederum gemittelt, um die Güte der verschiedenen Methoden zu vergleichen.

Die Auswahl des Gütekriteriums erfolgt dabei überwiegend ohne Berücksichtigung des Ziels der Regression. In vielen Fällen wird die L_2 -Norm verwendet (vgl. z.B. Wassermann 2006, Donoho und Johnstone 1994, Herrmann 1997a, Rissanen 2000), obwohl

diese nur für bestimmte Ziele wie etwa Vorhersagen als sinnvolles Kriterium dienen kann.

Zusätzlich besteht eine Diskrepanz zwischen dem, was die L_p -Normen messen und der visuellen Beurteilung durch einen Betrachter. Während die Normen nur den vertikalen Abstand von Signal- und Schätzfunktion verwenden, nimmt ein Betrachter automatisch ein Gesamtbild beider Funktionen wahr. Beispielsweise kann er falsch geschätzte Extremwerte visuell sofort erfassen und sie zur Bewertung der Schätzfunktion heranziehen. Insbesondere für Anwendungen, bei denen das korrekte Identifizieren von Extremwerten von Bedeutung ist, führen klassische Gütekriterien wie die L_p -Normen zu Fehleinschätzungen bezüglich der Eignung von Regressionsverfahren.

Ziel dieser Arbeit ist die Entwicklung eines Gütekriteriums für die Bewertung der Anpassung einer Schätzfunktion an eine Signalfunktion bezogen auf die korrekte Identifizierung von Extremwerten. Das Fehlverhalten einer Schätzfunktion bezüglich der Maximal- und Minimalstellen kann sich in zu vielen, zu wenigen oder an der falschen Stelle geschätzten Extremstellen zeigen. Diese verschiedenen möglichen Fehler werden in einem normierten Verlust zusammengefasst. Damit formalisiert der neue *Extremstellenverlust* die visuelle Beurteilung durch einen Betrachter und ist geeignet, die Güte einer Methode bezüglich ihrer Extremstellen in einem Verlustwert zu erfassen.

Des weiteren ist von Interesse, ob eine Regressionsmethode nicht nur die Extremstellen am richtigen Ort sondern auch die Extremwerte mit der richtigen Höhe schätzt. Daher wird ein weiterer Verlust definiert, der die Güte bezüglich der Richtigkeit geschätzter Extremwerte beurteilt.

Diese Arbeit gliedert sich wie folgt. In Kapitel 2 wird das nichtparametrische Regressionsmodell vorgestellt. Es werden zehn Signalfunktionen aus der Literatur eingeführt. Daneben erfolgt die Vorstellung von zehn Methoden der nichtparametrischen Regression und gängigen Gütekriterien.

In Kapitel 3 werden verschiedene Möglichkeiten eines Verlustes bezogen auf die Extremstellen diskutiert. Dazu wird zunächst mit Hilfe einer Toleranz definiert, ob Extremstellen der Schätzfunktion an den richtigen Stellen liegen. Damit lassen sich die möglichen Fehler in einer Schätzfunktion beschreiben. Es werden Anforderungen an eine Verlustfunktion, die das Fehlverhalten zusammenfasst, aufgestellt. Insgesamt können vier mögliche Verlustfunktionen definiert und verglichen werden. Darauf baut die Ent-

wicklung des eigentlichen Extremstellenverlustes auf. Dieser Extremstellenverlust ist abhängig von einer Toleranz für die Bestimmung *richtiger* Extremstellen. In Kapitel 4 werden die Auswirkungen verschieden großer Toleranzen auf den Extremstellenverlust untersucht.

Neben der Beurteilung der Güte einer Methode anhand der Extremstellen kann auch die Höhe eines Extremwerts von Bedeutung sein. In Kapitel 5 werden daher Verluste diskutiert, die als Zusatzinformation zum Extremstellenverlust verwendet werden können und die die Güte einer Schätzung bezüglich der richtigen Höhe der Extremwerte beurteilen.

Eine alternative Herangehensweise an die Herausforderung der Gütemessung bezüglich Extremwerten wird in Kapitel 6 vorgeschlagen. Ausgehend von der klassischen Gütemessung mittels L_p -Normen werden die Signal- und Schätzfunktion auf ihre jeweiligen Extremwerte reduziert. Nach der Faltung mit einem Gaußkern lassen sich beide Funktionen anhand der L_1 -Norm vergleichen.

Schließlich werden in Kapitel 7 die entwickelten Gütekriterien in einer großen Simulationsstudie auf ihre Eignung überprüft, einen Vergleich von Regressionsmethoden zur Erkennung von Extremwerten zu erzielen. Zudem werden die neuen Kriterien untereinander und mit bestehenden Gütekriterien verglichen.

Die Arbeit wird mit einer Zusammenfassung und einem Ausblick in Kapitel 8 abgeschlossen.

NICHTPARAMETRISCHE REGRESSION

In diesem Kapitel werden notwendige Begriffe aus der nichtparametrischen Regression erläutert. Unterkapitel 2.1 stellt das nichtparametrische Regressionsmodell und in dieser Arbeit verwendete Testbed-Funktionen vor. Anschließend werden in Kapitel 2.2 Methoden der nichtparametrischen Regression und in Kapitel 2.3 Gütekriterien eingeführt.

2.1 Regressionsmodell und Testbed-Funktionen

In der nichtparametrischen Regression werden Daten der Form $(t_i, y(t_i)) \in [0; 1] \times \mathbb{R}$, $i = 1, \dots, n$, betrachtet. Dabei sind t_i sogenannte *Designpunkte* und $y(t_i)$ zugehörige Beobachtungen. Das *nichtparametrische Regressionsmodell* ist gegeben durch

$$y(t_i) = f(t_i) + \varepsilon(t_i), i = 1, \dots, n. \quad (2.1)$$

Die Daten setzen sich aus einer zugrunde liegenden Funktion f und einem Rauschen ε zusammen. Ziel der nichtparametrischen Regression ist es, mit Hilfe der Daten eine Schätzung \hat{f} für f zu finden. In Simulationen kommen Beispieldatensätze zur Anwendung, für die die wahre zugrunde liegende Funktion f bekannt ist. Solche Funktionen heißen *Testbed-Funktion* oder *Signalfunktion*. Für einige der nachfolgend in Kapitel 2.2 vorgestellten Regressionsmethoden sind weitere Annahmen über die Designpunkte $t_i, i = 1, \dots, n$, die Stichprobengröße und das Rauschen notwendig. Daher werden hier äquidistante Designpunkte angenommen, d.h. $t_i = \frac{i}{n}, i = 1, \dots, n$. Das Rauschen wird als normalverteilt mit Mittelwert 0 und Varianz σ^2 angenommen.

Für nachfolgende Untersuchungen werden 10 verschiedene Testbed-Funktionen so ausgewählt, dass möglichst viele Eigenschaften von realen wahren Funktionen abgedeckt werden. Drei von ihnen sind stückweise konstant, zwei weitere enthalten Sprungstellen. Im Einzelnen werden vier Funktionen von Donoho und Johnstone (1994) verwendet: Doppler, Bumps, Heavisine und Blocks. Aus Davies, Gather und Weinert (2008) werden die Sine-Funktion, die äußerst viele Extremstellen hat, und eine Konstante, die keine Extremstellen hat, aufgenommen. Dazu kommen noch drei Funktionen von Marron et. al. (1998): Step, Blip und Wave. Als zehnte Funktion wird Spikes aus Cai (1999) aufgenommen. In den Tabellen 2.1 und 2.2 sind die Funktionen angegeben und die zugehörigen Graphen dargestellt, erstellt mit $n = 1024$ Designpunkten.

Die Funktionen werden so normiert, dass gilt:

$$SNR = \left[\int |f - \int f|^2 \right]^{\frac{1}{2}} = 7. \quad (2.2)$$

Dies entspricht dem häufig verwendeten Signal-To-Noise-Ratio von 7, wenn zu den Funktionen wie oben beschrieben standardnormalverteiltes Rauschen addiert wird.

Tabelle 2.3 zeigt für die 10 Testbed-Funktionen die Anzahl Extremstellen, wie sie im nächsten Kapitel in Definition 3.1 gegeben sind. Da die Funktionen an diskreten, äquidistanten Designpunkten ausgewertet werden, kann die Anzahl Designpunkte Einfluss auf die Anzahl der Extremwerte haben. Von den hier verwendeten Testbed-Funktionen betrifft das die Doppler- und die Sine-Funktion. Für $n = 256$ weist die Doppler-Funktion 27 Extremwerte auf, für $n = 1024$ sind es 39. Die Sine-Funktion zeigt 79 Extremwerte für $n = 256$ und 80 für $n = 1024$.

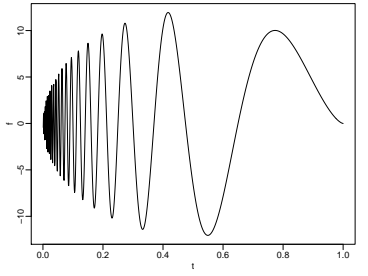
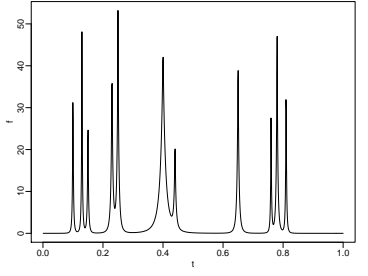
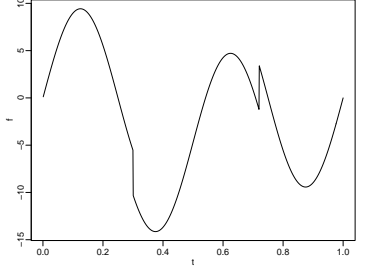
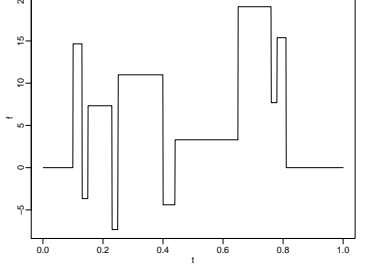
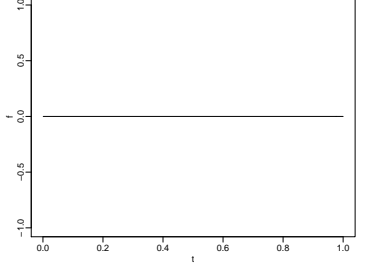
<p>Doppler:</p> $f(t) = (t(1-t))^{\frac{1}{2}} \sin\left(\frac{2\pi(1+\delta)t}{(t+\delta)}\right)$ <p>mit $\delta = 0,05$</p>	
<p>Bumps:</p> $f(t) = \sum h_j K\left(\frac{t-t_j}{w_j}\right), \quad K(t) = (1+ t)^{-4}$ <p>mit</p> $(t_j) = (10, 13, 15, 23, 25, 40, 44, 65, 76, 78, 81)/100,$ $(h_j) = (40, 50, 30, 40, 50, 42, 21, 43, 31, 51, 42)/10 \text{ und}$ $(w_j) = (5, 5, 6, 10, 10, 30, 10, 10, 5, 8, 5)/1000$	
<p>Heavisine:</p> $f(t) = 4 \sin 4\pi t - \operatorname{sgn}(t - 0,3) - \operatorname{sgn}(0,72 - t)$	
<p>Blocks:</p> $f(t) = \sum h_j K(t - t_j), \quad K(t) = \frac{1+\operatorname{sgn}(t)}{2}$ <p>mit</p> $(t_j) = (10, 13, 15, 23, 25, 40, 44, 65, 76, 78, 81)/100 \text{ und}$ $(h_j) = (40, -50, 30, -40, 50, -42, 21, 43, -31, 21, -42)/10$	
<p>Konstante:</p> $f(t) = 0$	

Tabelle 2.1: Testbed-Funktionen

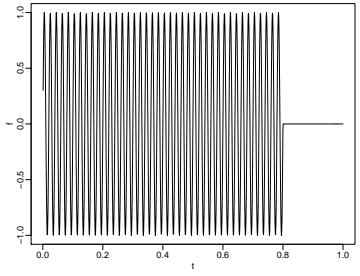
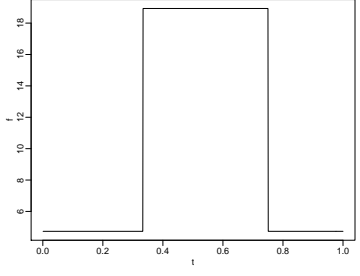
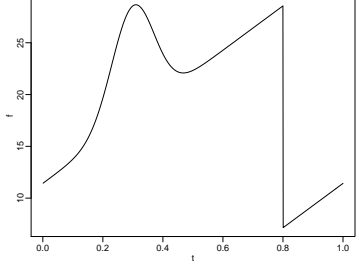
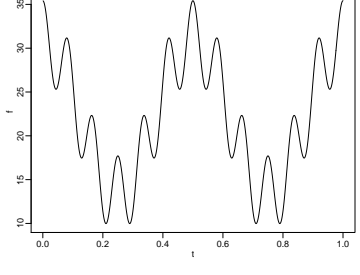
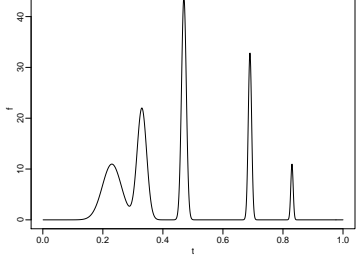
<p>Sine:</p> $f(t) = \begin{cases} \sin(100\pi t) & 0 \leq t \leq 0,8 \\ 0 & \text{sonst} \end{cases}$	
<p>Step:</p> $f(t) = 0,2 + (0,6)\mathbf{1}_{(\frac{1}{3};0,75)}(t)$	
<p>Blip:</p> $f(t) = (0,32 + 0,6t + 0,3e^{-100(t-0,3)^2})\mathbf{1}_{[0;0,8]}(t) + (-0,28 + 0,6t + 0,3e^{-100(t-1,3)^2})\mathbf{1}_{(0,8;1]}(t)$	
<p>Wave:</p> $f(t) = 0,5 + 0,2\cos(4\pi t) + 0,1\cos(24\pi t)$	
<p>Spikes:</p> $f(t) = 15.6676[e^{-500(t-0,23)^2} + 2e^{-2000(t-0,33)^2} + 4e^{-8000(t-0,47)^2} + 3e^{-16000(t-0,69)^2} + e^{-32000(t-0,83)^2}]$	

Tabelle 2.2: Testbed-Funktionen

Funktion	Anzahl Extremstellen		Funktion	Anzahl Extremstellen	
	$n=256$	$n=1024$		$n=256$	$n=1024$
Doppler	27	39	Sine	79	80
Bumps		21	Step		1
Heavisine		6	Blip		4
Blocks		9	Wave		23
Konstante		0	Spikes		9

Tabelle 2.3: Anzahl Extremstellen für alle Testbed-Funktionen

2.2 Methoden der Nichtparametrischen Regression

In dieser Arbeit werden insgesamt 10 aus der Literatur bekannte Methoden der nichtparametrischen Regression verwendet. Alle ausgewählten Methoden sind für Modell (2.1) geeignet. Einige Methoden basieren exakt darauf, während bei anderen Variationen im Modell erlaubt sind. Beispielsweise reicht für manche Methoden Sub-Gaußsches Rauschen aus, oder die Designpunkte können variabel sein.

2.2.1 Waveletmethoden (*WH* und *WN6*)

Ein Wavelet ist definiert durch

$$\Psi_{j,k}(t) = 2^{\frac{j}{2}} \Psi(2^j t - k), \quad j, k \geq 0.$$

Dabei ist Ψ das sogenannte Motherwavelet. Dieses wird so gewählt, dass die $\Psi_{j,k}$ eine Orthonormalbasis des $L^2(\mathbb{R})$ bilden. Daher kann jede Funktion f aus dem Raum $L^2(\mathbb{R})$ als Summe über verschobene und umskalierte Versionen des Motherwavelets dargestellt werden:

$$f(t) = \sum_{j,k} \bar{w}_{j,k} \Psi_{j,k}(t), \quad \bar{w}_{j,k} = \int_{-\infty}^{\infty} f(t) \Psi_{j,k}(t) dt.$$

Die Werte $\bar{w}_{j,k}$ werden Waveletkoeffizienten genannt (Daubechies 1992).

Um eine Schätzung einer Signalfunktion zu erhalten, muss $n = 2^{J+1}$ gelten, der Stichprobenumfang muss also eine Potenz von 2 sein. Zunächst wird eine Wavelettransformationsmatrix \mathcal{W} benutzt, um einen Vektor w empirischer Waveletkoeffizienten zu erhalten: $w = \mathcal{W}y$ (vgl. Daubechies 1992, Donoho und Johnstone 1994, Nason und Silverman 1994). Mit den $n = 2^{J+1}$ Elementen von w (bezeichnet mit $w_{j,k}$, $j = 0, \dots, J$; $k = 1, \dots, 2^j - 1$ und $w_{-1,0}$) kann nun die Waveletapproximation gebildet werden:

$$\hat{f}_\Delta(t_i) = \sum_{(j,k) \in \Delta} w_{j,k} \Psi_{j,k}(t_i), \quad i = 1, \dots, n, \quad (2.3)$$

wobei $\Psi_{-1,0}(t_i) \equiv 1$. Δ ist die Menge der verwendeten Waveletkoeffizienten. Bei Verwendung aller Koeffizienten werden die Daten rekonstruiert. Üblicherweise ist Δ eine Teilmenge der Paare (j, k) . Je nachdem, wie Δ gewählt wird, werden viele oder wenige Waveletkoeffizienten zur Schätzung der Signalfunktion verwendet. Die optimale Teilmenge Δ kann nach Donoho und Johnstone (1994) so geschätzt werden, dass nur solche Waveletkoeffizienten $w_{j,k}$ in (2.3) verwendet werden, die eine bestimmte Schranke überschreiten. Dahinter steht die Annahme, dass große Koeffizienten durch die Signalfunktion und kleine durch das Rauschen zustande kommen. Donoho und Johnstone (1994) schlagen dazu folgende sogenannte *hard thresholding* Regel vor: Δ in (2.3) beinhaltet diejenigen Koeffizienten, die

$$|w_{j,k}| > \hat{\sigma} \cdot \sqrt{2 \log(n)} \quad (2.4)$$

erfüllen. $\hat{\sigma}$ ist dabei die *mediane absolute Abweichung (MAD)* der Waveletkoeffizienten. Die zugehörige Schätzung ist der *Wavelet Hard Thresholding Estimator*. Die Zusammenfassung des Vorgehens lautet:

- (i) Wavelettransformation der Daten,
- (ii) Hard Thresholding der Waveletkoeffizienten,
- (iii) Rücktransformation mit den übrigen Waveletkoeffizienten.

Die Glattheit dieses Schätzers hängt vom verwendeten Motherwavelet und damit der verwendeten Waveletbasis ab. In Abbildung 2.2.1 sind die in dieser Arbeit verwendeten Motherwavelets dargestellt. Das Haarwavelet ist eine stückweise konstante Funktion

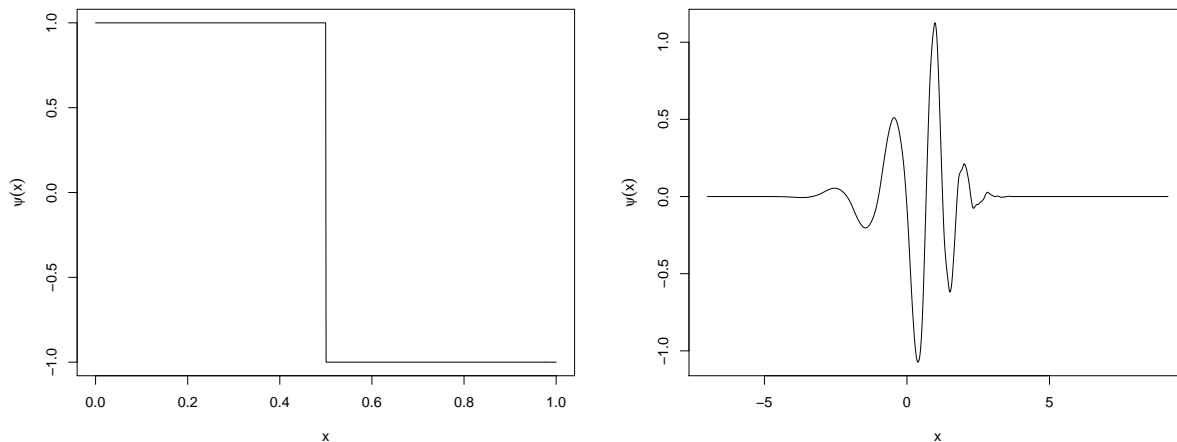


Abbildung 2.1: Verwendete Motherwavelets, links: Haar Wavelet, rechts: Daubechies Extremal Phase Family mit $N = 6$.

und führt somit auch zu stückweise konstanten Schätzern, hier mit WH abgekürzt. Das zweite Wavelet stammt aus Daubechies *Extremal Phase Family* mit $N = 6$ (vgl. Daubechies 1992) und führt zu glatteren Schätzfunktionen. Die Methode wird mit WN6 bezeichnet.

Details finden sich in Donoho und Johnstone (1994), Software ist im R-Package `wave-thresh` zu finden (Nason, Kovac und Maechler 2008, vgl. auch Nason und Silverman 1994).

2.2.2 *Unbalanced Haar-Methode (UH)*

Die *Unbalanced Haar-Methode* nach Fryzlewicz (2007) ist den Waveletmethoden sehr ähnlich. Anders als bei einer Wavelettransformation der Daten mit dem Haarwavelet wird hier die Basis so gebildet, dass die Sprünge in der Basisfunktion nicht immer in der Mitte des betrachteten Intervalls liegen, sondern unsymmetrisch an beliebiger Stelle liegen können. Diese Sprünge werden datengesteuert gesetzt. Die Basis, bestehend aus UH-Vektoren, wird wie folgt gebildet.

(i) Konstruktion eines Vektors

$$\psi^{0,1} = \begin{cases} \text{konstant und positiv} & i = 1, \dots, b^{0,1} \\ \text{konstant und negativ} & i = b^{0,1} + 1, \dots, n \end{cases},$$

wobei sich die Elemente von $\psi^{0,1}$ zu 0 addieren und die quadrierten Elemente zu 1. Der Wert $b^{0,1}$ gibt die erste *Sprungstelle* an.

(ii) Iteratives Fortsetzen dieser Konstruktionsweise auf den zwei entstandenen Teilen, wobei Vektoren $\psi^{j,k}$ mit Sprungstellen $b^{j,k}$ entstehen. Die Rekursion wird solange fortgesetzt, wie es möglich ist, Sprungstellen zu setzen.

Der Wert $b^{0,1}$ wird dabei wie folgt bestimmt: der *UH mother vector* $\psi_{s,b,e}(l)$ (für start, breakpoint und end, vgl. Fryzlewicz 2007) ist definiert als

$$\psi_{s,b,e} = \left\{ \frac{1}{b-s+1} - \frac{1}{e-s+1} \right\}^{\frac{1}{2}} \mathbf{1}_{(s \leq l \leq b)} - \left\{ \frac{1}{e-b} - \frac{1}{e-s+1} \right\}^{\frac{1}{2}} \mathbf{1}_{(b+1 \leq l \leq e)}.$$

Die erste Sprungstelle wird dann so gewählt, dass das Skalarprodukt zwischen dem Datenvektor und dem UH mother vector maximal wird: $b^{0,1} = \operatorname{argmax}_b |\langle \mathbf{y}, \psi_{1,b,n} \rangle|$. Die weiteren Sprungstellen werden in gleicher Weise gewählt.

Die so bestimmten Vektoren $\psi^{j,k}$ bilden eine Orthonormalbasis des \mathbb{R}^n . Danach entspricht das Vorgehen der UH-Methode dem Hard Thresholding:

(i) Diskrete Unbalanced-Haar-Transformation (DUHT) der Daten:

$$\text{DUHT}(\mathbf{y})^{j,k} = \langle \mathbf{y}, \psi^{j,k} \rangle$$

(ii) Hard-Thresholding der Koeffizienten $\text{DUHT}(\mathbf{y})^{j,k}$ mit Schranke $\hat{\sigma} \cdot \sqrt{2 \log(n)}$.

(iii) Rücktransformation mit den übrigen Koeffizienten.

Genauere Angaben zur Methode stehen in Fryzlewicz (2007), der Algorithmus ist im R-Package `unbalhaar` (Fryzlewicz 2006) implementiert.

2.2.3 Minimum Description Length Denoising (MDL)

Die Theorie des Minimum Description Length Denoising nach Rissanen (2000) basiert darauf, dass ein Codeschema mit einem Wahrscheinlichkeitsmodell korrespondiert. Daher kann der Vergleich von Wahrscheinlichkeitsmodellen über die Länge des Codes erfolgen, der zur Dekodierung der Daten benötigt wird. Enthalten Daten Regelmäßigkeiten, so wird ein gutes Modell diese erkennen und nutzen, um die Codelänge zu reduzieren. Beispielsweise lässt sich eine regelmäßige Folge $0, 0, 1, 1, 0, 0, 1, 1, \dots$ einfach beschreiben, während eine Folge zufällig aufeinanderfolgender Einsen und Nullen nur durch einen Code beschrieben werden kann, der genauso lang ist wie die Folge selbst. Das (einfache) MDL-Prinzip favorisiert dasjenige Modell (aus einer Menge von Kandidatenmodellen), das die kürzeste Dekodierung der Daten ermöglicht (Rissanen 1989). In diesen einfachen Ansatz geht die Komplexität des Modells jedoch nicht mit ein, die eine zusätzliche Dekodierung erfordert.

Dies kann als Spezialfall eines Variablenselektionsproblems in der multiplen linearen Regression betrachtet werden. Sei \mathcal{Z} eine gegebene $n \times n$ -Designmatrix für die lineare Regression. Jede Teilmenge $\gamma = \{h_1, \dots, h_k\}$ der Spalten von \mathcal{Z} definiert eine $n \times k$ -Matrix \mathbf{Z}_γ . Das zugehörige lineare Regressionsmodell lautet

$$\mathbf{Y}_n = \mathbf{Z}_\gamma \beta_k + \sigma \varepsilon, \quad \beta_k \in \mathbb{R}^k$$

wobei $\varepsilon = (\varepsilon(t_1), \dots, \varepsilon(t_n))^\top \in \mathbb{R}^n$ für standardnormalverteiltes weißes Rauschen steht. Das zugehörige Wahrscheinlichkeitsmodell für \mathbf{Y}_n hat die Parameter $(\gamma, \beta_k, \sigma)$ und die Dichte

$$M_\gamma = \left\{ p_\gamma(\mathbf{y}_n | \beta_k, \sigma) = (2\pi\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y}_n - \mathbf{Z}_\gamma \beta_k)^\top (\mathbf{y}_n - \mathbf{Z}_\gamma \beta_k) \right] \right\}.$$

Die Länge des Codes ist $-\log p_\gamma(\mathbf{y}_n | \beta_k, \sigma)$, d.h. die Dekodierung der Residuen. Für gegebenes γ führen die Maximum-Likelihood-Schätzer für β und σ zum kürzesten Code. Diese werden bezeichnet mit $\hat{\beta}_k \equiv \hat{\beta}_k(\mathbf{y}_n)$ und $\hat{\sigma}_n \equiv \hat{\sigma}(\mathbf{y}_n)$.

Noch nicht berücksichtigt ist hier die Codelänge des Modells bzw. der Modellklasse. Dies erfordert ein Modell für das Modell (vgl. Rissanen 1987). Durch Verwendung eines universellen Codes bzw. Wahrscheinlichkeitsmodells p_γ^* für die Modellklasse kann dieses Problem umgangen werden. Ein solcher universeller Code basiert auf der

Normalized Maximum Likelihood (NML) - Verteilung

$$p_\gamma^*(\mathbf{y}_n) = p_{NML,\gamma}(\mathbf{y}_n) = \frac{p_\gamma(\mathbf{y}_n | \hat{\beta}_k(\mathbf{y}_n), \hat{\sigma}_n)}{C_\gamma}, \quad (2.5)$$

wobei

$$C_\gamma = \int_{\mathbb{R}^n} p_\gamma(x_n | \hat{\beta}_k(x_n), \hat{\sigma}(x_n)) dx_n.$$

Für gegebene Daten \mathbf{y}_n sucht die MDL-Methode ein Wahrscheinlichkeitsmodell $p_\gamma(\cdot | \hat{\beta}_k, \hat{\sigma}_n)$ durch Minimierung des negativen Logarithmus von (2.5) über γ :

$$-\log p_\gamma(\mathbf{y}_n | \hat{\beta}_k(\mathbf{y}_n), \hat{\sigma}_n) + \log C_\gamma \equiv \text{Anpassung} + \text{Komplexität}. \quad (2.6)$$

Dies kann als Maximum-Likelihood-Ansatz mit Strafterm aufgefasst werden, mit einer Anpassungskomponente und einem Bestrafungsterm für die Komplexität (Grünwald 2000, Rissanen 1996).

Wird dieser Ansatz auf die nichtparametrische Regression angewandt, so dient die Designmatrix \mathcal{Z}^\top als Waveletmatrix \mathcal{W} . Die optimale Untermenge γ ergibt sich in diesem Fall als die Menge der k größten Koeffizienten.

Rissanen (2000) schlägt dann folgende Prozedur vor: Der Vektor $|w|_{(1)} \leq \dots \leq |w|_{(n)}$ enthalte die geordneten absoluten Waveletkoeffizienten. Die Wahl von k erfolgt nach dem NML-Prinzip durch

$$\hat{k} = \min_k \left((n-k) \log \frac{S_{(n)} - S_{(k)}}{n-k} + k \log \frac{S_{(k)}}{k} + \log k(n-k) \right) \quad (2.7)$$

mit $S_{(k)} = \sum_{i=1}^k |w|_{(i)}^2$ (vgl. Rissanen 2000). Die MDL-Methode ist damit ein Wavelet-Hard-Thresholding-Ansatz, wobei die Teilmenge Δ der Waveletkoeffizienten in (2.3) die der betragsmäßig größten \hat{k} Koeffizienten ist. Auf diese Weise ergibt sich eine Schranke λ_{MDL} und es werden Koeffizienten $|w|_{(\hat{k})} = |w_{j,k}| \geq \lambda_{\text{MDL}}$ verwendet. Rissanen (2000) zeigt, dass die MDL-Schranke λ_{MDL} asymptotisch kleiner ist als die aus (2.4) von Donoho und Johnstone (1994). Zur Berechnung wird die Schranke λ_{MDL} bestimmt und im Code des R-Packages `wavethresh` (Nason, Kovac und Maechler 2008) mit Daubechies *Extremal Phase Family*, $N = 6$ (vgl. Daubechies 1992) verwendet.

2.2.4 Lokaler Plug-in Kernschätzer (LPI)

Betrachtet wird der lokal adaptive Kernschätzer nach Herrmann (1997a):

$$\hat{f}(t; h_t) = \sum_{i=1}^n y(t_i) \int_{s_{i-1}}^{s_i} \frac{1}{h_t} K\left(\frac{t-u}{h_t}\right) du, \quad t \in [h_t; 1-h_t]. \quad (2.8)$$

Die Bandbreite h_t darf von t abhängen, d.h. sie kann lokal geschätzt werden. Die Punkte s_i sind wie folgt gewählt: $s_0 = 0$, $s_i = \frac{t_i+t_{i+1}}{2}$, $i = 1, \dots, n-1$, $s_n = 1$. Der Kern K hat Ordnung $k \geq 2$, d.h. die ersten $k-1$ Momente von K verschwinden während das k -te Moment ungleich 0 ist. Der mittlere quadratische Fehler (MSE)

$$MSE(\hat{f}(t; h_t)) = E(\hat{f}(t; h_t) - f(t))^2$$

wird als Kriterium für den lokalen Fehler eines Schätzers verwendet. Für den globalen Fehler lässt sich der integrierte mittlere quadratische Fehler (MISE) heranziehen:

$$MISE(\hat{f}) = E \int_0^1 w(t)(\hat{f}(t, h_t) - f(t))^2 dt.$$

Die Bandbreite h_t soll bezüglich MSE bzw. MISE optimal geschätzt werden. Die unbekanntenen Größen in MSE und MISE werden in der Prozedur geschätzt und eingesetzt (Plug-in Algorithmus). Im ersten Schritt wird ein initiiender Schätzer \hat{f} für f angenommen, der auf einer globalen Bandbreite \hat{h} basiert. Diese wird dann in die Formeln für die lokale Bandbreite eingesetzt und die wiederum in Gleichung (2.8). Der komplette Algorithmus ist (vgl. Brockmann et al. 1993, Herrmann 1997a, Gasser, Kneip, Köhler 1991):

- (i) Die erste Bandbreite auf $\hat{h}_0 = \frac{k-1}{n}$ setzen.
- (ii) Iteration, um eine globale Bandbreite zu erhalten:

$$\hat{g}_i = c \hat{h}_{i-1} n^{\frac{2}{(2k+1)(2k+3)}} \quad \text{und} \quad \hat{h}_i = \left(\frac{\hat{\sigma}^2 C(K)}{n \hat{I}_k(f; \hat{g}_i)} \right)^{\frac{1}{2k+1}}$$

für $i = 1, \dots, (k+1)(2k+1)$. Dabei ist $\hat{\sigma}^2$ ein Varianzschätzer (lokal oder global, vgl. Herrmann 1997a), $C(K)$ eine Kernkonstante, $\hat{I}_k(f; \hat{g}_i)$ eine Schätzung für die k -te Ableitung und c ist ein Tuning-Parameter.

- (iii) Setze nun $\hat{h} = \hat{h}_{(k+1)(2k+1)}$ als globale Plug-in Bandbreite ein.
- (iv) Setze die Bandbreiten $g_i = c_i \hat{h} n^{\frac{2}{(2k+1)(2k+3)}}$ für $i = 1, 2$ und den lokalen Plug-in Schätzer in (2.8) ein, also

$$\hat{h}(t_j) = \psi(t_j; g_1) \left(\frac{\hat{S}(t_j)C(K)}{n\tilde{r}_k^2(t_j; g_1, g_2)} \right)^{\frac{1}{2k+1}} + (1 - \psi(t_j; g_1))\hat{h}, \quad j = 1, \dots, n;$$

Dabei ist $\hat{S}(t)$ ein lokaler Varianzschätzer, $\tilde{r}_k^2(t_j; g_1, g_2)$ ein lokaler Schätzer für die k -te Ableitung, abhängig von einem weiteren Kern und Bandbreiten g_1 und g_2 , ψ eine Gewichtungsfunktion und c_1 und c_2 Tuning-Parameter.

Details zu der Prozedur beschreibt Herrmann (1997a). Eine Implementierung wird von Herrmann (1997b) gegeben.

2.2.5 Adaptive Weights Smoothing (AWSk und AWSg)

Adaptive Weights Smoothing (AWS) ist eine Methode nach Polzehl and Spokoiny (2000, 2003).

Das Ziel ist, für jeden Designpunkt t_i die größte Umgebung zu finden, in der die zugrunde liegende Signalfunktion $f(t_i)$ gut durch eine Konstante oder ein Polynom approximiert werden kann. Je nach dem, ob in der Schätzung lokal Konstanten oder Polynome verwendet werden, entstehen stückweise konstante oder stückweise polynomiale Schätzungen. Im Folgenden wird nur der Fall lokal konstanter Schätzungen beschrieben, da dies die Notation erheblich vereinfacht.

Für festes $i \in \{1, \dots, n\}$ wird der Schätzer $\hat{f}_\theta(t_i)$ für f an der Stelle t_i iterativ als gewichtetes Mittel der Beobachtungen berechnet. Die k -te Iteration berechnet den Schätzer

$$\hat{f}_\theta^{(k)}(t_i) = \frac{\sum_{j=1}^n w_{ij}^{(k)} \cdot y(t_j)}{\sum_{j=1}^n w_{ij}^{(k)}}. \quad (2.9)$$

Der Algorithmus beginnt mit einem üblichen Kernschätzer mit Kern K_l und Bandbreite $h^{(0)}$. Das erste Gewicht ist $w_{ij}^{(0)} = K_l(|\frac{t_i - t_j}{h^{(0)}}|^2)$. In jeder Iteration wird die Bandbreite $h^{(k)}$ erhöht. Mittels eines Gedächtnisparameters $\eta \in (0, 1)$ und eines Skalierungsfaktors

$a > 1$ werden neue Gewichte iterativ definiert:

$$w_{ij}^{(k)} = \eta \cdot w_{ij}^{(k-1)} + (1 - \eta) \cdot K_l \left(\frac{(t_i - t_j)^2}{(ah^{(k-1)})^2} \right) K_s(s_{ij}^{(k)}) K_e(e_{ij}^{(k)}),$$

Dabei sind $s_{ij}^{(k)}$ und $e_{ij}^{(k)}$, $j = 1, \dots, n$, Strafterme: In $s_{ij}^{(k)}$ geht ein Vergleich von $f_n^{(k-1)}(t_i)$ und $f_n^{(k-1)}(t_j)$ ein. Dadurch werden sozusagen zwei lokale Modelle verglichen und $s_{ij}^{(k)}$ als *Abstand* der beiden Modelle über einen lokalen Likelihoodansatz berechnet. Der Strafterm $e_{ij}^{(k)}$ beschränkt den Einfluss einer Beobachtung, beispielsweise wenn t_j als *high level point* für die lokale Anpassung an der Stelle t_i gilt. Die neue Bandbreite berechnet sich als Produkt der alten Bandbreite mit dem Skalierungsfaktor a : $h^{(k)} = ah^{(k-1)}$. Die Iteration endet, wenn $h^{(k)} > h_{\max}$ für eine zuvor festgelegte maximale Bandbreite h_{\max} gilt.

In dieser Arbeit werden zwei Versionen von AWS betrachtet: AWSk liefert eine stückweise konstante Schätzung, während AWSg stückweise polynomiale Approximationen ergibt.

Der Algorithmus und Angaben zur Wahl der Parameter und Kerne finden sich in Polzehl and Spokoiny (2000, 2003, 2006), Software ist im R-Package `aws` (Polzehl 2007) enthalten.

2.2.6 Taut String (TS)

Taut String basiert auf der Idee, eine möglichst *einfache* Schätzfunktion zu finden, die gleichzeitig eine *adäquate Approximation* der Daten zeigt (vgl. Davies und Kovac 2001). Dazu müssen die Begriffe *einfache Funktion* und *adäquate Approximation* definiert werden. Ersteres wird über die Anzahl lokaler Extremwerte erreicht: eine Funktion mit wenigen lokalen Extremwerten ist einfacher als eine mit vielen lokalen Extremwerten, ist also weniger komplex. Eine adäquate Approximation der Daten liegt vor, wenn die Schätzfunktion zu Residuen führt, die bestimmte Eigenschaften von weißem Rauschen haben. Der Taut String-Algorithmus erzeugt eine Kandidatenfunktion mit geringer Komplexität und überprüft dann, ob diese Funktion adäquat ist. Dabei wird wie folgt vorgegangen.

Verwendet werden die aufsummierten Daten $y^\circ(t_j) = \frac{1}{n} \sum_{i=1}^j y(t_i)$, $j = 1, \dots, n$. Dar-

um wird mittels einer Konstanten $C > 0$ ein Schlauch gelegt, durch den obere und untere Schranken l und u definiert werden: $l_n = y^\circ - \frac{C}{\sqrt{n}}$, $u_n = y^\circ + \frac{C}{\sqrt{n}}$. Nun wird in diesen Schlauch eine Saite (string) s° mit Endpunkten $(0, 0)$ und $(1, y^\circ(1))$ gelegt und 'gezogen', bis sie straff (taut) ist. s° ist diejenige Funktion auf $[0; 1]$ mit kürzester Länge

$$\text{length}(s^\circ) = \int_0^1 \sqrt{1 + s^{\circ(1)}(t)^2} dt, \quad (2.10)$$

und

$$s^\circ(0) = 0, \quad s^\circ(1) = y^\circ(1), \quad l(t) \leq s^\circ(t) \leq u(t), \quad 0 \leq t \leq 1. \quad (2.11)$$

Diese Konstruktion führt dazu, dass die Funktion innerhalb des Schlauchs stückweise linear ist und die obere und untere Schranke an einigen Stellen berührt. An den Stellen, an denen s° von der oberen Schranke u zur unteren Schranke l wechselt, hat die Ableitung $s^{\circ(1)}$ der Funktion s° ein lokales Maximum. Wechselt s° von der unteren Schranke l zur oberen Schranke u so hat $s^{\circ(1)}$ ein lokales Minimum. Die Ableitung $s^{\circ(1)}$ hat die kleinstmögliche Anzahl Extremwerte aller Funktionen, die (2.11) erfüllen. Als Ableitung einer stückweise linearen Funktion ist $s^{\circ(1)}$ stückweise konstant.

In der Taut String-Prozedur dient $\hat{f} = s^{\circ(1)}$ als eine Kandidatenfunktion für eine Schätzung der gesuchten Funktion f .

Um zu untersuchen, ob diese Kandidatenfunktion eine adäquate Schätzung liefert, wird überprüft, ob die Residuen $r(t_i) = y(t_i) - \hat{f}(t_i)$, $i = 1, \dots, n$, eine hinreichend kleine Multiresolutionsnorm haben. Die Multiresolutionsnorm ist definiert als

$$\|(x_1, \dots, x_n)\|_{MR} := \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} x_t \right|,$$

mit einem Intervallsystem \mathcal{I} (vgl. Mildenerger 2008). Davies und Kovac (2001) verwenden die Schranke $\hat{\sigma} \sqrt{\tau \log n}$ mit $\hat{\sigma} = \frac{1.48}{\sqrt{2}} \text{med} \{|y(t_2) - y(t_1)|, \dots, |y(t_n) - y(t_{n-1})|\}$. Ist die Multiresolutionsnorm für alle Intervalle kleiner als diese Schranke, so wird die Funktion \hat{f} als adäquate Approximation der Daten angesehen, mit anderen Worten sie erfüllt dann die *Multiresolutionsbedingung*. Dagegen wird der oben beschriebene Schlauch in denjenigen Intervallen, in denen diese Bedingung nicht erfüllt ist, verkleinert. Dadurch entstehen in der Schätzfunktion mehr Extremwerte, d.h. sie wird komplexer. Dieses Vorgehen wird solange fortgeführt, bis die Multiresolutionsbedingung in

allen Intervallen erfüllt ist. Bei großem n können auch dyadische Intervalle verwendet werden, um eine kürzere Rechenzeit zu erhalten. Genauere Angaben zu TS sind in Davies und Kovac (2001) zu finden, im R-Package `ftnonpar` (Davies und Kovac 2008) ist die Methode als `pmreg` implementiert.

2.2.7 Glatte Taut String-Methode (STS)

In Kovac (2007) wird eine glatte Version der Taut String-Methode vorgeschlagen. Als Grundlage dient eine Abwandlung der oben beschriebenen Methode, die die *Totale Variation* minimiert. Nach Kovac (2007) kann die Taut String-Methode auch als Algorithmus betrachtet werden, der die durch die Totale Variation von f bestrafte L_2 -Norm der Residuen minimiert:

$$T^{TS}(f) = \sum_{i=1}^n (y(t_i) - f(t_i))^2 + \lambda \sum_{i=1}^{n-1} |f(t_{i+1}) - f(t_i)| \longrightarrow \min! \quad (2.12)$$

Durch den Strafterm der Totalen Variation entstehen stückweise konstante Schätzungen. Um glattere Schätzungen zu erhalten, kann ein anderer Strafterm benutzt werden, zusätzlich kann der Strafterm lokal gewählt werden:

$$T^{SM}(f) = \sum_{i=1}^n (y(t_i) - f(t_i))^2 + \sum_{i=1}^{n-1} \lambda_i \sqrt{(t_{i+1} - t_i)^2 + (f(t_{i+1}) - f(t_i))^2}. \quad (2.13)$$

Die Lösung dieses Problems erfolgt durch einen ähnlichen Algorithmus wie die normale Taut String-Methode. Für das Funktional (2.13) lässt sich iterativ der Minimierer finden, indem ebenso ein Schlauch benutzt wird. Um auch hier die Multiresolutionsbedingung zu erfüllen, werden die λ_i beginnend mit einem großen Wert, schrittweise angepasst, bis die zu diesem λ_i gehörige Lösung von (2.13) die Multiresolutionsbedingung erfüllt. Zusätzlich werden noch Bedingungen an das Monotonieverhalten der Schätzfunktion gestellt, die verhindern sollen, dass zu viele lokale Extremwerte entstehen. Hierfür wird das Monotonieverhalten einer Taut String-Schätzung mit Totaler Variation nach Formel (2.12) für die glatte Taut String-Schätzung übernommen. Details sind in Kovac (2007) erschienen, das Verfahren ist als `smqreg` im R-Package `ftnonpar` (Davies und Kovac 2008) implementiert.

2.2.8 Potts-Schätzer (POT)

Die Potts-Methode rekonstruiert ein Signal durch stückweise konstante Funktionen, indem eine Kleinste-Quadrate-Anpassung mit der anfallenden Anzahl Sprungstellen bestraft wird (vgl. Boysen et al. 2009, Winkler und Liebscher 2002, Winkler et al. 2003, Wittich et al. 2008). Das Potts-Funktional ist definiert als:

$$P_\gamma : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, (\hat{f}, y) \longmapsto \gamma |J(\hat{f})| + \sum_{i=1}^n (y(t_i) - \hat{f}(t_i))^2, \gamma \geq 0, \quad (2.14)$$

wobei $J(\hat{f}) = \{i = 1, \dots, n-1 : f(t_i) \neq f(t_{i+1})\}$ die Sprungstellen von \hat{f} sind und $|A|$ die Anzahl Elemente in A angibt. Dabei sind y die Daten und \hat{f} eine Kandidatenfunktion zur Schätzung von f . Ein Potts-Schätzer für den Parameter γ ist definiert als Abbildung

$$\Phi_\gamma : \mathbb{R}^n \longrightarrow \mathbb{R}, y \longmapsto \Phi_\gamma(\cdot, y) \in \operatorname{argmin} P_\gamma(\cdot, y), \gamma \geq 0. \quad (2.15)$$

Wittich et al. (2008) untersuchen das Verhalten des Schätzers für variables γ . Betrachtet werden die Parameter

$$0 = \gamma_0 < \gamma_1(y) < \dots < \gamma_{m(y)}(y) < \gamma_{m(y)+1}(y) = \infty.$$

Nun gilt:

- Die Schätzung \hat{f} ist für jedes γ aus einem Intervall $I_j = (\gamma_{j-1}(y), \gamma_j(y))$ gleich.
- Die Schätzung ist stückweise konstant mit $\hat{f} = \frac{1}{|I_j|} \sum_{i \in I_j} y(t_i)$, d.h. in jedem Intervall I_j ist sie der Mittelwert seiner y -Werte.
- Für die kleinsten γ -Werte, d.h. $\gamma \in (0, \gamma_1(y))$, ergeben sich die Daten als Schätzung, d.h. $\hat{f} = \mathbf{y}$. Für die größten Werte, also $\gamma \in (\gamma_m(y), \infty)$ ist \hat{f} eine Konstante, nämlich der Mittelwert der Daten $\hat{f} = \frac{1}{n} \sum_{i=1}^n y(t_i)$.
- Zwischen kleinstem und größtem γ sinkt die Anzahl Sprünge mit steigendem γ .

In Winkler und Liebscher (2002) wird ein Algorithmus angegeben, der den Schätzer für jedes γ berechnet. Um ein γ auszuwählen wird beispielsweise vorgeschlagen, ein γ aus

demjenigen Intervall zu verwenden, dass die größten Veränderungen von γ -Werten erlaubt, so dass die Schätzung noch gleich bleibt (Wittich et al. 2008, Winkler et al. 2005). Boysen et al. (2009) schlagen vor, γ nach der Multiresolutionsbedingung zu wählen, d.h. das größtmögliche γ zu verwenden, so dass die Rekonstruktion die Bedingung noch erfüllt. Diese Version wird auch in dieser Arbeit verwendet. Um die Schätzung für alle γ -Werte zu erhalten, wird das R-package `tssegmentation` (Liebscher 2003) verwendet (vgl. auch Liebscher und Winkler 1999). Die Berechnung des Multiresolutionskriteriums folgt einem Code von Bernholt (Bernholt und Hofmeister 2006, Bernholt, Eisenbrand und Hofmeister 2007).

2.3 Gütekriterien

Wie Kapitel 2.2 zeigt, erfolgt die Herleitung einer nichtparametrischen Regressionsmethode häufig über ein Gütemaß. Zur späteren Beurteilung der Güte einer Methode werden neben asymptotischen Betrachtungen Simulationsstudien herangezogen, in denen wiederum verschiedene Gütemaße von Bedeutung sind. Diese Maße variieren nach dem Ziel der Regression. Die im Folgenden vorgestellten Gütekriterien werden in Kapitel 7 mit dem in dieser Arbeit hergeleiteten Extremstellenverlust verglichen.

- Häufig lassen sich L_p -Normen, speziell die L_2 -Norm oder davon abgeleitete Kriterien wie der Mittlere Quadratische Fehler als Gütekriterium, sowohl in der Herleitung als auch in der Beurteilung einer Regressionsmethode, heranziehen (vgl. z.B. Härdle et al. 2004, Donoho und Johnstone 1994, Herrmann 1997a, Rissanen 2000). Da sich diese Arbeit auf diskrete Designpunkte bezieht, wird auch folgende diskretisierte Version der L_1 - und L_2 -Norm zweier Funktionen f und \hat{f} verwendet:

$$d_p(f, \hat{f}) = \left(\frac{1}{n} \sum_{i=1}^n |f(t_i) - \hat{f}(t_i)|^p \right)^{\frac{1}{p}}, p = 1, 2. \quad (2.16)$$

- Boysen et al. (2009) verwenden den Hausdorff-Abstand, um die Sprungstellen einer wahren mit denen einer geschätzten Funktion zu vergleichen (siehe auch Huttenlocher, Klanderman und Rucklidge 1993). Der *Hausdorff-Abstand* ρ_H zweier

Mengen A und B (beispielsweise von wahren und geschätzten Sprungstellen) ist definiert als

$$\rho_H(A, B) = \max \left\{ \max_{a \in A} \min_{b \in B} |b - a|, \max_{b \in B} \min_{a \in A} |b - a| \right\}, \quad (2.17)$$

wobei $\rho_H(A, \emptyset) = \rho_H(\emptyset, A) = 1$.

- In Kovac (2007) ist das Ziel der Regression das Erkennen von Extremstellen. Als Gütekriterium in einer Simulation dient der *Median der Anzahl der geschätzten Maxima*. Gleichzeitig wird die wahre Anzahl Maxima genannt, damit beim Lesen ein Vergleich möglich ist. Ähnlich gibt Fryzlewicz (2007) für eine Simulation die *Anzahl Fälle an, in denen die korrekte Anzahl Maxima geschätzt wurde*, zusätzlich noch die Anzahl Fälle, in denen ein Maximum zu viel bzw. zu wenig geschätzt wurde.

EXTREMSTELLENVERLUST

Ein typisches Szenario aus Simulationsstudien in der nichtparametrischen Regression besteht darin, eine Signalfunktion f wie in Modell (2.1) mit normalverteiltem Rauschen zu überlagern. Mittels einer Regressionsmethode wird aus den so erzeugten Daten die unterliegende und im Rahmen von Simulationsstudien somit bekannte Funktion f durch eine Funktion \hat{f} geschätzt, um Eigenschaften der Regressionsmethode zu untersuchen und Aussagen über deren Güte ableiten zu können. So zum Beispiel in Abbildung 3.1

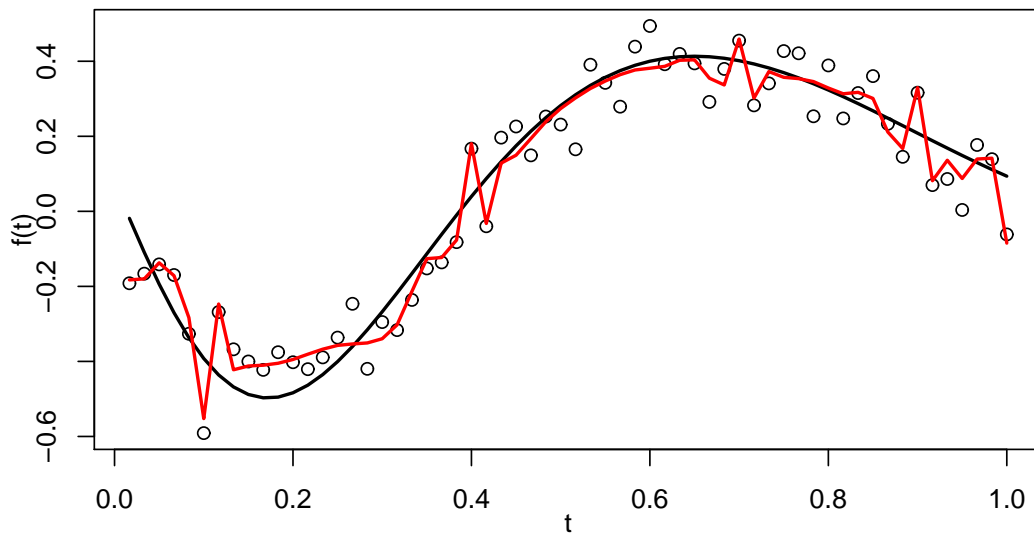


Abbildung 3.1: Datenpunkte (Kreise) mit zugehöriger Signalfunktion (schwarz) und Schätzfunktion (rot)

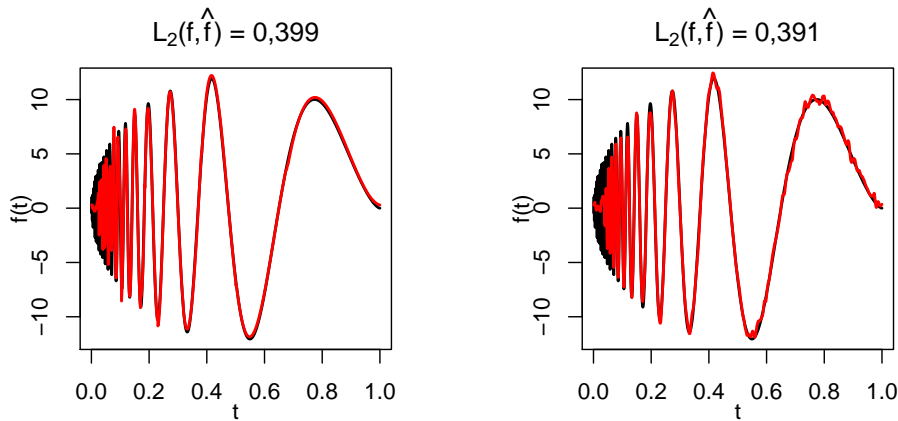


Abbildung 3.2: Zwei verschiedene Schätzfunktionen (rot) für die gleiche Signalfunktion (schwarz)

gezeigt: schwarz gezeichnet ist eine Signalfunktion mit um sie herum streuenden simulierten Datenpunkten und rot eine zugehörige Schätzfunktion. Für das menschliche Auge ist offensichtlich, dass sich Artefakte, d.h. überflüssige Extremwerte, in der Schätzung häufen. Für bestimmte Anwendungen ist das Erkennen von Extremwerten wichtig (vgl. Kovac 2007). Diese Arbeit beschäftigt sich damit, eine Verlustfunktion zu finden, die die Güte der Schätzung im Hinblick auf das Auffinden der Extremwerte misst. Zumeist werden zum Vergleich verschiedener Regressionsmethoden L_p -Normen genutzt (vgl. Kapitel 2.3). Jedoch haben Artefakte wie in Abbildung 3.1 oft keinen negativen Einfluss auf dieses Gütekriterium. Dies ist in Abbildung 3.2 dargestellt. Schwarz gezeichnet findet sich wieder eine Signalfunktion, rot die zugehörigen Schätzungen. Die L_2 -Norm der beiden verschiedenen Schätzfunktionen im linken und rechten Graphen von Abbildung 3.2 ist nahezu gleich, dennoch hat die rechts dargestellte Schätzung deutlich erkennbare überflüssige Extremwerte im Gegensatz zu der im linken Graphen. Dieses Beispiel zeigt, dass die L_2 -Norm nicht als Verlust bezüglich des Erkennens von Extremwerten geeignet ist und verdeutlicht die Notwendigkeit einer solchen Funktion, die die Güte einer Schätzung bezogen auf das Erkennen von Extremwerten beurteilt.

Im folgenden Unterkapitel 3.1 werden für die Herleitung eines Extremstellenverlustes benötigte Begriffe definiert. Das Unterkapitel 3.2 beinhaltet darauf aufbauend die Her-

leitung des neuen Verlustes. Dazu werden gewünschte Eigenschaften eines Extremstellenverlustes bestimmt, dann verschiedene Möglichkeiten der Definition vorgestellt, und schließlich die möglichen Verlustfunktionen verglichen. Zuletzt wird basierend auf dem am besten geeigneten Verlust der endgültige Extremstellenverlust definiert.

3.1 Definitionen und Notation

Zunächst erfolgt eine Definition von Extremstellen, die auf die speziellen Anforderungen in der Extremstellenerkennung im Kontext der nichtparametrischen Regression zugeschnitten ist. Dazu werden für eine Funktion f die *Veränderungen* der Funktionswerte betrachtet. Die Menge der Stellen, an denen sich die Funktionswerte von f ändern, ist gegeben durch

$$\{j_2 < \dots < j_k\} = \{i | f(t_i) \neq f(t_{i-1})\} \subset \{2, \dots, n\}. \quad (3.1)$$

Der Vektor, der diese Stellen enthält, ist nun $\mathbf{z} = (z_1, \dots, z_k) = (t_1, t_{j_2}, \dots, t_{j_k})$. Der erste Funktionswert wird als Veränderung betrachtet.

Bei vielen Testbed-Funktionen ändert sich der Funktionswert $f(t_i)$ an jeder Stelle t_i und damit ist $\mathbf{z} = \mathbf{t}$. Bei stückweise konstanten Funktionen gibt \mathbf{z} den Vektor der Sprungstellen an. Bei diesen Funktionen ergibt sich durch nachfolgende Definition eine Extremstelle genau in der Mitte jenes Abschnitts, dessen Werte größer bzw. kleiner als die Werte der danebenliegenden Abschnitte sind. Mithilfe des Vektors der Veränderungen werden Extremstellen wie folgt definiert:

Definition 3.1 (Extremstellen)

Sei $l \in \{1, \dots, k\}$ und $a = \frac{z_l + z_{l+1} - \frac{1}{n}}{2}$. Dann heißt a lokale Minimalstelle von f , wenn gilt

$$f(z_{l-1}) > f(z_l) \text{ und } f(z_{l+1}) > f(z_l).$$

Entsprechend heißt a lokale Maximalstelle von f , wenn

$$f(z_{l-1}) < f(z_l) \text{ und } f(z_{l+1}) < f(z_l).$$

Minimal- und Maximalstellen werden zu Extremstellen zusammengefasst. Der erste und letzte Datenpunkt, also $t_1 = z_1$ und t_n gelten nicht als Extremstellen.

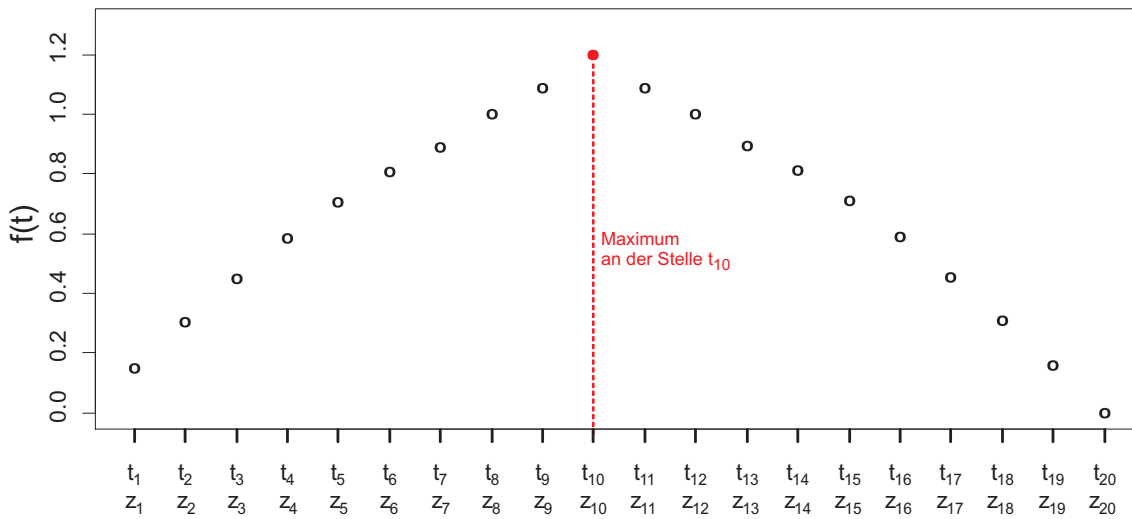


Abbildung 3.3: Beispiel für ein Maximum einer stetigen Funktion bei Verwendung von 20 Designpunkten

Beispiel 3.2 (Extremstellen)

In den Abbildungen 3.3 und 3.4 sind Beispiele für Extremstellen nach Definition 3.1 gegeben. Dargestellt ist jeweils eine Funktion $f(\mathbf{t})$ mit 20 Designpunkten $t_1 = \frac{1}{20}, \dots, t_{20} = \frac{20}{20} = 1$. In Abbildung 3.3 ändert sich der Funktionswert bei jedem Designpunkt. Da $f(z_9) < f(z_{10})$ und $f(z_{10}) > f(z_{11})$ ist $\frac{z_{10} + z_{10+1} - \frac{1}{n}}{2} = \frac{z_{10} + z_{10}}{2} = z_{10} = t_{10}$ eine Maximalstelle. In Abbildung 3.4 ist $f(\mathbf{t})$ stückweise konstant und nimmt zwei verschiedene Werte an: $f(t_1) = \dots = f(t_6) = f(t_{16}) = \dots = f(t_{20}) = 0.2$ und $f(t_7) = \dots = f(t_{15}) = 0.8$. Es gibt drei Stellen, an denen sich der Funktionswert verändert: $z_1 = t_1$ (nach Definition), $z_2 = t_7$ und $z_3 = t_{16}$, d.h. $j_2 = 7$ und $j_3 = 16$. Es gilt $f(z_1) < f(z_2)$ und $f(z_2) > f(z_3)$ und daher ist

$$\frac{z_2 + z_{2+1} - \frac{1}{n}}{2} = \frac{t_7 + t_{16} - \frac{1}{20}}{2} = \frac{t_7 + t_{15}}{2} = \frac{\frac{7}{20} + \frac{15}{20}}{2} = \frac{11}{20} = t_{11}$$

eine Maximalstelle.

Fällt eine Extremstelle bei stückweise konstanten Funktionen in einen Abschnitt mit einer geraden Anzahl von Designpunkten, so liegt sie nach Definition 3.1 zwischen zwei Designpunkten. Dies kann beispielsweise bei der Blocks-Funktion der Fall sein (vgl. Kapitel 2.1).

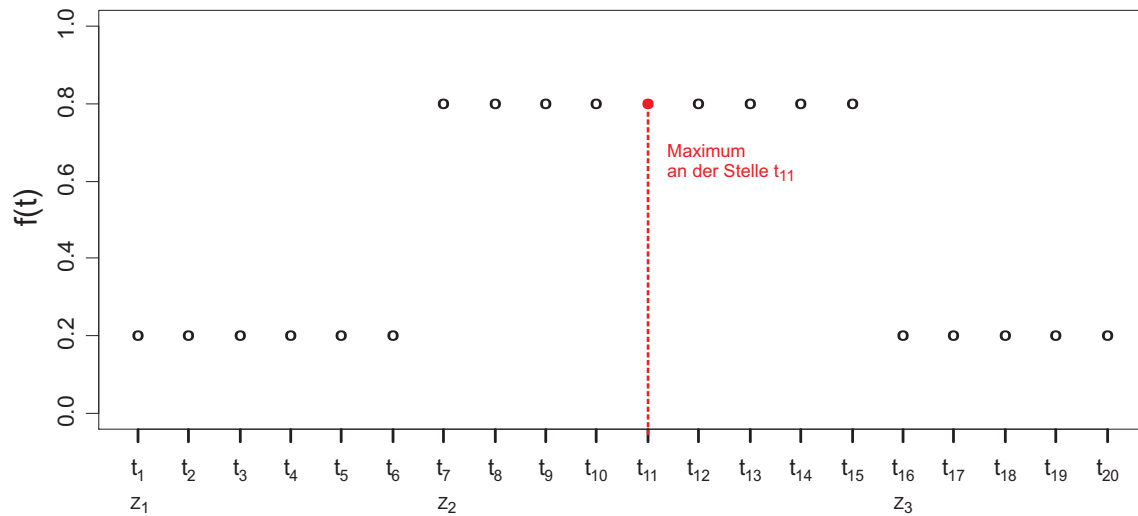


Abbildung 3.4: Beispiel für ein Maximum bei einer stückweise konstanten Funktion mit 20 Designpunkten

Die auf dem kompletten Intervall $[0; 1]$ konstante Funktion aus Kapitel 2.1 hat keine Extremstellen.

Die Extremstellendefinitionen 3.1 unterscheidet sich von denen der Analysis (vgl. z.B. Barner und Flohr 1991) in der Behandlung konstanter Abschnitte, die dort (im stetigen Fall) unendlich viele Extremstellen haben. Da eine Verlustfunktion, die über richtig geschätzte Extremstellen definiert wird, nur sinnvoll ist, wenn nicht an jedem Designpunkt eine Extremstelle liegt, ist diese klassische Definition ungeeignet (vgl. auch Donoho 1988).

Definition 3.3 (Extremwerte)

Sei a eine Minimalstelle von f . Der Funktionswert $f(a)$ heißt Minimum. Entsprechend heißt $f(a)$ Maximum, wenn a eine Maximalstelle von f ist. Minima und Maxima werden Extremwerte genannt.

Für den neuen Verlust sollen die Extremstellen einer geschätzten Funktion mit denen der zu schätzenden Testbed-Funktion verglichen werden. Wenn der Abstand zwischen wahren und geschätzten Extremstellen klein ist, so ist die Schätzung bezüglich Extremstellen gut. Daher wird eine wahre Extremstelle als *identifiziert* definiert, wenn

der Abstand zu einer geschätzten Extremstelle kleiner als eine Toleranz um die wahre Extremstelle ist:

Definition 3.4 (identifizierte Extremstelle)

Sei $x_{true}^\ell, \ell \in \{\min, \max\}$, eine Extremstelle der Signalfunktion f und $\delta^\ell \in \mathbb{R}$ eine gegebene Toleranz. Sei weiterhin x_{est}^ℓ eine Extremstelle der geschätzten Funktion. Dann heißt x_{true}^ℓ identifiziert durch x_{est}^ℓ , wenn

$$|x_{true}^\ell - x_{est}^\ell| \leq \delta^\ell. \quad (3.2)$$

Gibt es $k^{\ell*}$ geschätzte Extremstellen $x_{est,i}^\ell, i \in \{1, \dots, k^{\ell*}\} \subseteq \{1, \dots, k^\ell\}$, die (3.2) erfüllen, wird x_{true}^ℓ nur durch $x_{est,j}^\ell$ mit $|x_{true}^\ell - x_{est,j}^\ell| < |x_{true}^\ell - x_{est,i}^\ell|$ für alle $i \neq j, i = 1, \dots, k^{\ell*}$, identifiziert.

Es wird also zwischen Minimal- und Maximalstellen unterschieden und jede Extremstelle der wahren Funktion kann nur durch genau eine Extremstelle der geschätzten Funktion identifiziert werden. Nur wenn im entsprechenden Toleranzintervall um eine wahre Minimalstelle auch eine geschätzte liegt, ist die wahre identifiziert, entsprechendes gilt für Maximalstellen. Die Differenzierung von Minimalstelle und Maximalstelle wird mit *Extremstellen verschiedener Art* bezeichnet, zwei Minimalstellen oder zwei Maximalstellen haben *die gleiche Art*.

Die Toleranz wird in Kapitel 4 definiert und es werden verschiedene mögliche Toleranzen vorgeschlagen. Die Wahl der Toleranz δ hat Einfluss auf die Menge der identifizierten Extremstellen. Große Toleranzen führen zu vielen identifizierten Extremstellen, kleine zu wenigen. Damit Definition 3.4 eindeutig ist, ist die Toleranz um eine Extremstelle so zu wählen, dass keine weitere wahre Extremstelle der gleichen Art im Toleranzbereich liegt, denn es muss sicher gestellt sein, dass eine wahre Extremstelle immer nur durch genau eine geschätzte Extremstelle identifiziert werden kann. Durch Definition 4.1 wird dies gewährleistet.

3.2 Herleitung des Extremstellenverlustes

In diesem Kapitel wird ein Extremstellenverlust hergeleitet, der die Güte einer Schätzfunktion bezogen auf die Extremstellen in eine Maßzahl fasst. Hierfür muss in den Verlust eingehen, wenn in der Schätzfunktion zu viele Extremstellen enthalten sind, wenn sie an der falschen Stelle liegen und wenn zu wenig Extremstellen geschätzt werden. Diese Werte lassen sich aus folgenden drei Größen bestimmen:

- n_{true} : Anzahl der wahren Extremstellen,
- n_{est} : Anzahl der geschätzten Extremstellen,
- n_{id} : Anzahl der identifizierten Extremstellen.

Prinzipiell kann an jedem der Designpunkte t_2, \dots, t_{n-1} , eine Extremstelle von f bzw. \hat{f} liegen. Daher gilt $0 \leq n_{true} \leq n-2$ und $0 \leq n_{est} \leq n-2$. Außerdem gilt $0 \leq n_{id} \leq n_{true}$, da nicht mehr Extremstellen identifiziert werden können als es wahre Extremstellen gibt. Entsprechend gilt $0 \leq n_{id} \leq n_{est}$. n_{true} nimmt den Wert 0 beispielsweise bei einer konstanten Funktion an. $n_{est} = 0$ gilt, wenn die Schätzfunktion beispielsweise konstant ist. Der Fall $n_{id} = 0$ tritt ein, wenn keine wahren oder keine geschätzten Extremstellen existieren oder wenn alle geschätzten Extremstellen an den falschen Stellen liegen. Idealerweise wird die richtige Anzahl von Extremstellen an den richtigen Stellen geschätzt. In diesem Fall ist $n_{true} = n_{est} = n_{id}$.

Es muss zwischen *zu wenig identifizierten* und *zu viel bzw. wenig geschätzten* Stellen unterschieden werden. Zu wenig geschätzte Stellen bedeuten $n_{est} < n_{true}$, zu viele $n_{est} > n_{true}$. Die Differenz $\xi = n_{est} - n_{true}$ ist zur Messung des Fehlverhaltens nicht geeignet, da nur die Anzahl und nicht die Stellen der geschätzten Extremstellen Einfluss haben. Beispielsweise ist $\xi = 0$ möglich, auch wenn die geschätzten Extremstellen alle an der falschen Stelle liegen oder nicht die richtige Art haben. ξ ist daher ausschließlich geeignet, um eine Tendenz für eher zu viele – positives Vorzeichen – oder eher zu wenige – negatives Vorzeichen – geschätzte Extremstellen anzuzeigen.

Besser geeignet sind zu wenig *identifizierte* Extremstellen: $F_1 = n_{true} - n_{id}$. Über n_{id} geht in diesen Fehler indirekt auch n_{est} ein. Der Fehler für die *zu viel geschätzten* Extremstellen ist $F_2 = n_{est} - n_{id}$. F_2 gibt die Anzahl der überflüssigerweise geschätzten Extremstellen an. In diesem Fall geht über n_{id} indirekt n_{true} ein. Beide Fehler sind nicht negativ, da es mehr (oder gleich viele) wahre als identifizierte Extremstellen gibt,

also $n_{true} \geq n_{id}$, gleiches gilt für geschätzte Extremstellen $n_{est} \geq n_{true}$. Es ist möglich, zu viele Extremstellen zu schätzen und gleichzeitig zu wenige zu identifizieren.

Diese einfachen Differenzen reichen offensichtlich nicht aus, um die Güte bezüglich des Erkennens von Extremstellen zu beschreiben. Eine differenziertere Betrachtung des Problems ist erforderlich.

3.2.1 *Gewünschte Eigenschaften eines Extremstellenverlustes*

Vor der Definition eines neuen *Extremstellenverlustes*, werden nachfolgend zunächst die gewünschten Eigenschaften eines solchen Verlustes festgehalten. Wie bereits zu Anfang dieses Kapitels dargestellt, soll es der Verlust ermöglichen, als einzelne Maßzahl eine Schätzfunktion anhand fehlender, überflüssiger oder an der falschen Stelle liegender Extremstellen zu beurteilen. Daneben sollten idealerweise folgende Eigenschaften vorliegen.

Bemerkung 3.1 (Gewünschte Eigenschaften)

- (i) **Normierung:** *Der Verlust sollte nach oben und unten beschränkt sein, damit das Ergebnis des Verlustes einer bestimmten Situation beurteilt werden kann, ohne Vergleichswerte heranzuziehen. Ein Verlust sollte als bestmöglicher Wert 0 haben und dieser sollte angenommen werden, wenn eine Methode einwandfrei Extremstellen erkennt, d.h. $n_{true} = n_{est} = n_{id}$. Wann der Verlust den schlechtest möglichen Wert (beispielsweise 1) annehmen soll, ist nicht so offensichtlich, da ein schlimmster Fall des Nicht-Erkennens von Extremstellen erst bestimmt werden muss.*
- (ii) **Wenig Informationsverlust:** *Die Information, ob ein Verlust durch zu viele oder zu wenige geschätzte Extremstellen entstanden ist, soll erhalten bleiben.*
- (iii) **Keine Definitionslücken:** *Der Verlust soll für alle Fälle berechenbar sein, d.h. insbesondere für die Fälle $n_{true} = 0$ und $n_{est} = 0$.*

- (iv) **Definitionsbereich:** Der Definitionsbereich der Verlustfunktion soll voll ausgeschöpft werden. Insbesondere soll die Verlustfunktion ihren besten Wert annehmen, wenn $n_{true} = n_{est} = n_{id}$.
- (v) **Bestrafung:** Die verschiedenen Fälle, die eine Schätzung verschlechtern, sollen berücksichtigt werden. Es muss den Verlust vergrößern, wenn keine oder zu wenig Extremstellen identifiziert werden, wenn zu viele Extremstellen geschätzt werden und wenn sie an der falschen Stelle geschätzt werden.
- (vi) **Verhältnis:** Das Verhältnis der Werte für verschiedene Fehler in der Schätzung soll stimmen: beispielsweise sollte der Verlust entsprechend weniger groß sein, wenn wenige wahre Extremstellen nicht identifiziert werden, als wenn viele wahre Extremstellen nicht identifiziert werden. Genauso sollen sich wenige zu viel geschätzte Extremstellen nicht so negativ auswirken wie viel zu viele.
- (vii) **Zusammenfassung verschiedener Verlustwerte:** Es sollte die Möglichkeit bestehen, sinnvoll einen Mittelwert oder Median über mehrere Werte des Verlustes bei einer Testbed-Funktion zu bilden. In der Untersuchung von Regressionsmethoden mittels Simulationen finden üblicherweise viele Datensätze Verwendung, für die auch ein gemeinsamer oder mittlerer Verlust von Interesse ist.

3.2.2 Herleitung verschiedener Verlustfunktionen

Peak Identification Loss

In Davies, Gather und Weinert (2008) wird der *Peak Identification Loss (PID)* verwendet.

Definition 3.2 (Peak Identification Loss PID)

Der Peak Identification Loss ist definiert als

$$PID = \operatorname{sgn}(n_{est} - n_{true})((n_{true} - n_{id}) + (n_{est} - n_{id})) \quad (3.3)$$

$$= \operatorname{sgn}(n_{est} - n_{true})(n_{true} + n_{est} - 2n_{id}). \quad (3.4)$$

Der PID setzt sich also aus den Fehlern F_1 und F_2 und dem Vorzeichen von ξ zusammen. Addiert ergeben die beiden Fehler die Verlustfunktion. Ein ähnlicher Verlust wird in Davies, Gather, Nordman und Weinert (2009) im Kontext der Histogrammkonstruktion verwendet.

Die Eigenschaften des PID lassen sich wie folgt zusammenfassen (vgl. auch Beispiel 3.10):

Bemerkung 3.3 (Eigenschaften des PID)

- $PID = 0$ für $n_{true} = n_{est} = n_{id}$, d.h. alle wahren Extremstellen sind identifiziert und keine überflüssigen Extremstellen geschätzt.
- PID ist nicht normiert.
- PID ist nach oben nur durch $2(n-2)$ beschränkt, denn theoretisch kann an jedem der Designpunkte t_2, \dots, t_{n-1} eine wahre Extremstelle liegen und eine weitere geschätzt werden, die sich jedoch in der Art unterscheidet, so dass $n_{id} = 0$ gilt. Der Definitionsbereich von PID ist damit $[-2(n-2), 2(n-2)]$ und damit abhängig von n .
- Der Stichprobenumfang n wird bei der Definition von PID nicht berücksichtigt.
- PID ist immer berechenbar, auch wenn eine oder mehrere der eingehenden Größen 0 werden.
- Der Betrag von PID ist klein, wenn die drei Werte n_{true} , n_{est} und n_{id} nah beieinander liegen, d.h. die Schätzung gut ist.

- Der Betrag von PID wird groß, wenn n_{true} oder n_{est} (oder beide) groß sind und n_{id} klein ist. Bestraft wird also, wenn keine oder überflüssige Extremstellen auftreten oder keine identifiziert werden.
- $PID > 0$ deutet auf zu viele geschätzte Extremstellen, $PID < 0$ auf zu wenige geschätzte Extremstellen hin.

Ein wesentlicher Nachteil des PID ist, dass er nicht normiert ist. Dadurch ist es schwierig, die Güte einer Schätzung zu erfassen, da derselbe PID -Wert je nach Signalfunktion auf eine gute oder schlechte Erkennung der Extremstellen hindeuten kann. Auch Vergleiche zwischen Schätzmethoden werden dadurch erschwert. Andere Eigenschaften des Verlustes sind positiv zu werten. Beispielsweise ist es ein Vorteil, dass PID immer berechenbar ist. Auch die Information, ob der Verlust eher durch zu viele oder zu wenige Extremstellen der Schätzfunktion entsteht, bleibt erhalten.

Abbildung 3.5 liefert einen Überblick über die Beträge der PID -Werte bei variierenden n_{est} und n_{id} . Für $n_{true} = 30$ wird der n_{est} -Wert angefangen bei 0 in Zehnerschritten bis auf 200 erhöht, gleichzeitig wird der n_{id} -Wert von 0 in Zehnerschritten bis auf 30 erhöht. Die dritte Achse gibt für die jeweiligen n_{est} - und n_{id} -Werte das Ergebnis von $|PID|$ an. Der Betrag wird aufgrund der besseren Darstellbarkeit verwendet. Der Verlust ist 0, wenn n_{true} , n_{est} und n_{id} gleich 30 sind. Mit zunehmender Abweichung von n_{est} und n_{id} vom wahren Wert n_{true} steigt der $|PID|$ -Wert linear an.

Mittel aus den Fehlern F_1 und F_2

Um einen normierten Verlust zu erhalten, ist es naheliegend, zunächst die Fehler F_1 und F_2 zu normieren:

$$\tilde{F}_1 = \frac{n_{true} - n_{id}}{n_{true}} \quad \text{und} \quad \tilde{F}_2 = \frac{n_{est} - n_{id}}{n_{est}}.$$

Da $0 \leq n_{id} \leq n_{true}$ und $0 \leq n_{id} \leq n_{est}$ gilt, liegen \tilde{F}_1 und \tilde{F}_2 jeweils zwischen 0 und 1, wenn sie definiert sind, also wenn $n_{true} > 0$ bzw. $n_{est} > 0$ gilt.

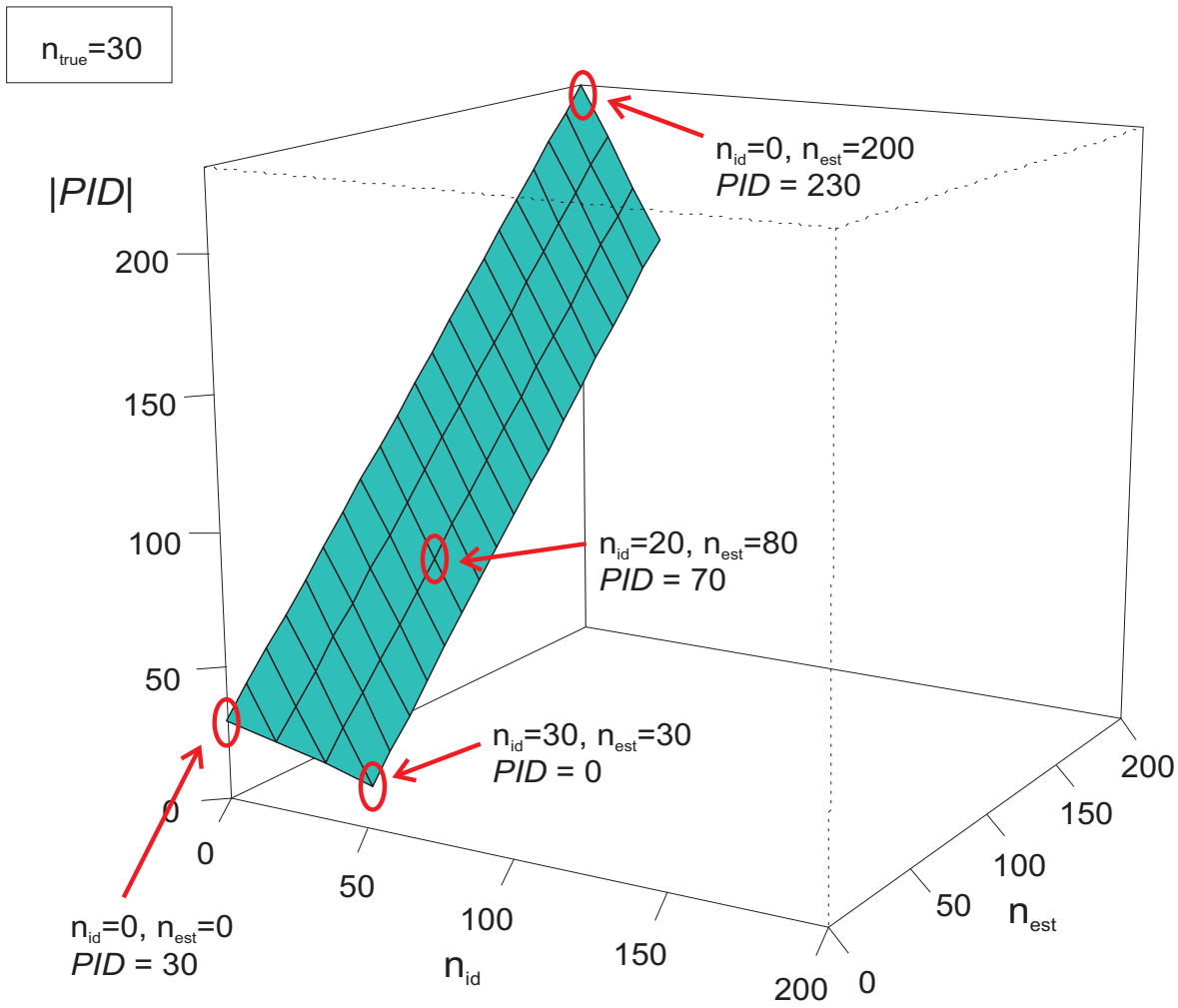


Abbildung 3.5: Ergebnisse von PID bei variierenden n_{est} und n_{id}

Definition 3.4 ($loss_1$)

Als Verlustfunktion $loss_1$ wird das Mittel aus diesen normierten Fehlern definiert:

$$loss_1 = \frac{\frac{n_{true}-n_{id}}{n_{true}} + \frac{n_{est}-n_{id}}{n_{est}}}{2}. \quad (3.5)$$

Bemerkung 3.5 (Eigenschaften von $loss_1$)

- Der bestmögliche Wert $loss_1 = 0$ wird für $n_{true} = n_{est} = n_{id}$ angenommen, also wenn die Schätzung richtig ist.
- Der schlechteste Wert $loss_1 = 1$ wird erreicht, wenn $n_{true} > 0, n_{est} > 0$ und $n_{id} = 0$, also wenn (beliebig viele) Extremstellen vorhanden sind und geschätzt

aber nicht identifiziert werden.

- $loss_1$ ist unabhängig von n .
- Es sind Definitionslücken vorhanden: für $n_{true} = 0$ oder $n_{est} = 0$ ist $loss_1$ nicht definiert.
- Der Definitionsbereich wird nicht gut ausgeschöpft. Dies liegt daran, dass $loss_1$ ein Mittel aus zwei Fehlern ist. Ist bei richtiger Anzahl von n_{id} der erste Fehler $\tilde{F}_1 = 0$, kann $loss_1$ nicht größer als 0,5 sein. Durch einen großen $loss_1$ -Wert nahe 1 wird nur bestraft, wenn n_{est} und n_{true} groß sind und n_{id} klein ist.
- Ob sich der Verlust durch zu viele oder zu wenige geschätzte Extremstellen ergibt, ist nicht erkennbar.

Der Verlust $loss_1$ differenziert nicht zwischen wenigen und vielen wahren Extremstellen, die nicht identifiziert werden, da in beiden Fällen $\tilde{F}_1 = 1$ ist. Ebenso wird zwischen wenigen und vielen falsch geschätzten Extremstellen nicht unterschieden, da sich hierfür $\tilde{F}_2 = 1$ ergibt (vgl. auch Beispiel 3.10).

Auch für diesen Verlust wird ein Überblick der Ergebnisse bei variierenden n_{est} und n_{id} in Abbildung 3.6 gegeben. Wie bereits in Abbildung 3.5 ist $n_{true} = 30$ und n_{est} wird von 0 bis auf 200 in Zehnerschritten erhöht, gleichzeitig wird n_{id} von 0 bis auf 30 in Zehnerschritten erhöht. Die dritte Achse gibt für die jeweiligen n_{est} - und n_{id} -Werte das Ergebnis von $loss_1$ an. Sind $n_{true} = n_{est} = n_{id} = 30$, so gilt $loss_1 = 0$. Der Anstieg des $loss_1$ -Wertes bei zunehmender Abweichung von n_{est} und n_{id} vom wahren Wert n_{true} ist nicht linear. Deutlich wird, dass $loss_1$ für $n_{id} = 30$ auch bei hohem n_{est} keine Werte über 0,5 annimmt. Für $n_{id} = 0$ ist $loss_1 = 1$, unabhängig davon, welchen Wert n_{est} annimmt.

An dieser Stelle sei noch die Möglichkeit erwähnt, \tilde{F}_1 und \tilde{F}_2 verschieden zu gewichten, also $r\tilde{F}_1 + (1 - r)\tilde{F}_2$. In diesem Fall muss jedoch r bestimmt werden. Das erfordert die Entscheidung, welcher der Fehler wichtiger ist, also stärker einfließen soll. Wenn bei einem endgültigen Verlust erkennbar ist, worin die Schwächen der Schätzfunktion liegen, ist dies nicht mehr notwendig.

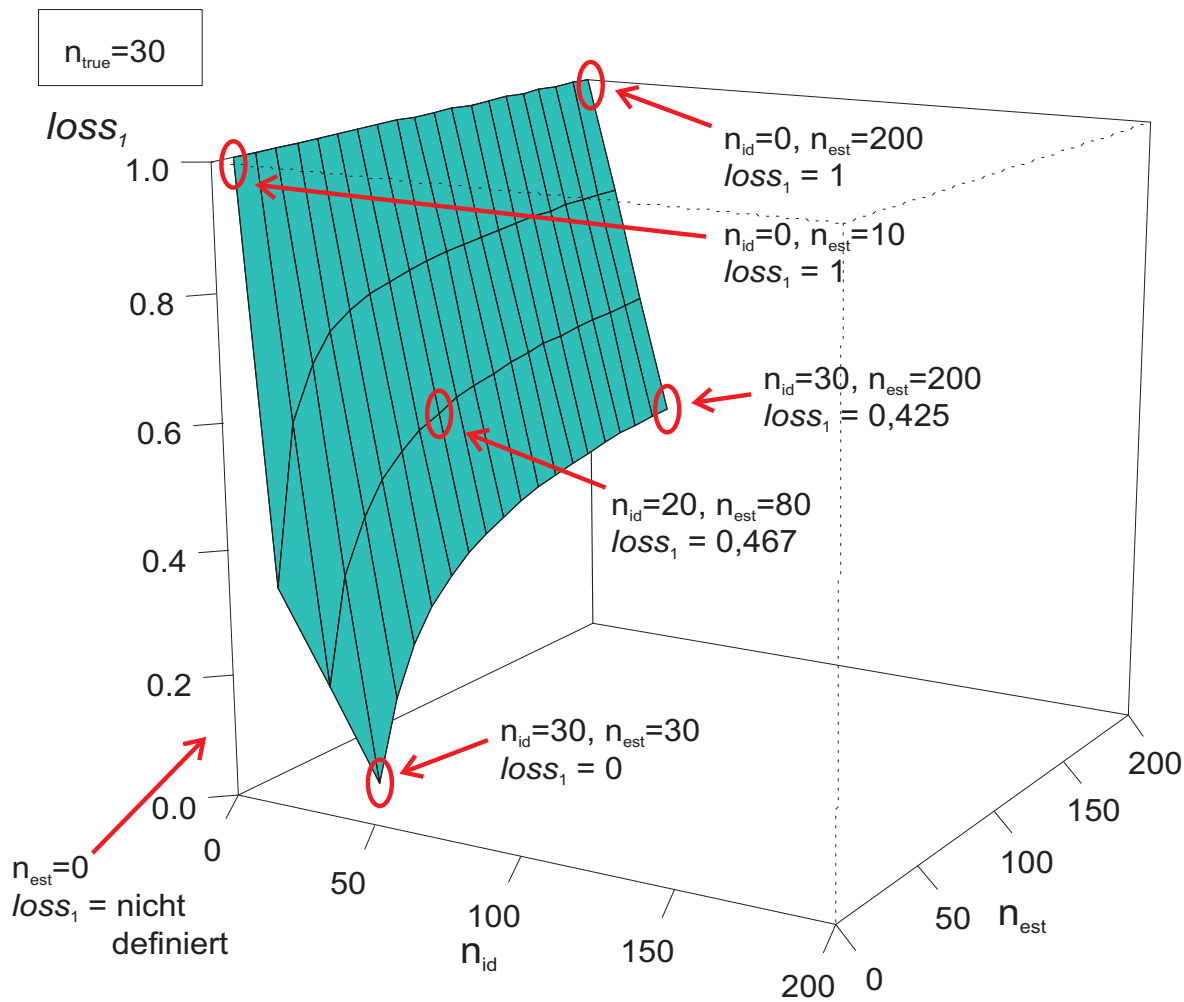


Abbildung 3.6: Ergebnisse von $loss_1$ bei variierenden n_{est} und n_{id}

Normierung von PID

Da PID durchaus positive Eigenschaften hat (vgl. Bemerkung 3.3), ist es wünschenswert, diesen Verlust zu normieren. Die eigentliche Schwierigkeit liegt darin, den *schlimmsten Fall* zu erfassen. Wie bereits erwähnt ist $2(n - 2)$ der theoretisch größtmögliche Wert von PID . Eine weitere Verlustfunktion entsteht folgendermaßen.

Definition 3.6 ($loss_2$)

Der Verlust $loss_2$ ist definiert als:

$$loss_2 = \frac{n_{true} + n_{est} - 2n_{id}}{2(n-2)} \quad (3.6)$$

Auf das Vorzeichen von PID wird hier zunächst verzichtet, da das Kriterium der Zusammenfassung aus Bemerkung 3.1 (vii) dadurch nicht gegeben ist.

Bemerkung 3.7 (Eigenschaften von $loss_2$)

- *Es gilt $loss_2 = 0$, wenn $n_{true} = n_{est} = n_{id}$.*
- *Der Verlust $loss_2$ ist immer berechenbar, es gibt keine Definitionslücken.*
- *Der Stichprobenumfang n geht mit ein.*
- *Da der theoretisch größtmögliche Wert von $2(n-2)$ nur bei sehr schlechten Schätzmethoden auftritt, ist diese Zahl zum Normieren in den meisten Fällen viel zu groß.*
- *Viel zu viele n_{est} werden stärker bestraft als wenig zu viele, denn durch wachsendes n_{est} wird der Zähler größer.*

Im Gegensatz zu den anderen vorgestellten Verlusten hängt $loss_2$ vom Stichprobenumfang n ab. Hier sind also die Ergebnisse unterschiedlich, wenn verschiedene n in die Berechnung eingehen, aber dieselben Größen für n_{true} , n_{est} und n_{id} bestehen bleiben (vgl. auch Beispiel 3.10). Die Werte von $loss_2$ sind häufig sehr klein, da die Stichprobengröße im Gegensatz zur Anzahl Extremstellen oft sehr groß ist. Wenn der Zähler kleiner als etwa 20 % des Stichprobenumfangs n ist, so gilt $loss_2 < 0,1$. Bei vielen Testbed-Funktionen tritt dieser Fall wegen $n_{true} < 10$ und damit $n_{id} < 10$ bereits bei relativ geringem n auf.

Abbildung 3.7 gibt für zwei verschiedene Stichprobengrößen $n = 200$ im linken und $n = 500$ im rechten Graphen Ergebnisse für $loss_2$ an. Bei $n_{true} = 30$ variieren wieder die n_{est} -Werte zwischen 0 und 200 und die n_{id} -Werte zwischen 0 und 30 jeweils in Zehnerschritten. Die dritte Achse gibt für die jeweiligen n_{est} - und n_{id} -Werte das

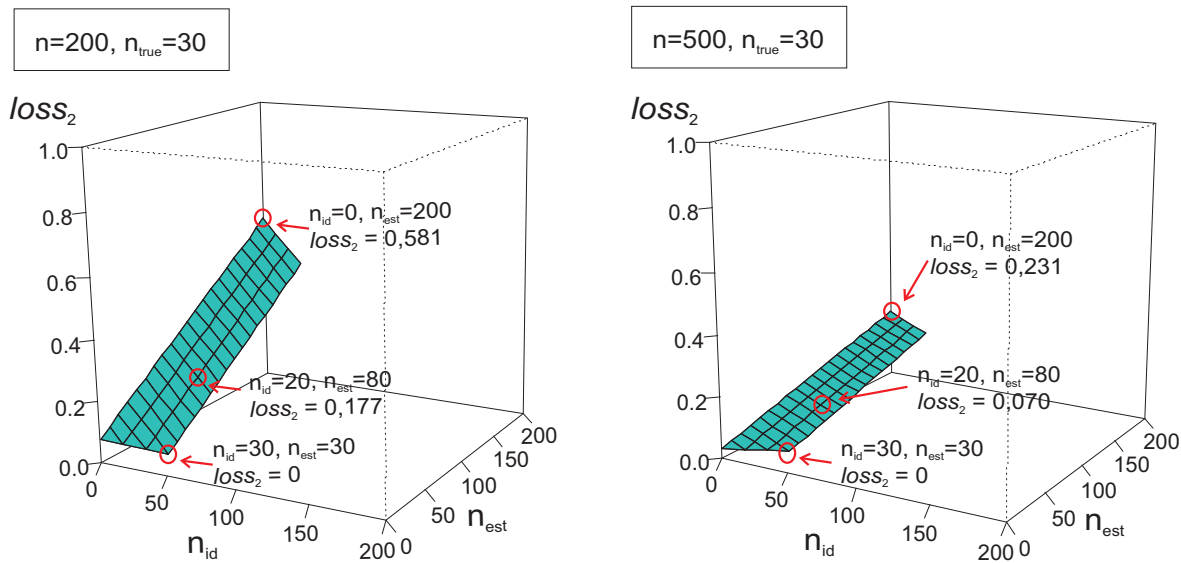


Abbildung 3.7: Ergebnisse von $loss_2$ bei variierenden n_{est} und n_{id}

Ergebnis von $loss_2$ an. Bei richtiger Anzahl geschätzter und identifizierter Extremstellen, $n_{true} = n_{est} = n_{id} = 30$, gilt $loss_2 = 0$. Bei zunehmender Abweichung von n_{est} und n_{id} vom wahren Wert n_{true} steigt $loss_2$ linear. Beide Graphen zusammen ergeben eine Vorstellung davon, wie stark der Einfluss des Stichprobenumfangs auf $loss_2$ ist. Für $n = 200$ ist der größte $loss_2$ -Wert 0,581 für $n_{est} = 200$ und $n_{id} = 0$. Ein solches schlechtes Abschneiden bezogen auf das Erkennen von Extremstellen durch eine Schätzfunktion lässt sich am Ergebnis nicht erkennen. Bei $n = 500$ ist der größte Wert mit 0,231 noch deutlich kleiner.

Zu bemerken ist hier noch, dass es die Möglichkeit gibt, PID nur durch n zu teilen. Dies führt allerdings nicht zu einem normierten Verlust, da $PID > n$ möglich ist, auch wenn dieser Fall in der Praxis selten vorkommt.

Bei der Normierung durch $2(n - 2)$ wird als schlimmster Fall angenommen, dass an allen Designpunkten falsch geschätzte Extremstellen liegen. Bei den meisten Schätzmethoden der nichtparametrischen Regression ist das nicht realistisch. Anstatt auf die Anzahl Designpunkte kann der Wert auch auf n_{true} und n_{est} bezogen werden. Bestraft wird dadurch ein schlechtes Verhältnis zwischen wahren, geschätzten und identifizierten Extremstellen.

Definition 3.8 ($loss_3$)

Der Verlust $loss_3$ ist definiert als:

$$loss_3 = \frac{n_{true} + n_{est} - 2n_{id}}{n_{true} + n_{est}}. \quad (3.7)$$

Da $n_{id} \leq n_{true}$ und $n_{id} \leq n_{est}$ liegt $loss_3$ zwischen 0 und 1.

Bemerkung 3.9 (Eigenschaften von $loss_3$)

- $loss_3 = 0$ wenn $n_{true} = n_{est} = n_{id}$.
- Eine Definitionslücke entsteht nur, wenn $n_{true} = n_{est} = 0$ ist, also weder Extremstellen vorhanden sind noch geschätzt werden.
- $loss_3 = 1$ ergibt sich, wenn keine Extremstellen identifiziert werden, obwohl es sie gibt oder welche geschätzt werden.
- Der Verlust nutzt den Bereich von 0 bis 1 aus, insbesondere wird angemessen bestraft, wenn sehr viele überflüssige Extremstellen geschätzt werden.
- $loss_3$ ist unabhängig von n .
- Für $n_{true} = 0$ und $n_{est} > 0$ gilt immer $loss_3 = 1$, unabhängig davon, ob n_{est} groß oder klein ist.

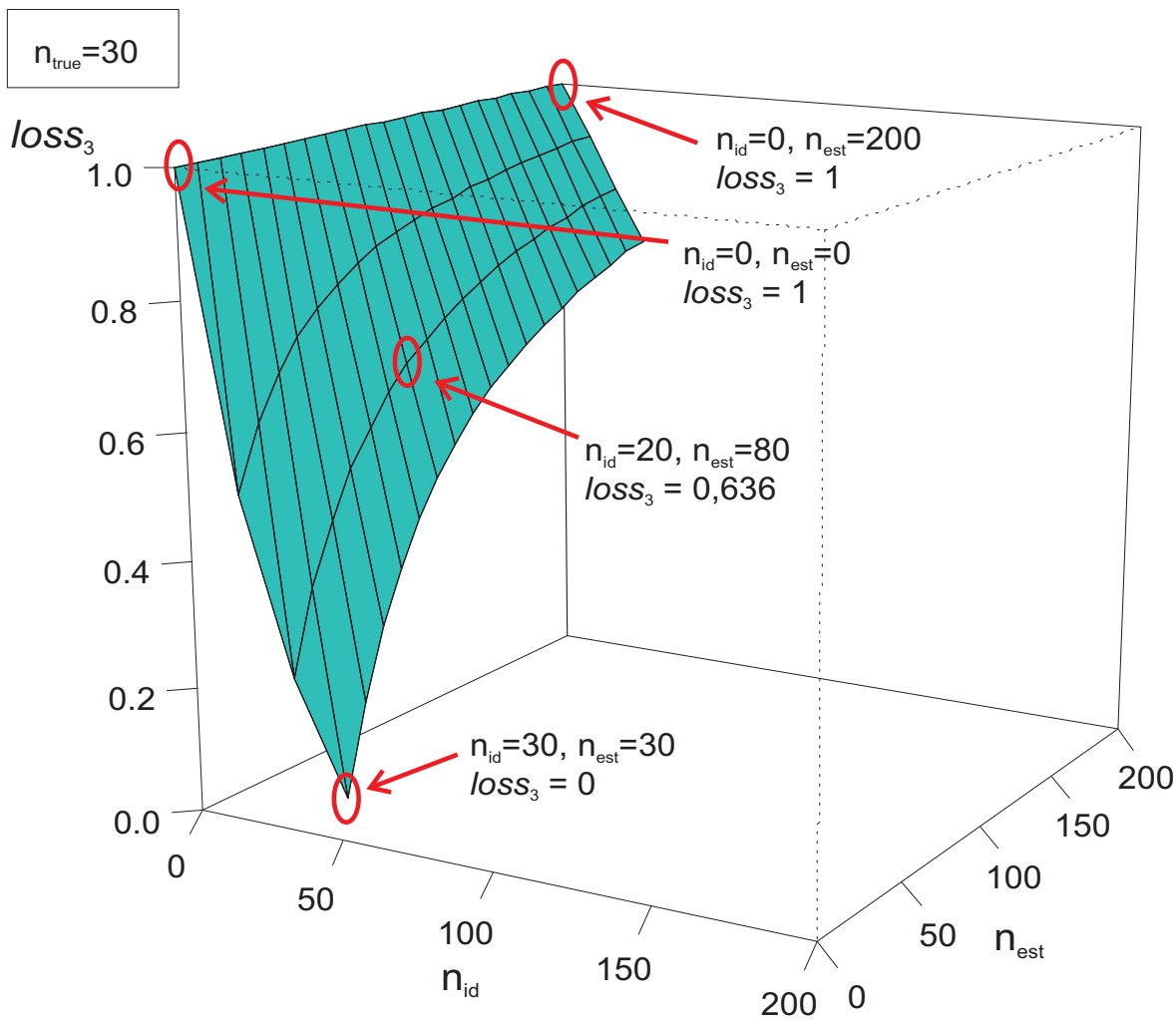


Abbildung 3.8: Ergebnisse von $loss_3$ bei variierenden n_{est} und n_{id}

Ergebnisse von $loss_3$ bei variierenden n_{est} und n_{id} sind in Abbildung 3.6 dargestellt. Es gilt $n_{true} = 30$ und der n_{est} -Wert wird wiederum von 0 bis auf 200, der n_{id} -Wert von 0 bis auf 30 in Zehnerschritten erhöht. An der dritten Achse sind für die jeweiligen n_{est} - und n_{id} -Werte die Ergebnisse von $loss_3$ aufgetragen. Für $n_{true} = n_{est} = n_{id} = 30$ ergibt sich $loss_3 = 0$. Wieder wird der Anstieg des $loss_3$ -Wertes bei zunehmender Abweichung von n_{est} und n_{id} vom wahren Wert n_{true} sichtbar. Es fällt auf, dass $loss_3 = 1$ für $n_{id} = 0$ gilt, unabhängig vom n_{est} -Wert. Gleichzeitig ist erkennbar, dass $loss_3$ für steigende n_{est} -Werte immer größere Werte annimmt und seinen ganzen Definitionsbereich von $[0; 1]$ ausfüllt.

Sensitivität und Spezifität

Bei Betrachtung der beiden genannten Fehler F_1 und F_2 fällt eine Ähnlichkeit zur Bewertung von Testverfahren durch *Sensitivität* und *Spezifität* auf (vgl. Zhou, Obuchowski und McClish 2002). Dabei werden Untersuchungseinheiten betrachtet, bei denen ein bestimmtes Merkmal auftritt oder nicht. Beispielsweise können die Einheiten Patienten sein und das Merkmal eine bestimmte Krankheit. Zusätzlich erfolgt ein Test, der für jeden Merkmalsträger angibt, ob das Merkmal auftritt. Die Testentscheidung muss nicht richtig sein. Für alle Untersuchungseinheiten wird notiert, ob sie in der Realität ein bestimmtes Merkmal aufweisen und welches Ergebnis der Test liefert. Die absoluten Häufigkeiten (a, b, c, d) der vier Zuordnungsfälle sind:

	Merkmal vorhanden ($M = 1$)	Merkmal nicht vorhanden ($M = 0$)	Summe
Test positiv ($T = 1$)	a (<i>richtig positiv</i>)	b (<i>falsch positiv</i>)	$a + b$
Test negativ ($T = 0$)	c (<i>falsch negativ</i>)	d (<i>richtig negativ</i>)	$c + d$
Summe	$a + c$	$b + d$	n

Es gibt verschiedene Kenngrößen, die als Maßzahlen für die Güte eines Tests verwendet werden. Dazu gehören die Sensitivität (Se) und die Spezifität (Sp). Die Sensitivität beschreibt die Wahrscheinlichkeit dafür, dass der Test eine Untersuchungseinheit, die ein Merkmal aufweist, auch richtigerweise als solche erkennt: $Se = P(T = 1|M = 1)$. Die Spezifität ist dagegen die Wahrscheinlichkeit dafür, dass der Test eine Untersuchungseinheit, die das Merkmal nicht aufweist, richtigerweise als solche erkennt: $Sp = P(T = 0|M = 0)$. Diese Wahrscheinlichkeiten werden meist durch die entsprechenden bedingten relativen Häufigkeiten geschätzt:

$$\hat{Se} = \frac{a}{a + c} \quad \text{und} \quad \hat{Sp} = \frac{d}{b + d}. \quad (3.8)$$

Eine Übertragung von Sensitivität und Spezifität auf die Extremstellenerkennung scheint zunächst vielversprechend, zumal es verschiedene Möglichkeiten gibt, Sensitivität und Spezifität zu einer Maßzahl zusammenzufassen (vgl. Zhou, Obuchowski und McClish 2002). Die Designpunkte t_1, \dots, t_n können als Merkmalsträger für das Merkmal t_i ist *Extremstelle* betrachtet werden. Der wahre Zustand resultiert aus der

Signalfunktion und das Ergebnis des Tests aus der Schätzfunktion. Damit lässt sich folgende Häufigkeitstabelle aufstellen:

Schätzfunktion	Signalfunktion		Summe
	Extremwert vorhanden	Extremwert nicht vorhanden	
Extremwert geschätzt	n_{id}	$n_{est} - n_{id}$	n_{est}
Extremwert nicht geschätzt	$n_{true} - n_{id}$	$n - n_{true} - n_{est} + n_{id}$	$n - n_{est}$
Summe	n_{true}	$n - n_{true}$	n

Dann ergeben sich als Schätzer für Sensitivität und Spezifität:

$$\hat{S}e = \frac{n_{id}}{n_{true}} \quad \text{und} \quad (3.9)$$

$$\hat{S}p = \frac{n - n_{true} - n_{est} + n_{id}}{n - n_{true}} = 1 - \frac{n_{est} - n_{id}}{n - n_{true}}. \quad (3.10)$$

Dabei ist die Sensitivität vergleichbar mit dem oben genannten Fehler \tilde{F}_1 , denn es gilt $\tilde{F}_1 = 1 - Se$. Die Sensitivität gibt an, wie hoch der Anteil richtig geschätzter Extremstellen an wahren Extremstellen ist. Die Spezifität gibt an, wieviele wahre *Nicht-Extremstellen* auch als solche erkannt werden. Darin liegt die Schwierigkeit dieses Ansatzes: Probleme treten auf, da hier alle Designpunkte, die keine Extremstellen sind, als *Nicht-Extremstelle* gezählt werden. Hier müsste äquivalent zur Definition der identifizierten Extremstellen über Toleranzintervalle um jede Extremstelle auch um jede Nicht-Extremstelle ein Intervall gelegt werden, bzw. die Designpunkte müssten in *Extremstellenintervalle* und *Intervalle ohne Extremstelle* eingeteilt werden. Dies könnte ähnlich wie bei einer Anwendung von Daten aus der Intensivmedizin durchgeführt werden, bei der die Zeit in alarmrelevante und nicht alarmrelevante Patientenzustände eingeteilt wird (vgl. Kuhls 2008). Da auf diese Weise nicht wie gewünscht ein sondern zwei Maßzahlen die Güte der Extremstellenerkennung beschreiben, und die zusätzliche Schwierigkeit der Wahl geeigneter Toleranzen für Nicht-Extremstellen auftritt, ist dieser Ansatz wenig praktikabel.

Stattdessen können zwei verschiedene Sensitivitäten betrachtet werden: die obige Sensitivität $Se = \frac{n_{id}}{n_{true}}$ bezüglich *Anteil richtig geschätzter an allen wahren Extremstellen* und $Se_2 = \frac{n_{id}}{n_{est}}$, also eine Sensitivität bezüglich des *Anteils richtig geschätzter an allen*

geschätzten Extremstellen. Dadurch ergibt sich jedoch nichts anderes als die beiden Fehler \tilde{F}_1 und \tilde{F}_2 , denn es gilt: $\tilde{F}_1 = 1 - Se$ und $\tilde{F}_2 = 1 - Se_2$.

Obwohl Parallelen zu Sensitivität und Spezifität vorhanden sind, lassen sich daraus keine weiteren sinnvollen Verlustfunktionen ableiten.

3.2.3 Vergleich und Auswahl der neuen Verlustfunktion

Im Weiteren werden die Verlustfunktionen PID , $loss_1$, $loss_2$ und $loss_3$ betrachtet und verglichen. Zunächst werden anhand eines Beispiels Ergebnisse dieser Verluste aufgezeigt, um einen Eindruck zu verschaffen, in welchen Verlustwerten verschiedene Szenarien resultieren. Das hilft einzuschätzen, welcher Verlust bei einer bestimmten Konstellation von n_{true} , n_{est} und n_{id} das angemessenste Ergebnis liefert.

Beispiel 3.10 *In Tabelle 3.1 sind verschiedene Szenarien aufgeführt, die verdeutlichen, welche Ergebnisse bei den vier Verlustfunktionen vorkommen. Die Werte für n_{true} , n_{est} und n_{id} sind hier exemplarisch. '–' bedeutet, dass der Verlust für die jeweilige Kombination nicht definiert ist. Für $loss_2$ werden die Ergebnisse für zwei verschiedene Stichprobenumfänge n angegeben. Für die anderen Verluste ist dies nicht nötig, da sie unabhängig vom Stichprobenumfang sind.*

- Zeile 1 und Zeile 8 zeigen Situationen, in denen alles richtig geschätzt wird. Ein Verlust sollte in diesen Situationen seinen besten Wert, d.h. im Allgemeinen 0, annehmen. Dies ist für die Verluste weitgehend erfüllt, nur in der Situation in Zeile 1 sind $loss_1$ und $loss_3$ nicht definiert, da $n_{true} = 0$ und $n_{est} = 0$.
- $loss_1$ ist nicht definiert, wenn $n_{true} = 0$ oder $n_{est} = 0$, also in den Zeilen 2–5 und 15.

Zeile	n_{true}	n_{est}	n_{id}	PID	$loss_1$	$loss_2$		$loss_3$
						$n = 200$	$n = 1000$	
1	0	0	0	0	–	0	0	–
2	0	20	0	20	–	0,05	0,01	1,00
3	0	50	0	50	–	0,13	0,03	1,00
4	0	150	0	150	–	0,38	0,08	1,00
5	20	0	0	-20	–	0,05	0,01	1,00
6	20	20	0	40	1,00	0,10	0,02	1,00
7	20	20	5	30	0,75	0,08	0,02	0,75
8	20	20	20	0	0	0	0	0
9	20	50	0	70	1,00	0,18	0,04	1,00
10	20	50	5	60	0,82	0,15	0,03	0,86
11	20	50	20	30	0,30	0,08	0,02	0,43
12	20	150	0	170	1,00	0,43	0,09	1,00
13	20	150	5	160	0,86	0,40	0,08	0,94
14	20	150	20	130	0,43	0,33	0,07	0,76
15	150	0	0	-150	–	0,38	0,08	1,00
16	150	20	0	-170	1,00	0,43	0,09	1,00
17	150	20	5	-160	0,86	0,40	0,08	0,94
18	150	20	20	-130	0,43	0,33	0,07	0,76
19	150	50	0	-200	1,00	0,51	0,10	1,00
20	150	50	5	-190	0,93	0,48	0,10	0,95
21	150	50	20	-160	0,73	0,40	0,08	0,80
22	150	150	0	300	1,00	0,76	0,15	1,00
23	150	150	5	290	0,97	0,73	0,15	0,97
24	150	150	20	260	0,87	0,66	0,13	0,87

Tabelle 3.1: Beispielergebnisse für die Verlustfunktionen PID , $loss_1$, $loss_2$ und $loss_3$ bei verschiedenen Werten von n_{true} , n_{est} und n_{id}

- In den Zeilen 11, 14 und 18 liefert $loss_1$ verhältnismäßig kleine Werte, insbesondere, wenn n_{true} oder n_{est} groß werden: in Zeile 14 und 18 gibt der Wert von $loss_1$ die Situation nicht gut wieder. $loss_3$ weist hier bessere Ergebnisse auf. Insgesamt füllt $loss_3$ den Definitionsbereich besser aus als $loss_1$.
- Die Ergebnisse von PID erfassen gut die verschiedenen Situationen des Beispiels. Für viele falsch geschätzte n_{est} und große Unterschiede von n_{true} und n_{est} ergibt sich auch ein großer Verlustwert (Zeile 12-24).
- $loss_2$ nimmt für $n = 200$ besonders in Situationen, in denen n_{true} und n_{est} relativ klein sind (Zeile 1 bis 11), sehr kleine Werte an. Für $n = 1000$ werden die Ergebnisse nochmals deutlich kleiner. So fällt das Ergebnis in keinem Fall größer aus als 0,15 (Zeile 22 und 23), zumeist sogar unter 0,1.
- Ist $n_{id} = 0$ (Zeilen 2-6, 9, 12, 15-16, 19 und 22) so gilt $loss_1 = loss_3 = 1$, beide Verluste nehmen also ihren schlechtesten Wert an (wenn sie definiert sind).
- Die Zeilen 2-4 zeigen die fehlende Differenzierung von $loss_3$ bei verschiedenen n_{est} und $n_{true} = 0$. In dieser Situation gilt immer $loss_3 = 1$.

In wieweit die vorgestellten Verlustfunktionen die gewünschten Eigenschaften aus Bemerkung 3.1 erfüllen, ist in Tabelle 3.2 dargestellt. '+' bedeutet erfüllt, '-' nicht erfüllt, 'o' teilweise erfüllt. '*' heißt, dass '+' unter bestimmten Voraussetzungen erreichbar ist. Das Kriterium der *Normierung* ist nur durch PID nicht erfüllt, damit fällt auch die Beurteilung von *Definitionsbereich* für PID negativ aus. Die weiteren gewünschten Eigenschaften werden von PID erfüllt, bis auf das Kriterium *Zusammenfassung*. Hier kann jedoch durch Verwendung des Betrages Abhilfe geschaffen werden, denn dann lässt sich sinnvoll ein Mittelwert über verschiedene PID -Werte für eine Testbed-Funktion bilden. Gleichzeitig wird aber die Erfüllung von *Wenig Informationsverlust* verloren gehen. Genau umgekehrt verhält es sich mit diesen beiden Kriterien bei den drei normierten Verlustfunktionen. Es ist möglich, jeweils dasselbe Vorzeichen wie bei PID zu verwenden, nämlich $sgn(n_{est} - n_{true})$ und damit das Kriterium *Wenig Informationsverlust* zu erfüllen. Wegen des Vorzeichens ist dann aber das Kriterium *Zusammenfassung*

	PID	$loss_1$	$loss_2$	$loss_3$
Normierung	–	+	+	+
Wenig Informationsverlust	+	–*	–*	–*
Keine Definitionslücken	+	–*	+	o*
Definitionsbereich	–	–	–	+
Bestrafung	+	o	–	+
Verhältnis	+	o	–	o*
Zusammenfassung	–*	+	+	+

Tabelle 3.2: Erfüllung der gewünschten Eigenschaften einer Verlustfunktion durch die vorgestellten Verluste.

nicht mehr erfüllt. An dieser Stelle muss entschieden werden, welches der Kriterien wichtiger ist.

Weil $loss_1$ ein Mittel aus zwei Fehlern ist, kann $loss_1 > 0,5$ nur vorkommen, wenn beide Fehler größer 0,5 sind. Daher erfüllt $loss_1$ die Kriterien *Bestrafung* und *Verhältnis* nur mittelmäßig gut. $loss_2$ schneidet durch die Abhängigkeit von n bei diesen Kriterien schlecht ab, weil dadurch oft sehr kleine Werte geliefert werden. $loss_3$ zeigt gute Ergebnisse. Der *Definitionsbereich* wird nur von $loss_3$ gut ausgefüllt, die anderen Verluste tendieren dazu, zu kleine Ergebnisse bei schlechten Schätzfunktionen zu liefern (vgl. auch Beispiel 3.10).

Die Definitionslücken von $loss_1$ und $loss_3$ lassen sich durch eine Sonderbehandlung der Fälle, in denen die Verluste nicht definiert sind, beheben. Dabei reicht es aus, für $loss_3$ nur den Fall $n_{true} + n_{est} = 0$ zu betrachten. In diesem Fall hat die Schätzfunktion jedoch keine Fehler bezogen auf die Extremstellen, da weder wahre noch geschätzte existieren. Somit ist es in diesem Fall sinnvoll, $loss_3 = 0$ zu definieren. Für $loss_1$ muss nur einer der beiden Werte n_{true} und n_{est} gleich 0 sein, um eine Definitionslücke hervorzurufen.

Insgesamt ist $loss_3$ von den drei normierten Verlusten derjenige mit den meisten erfüllten Forderungen. Daher dient $loss_3$ als Grundlage für einen normierten Extremstellenverlust. Die Definitionslücke wird als Sonderfall behandelt. Durch einen zweiten

Sonderfall für $n_{true} = 0$ und $n_{est} > 0$ lässt sich berücksichtigen, ob nur wenige oder viel zu viele Extremstellen geschätzt werden. Ohne die gesonderte Betrachtung wäre der Verlust immer 1, wenn $n_{true} = 0$ gilt. Der Fall $n_{true} > 0$ und $n_{est} = 0$ wird nicht als Sonderfall aufgenommen, obwohl auch für dieses Szenario der Verlust immer 1 ist. Es kann aber vorkommen, dass in einigen Datensätzen einer Simulation von einer festen Testbed-Funktion Extremstellen geschätzt werden, in anderen Datensätzen nicht. Eine Mittelwertbildung über beide Situationen wäre nicht mehr sinnvoll, wenn eine davon als Sonderfall behandelt würde.

Damit wird der normierte *Extremstellenverlust* (*ESV*) wie folgt definiert:

Definition 3.11 (Extremstellenverlust (ESV))

Der Extremstellenverlust ist definiert als:

$$ESV = \begin{cases} \operatorname{sgn}(n_{est} - n_{true}) \frac{n_{true} + n_{est} - 2n_{id}}{n_{true} + n_{est}} & n_{true} > 0, n_{est} > 0 \\ 0 & n_{true} = 0, n_{est} = 0 \\ \frac{n_{est}}{n} & n_{true} = 0, n_{est} > 0 \end{cases} \quad (3.11)$$

Für $n_{true} = 0$ und $n_{est} > 0$ ergibt sich als Verlust der Anteil geschätzter Extremstellen an allen Designpunkten. Dieser Fall tritt z.B. bei einer konstanten Signalfunktion auf. Das Vorzeichen $\operatorname{sgn}(n_{est} - n_{true})$ von *PID* wird hier aufgenommen, um die Herkunft des Verlustes durch eher zu viele oder eher zu wenige geschätzte Extremstellen erkennen zu können.

Der neue Verlust *ESV* ist im Kern gleich dem in Kapitel 3.2.2 vorgestellten $loss_3$ und hat daher auch die gleichen positiven Eigenschaften. Durch Schließen der Definitionslücken, Einführen des Vorzeichens und des Sonderfalls konnten die Nachteile von $loss_3$ verbessert werden.

In den Situationen von Beispiel 3.10 hat der ESV in den ersten vier Zeilen andere Werte als $loss_3$:

Zeile	n_{true}	n_{est}	n_{id}	ESV	
				$n = 200$	$n = 1000$
1	0	0	0	0	0
2	0	20	0	0,1	0,02
3	0	50	0	0,25	0,05
4	0	150	0	0,75	0,15

In den Zeilen 5 und 15 - 21 hat der ESV den Wert von $loss_3$ mit anderem Vorzeichen.

Damit erfüllt der neue Verlust alle genannten Kriterien:

- ESV ist normiert,
- ESV hat wenig Informationsverlust,
- ESV hat keine Definitionslücken,
- die Bestrafung *schlechter* Situationen ist durch ESV gut erfüllt,
- das Verhältnis verschiedener Verlustwerte in verschiedenen Situationen wird durch ESV richtig wiedergegeben,
- die Zusammenfassung verschiedener Verlustwerte in Simulationen kann durch Verwendung des Betrags von ESV gewährleistet werden.

TOLERANZEN FÜR EXTREMSTELLEN

Wie bereits in Kapitel 3 erwähnt, spielen die Toleranzen für Extremstellen eine große Rolle. Dieses Kapitel untersucht, welchen Einfluss sie auf den berechneten Extremstellenverlust haben. Die Frage, ob eine andere Toleranz zu einem gänzlich anderen Ergebnis einer Simulation führen kann, muss beantwortet werden.

Es muss sichergestellt sein, dass im Toleranzbereich um eine wahre Extremstelle keine weitere Extremstelle der gleichen Art liegt, damit die Definition 3.4 identifizierter Extremstellen eindeutig ist (vgl. Kapitel 3.1). Daher wird eine Toleranz für eine wahre Extremstelle wie folgt definiert:

Definition 4.1 (Toleranz)

Seien x_1, \dots, x_k mit $x_1 \leq x_2 \leq \dots \leq x_k$ die geordneten Extremstellen einer Funktion.

Eine Toleranz δ_i für die Extremstelle x_i ist jeder Wert, für den gilt:

$$x_i - \delta_i > x_{i-2} \quad \text{und} \quad x_i + \delta_i < x_{i+2} \quad \text{für } i \in \{3, \dots, k\}.$$

Für die jeweils ersten und letzten beiden Toleranzen muss gelten:

$$\begin{aligned} x_1 + \delta_1 &< x_3, & x_2 + \delta_2 &< x_4 \\ x_{k-1} + \delta_{k-1} &> x_{k-3}, & x_k + \delta_k &> x_{k-2}. \end{aligned}$$

Der Vektor der Toleranzen für alle Extremstellen einer Funktion wird mit $\boldsymbol{\delta}$ bezeichnet.

In Abbildung 4.1 ist ein Beispiel für den erlaubten Toleranzbereich um ein Maximum x_i rot dargestellt. Jeder Wert δ_i , der die Hälfte dieses Intervalls nicht übersteigt, ist als Toleranz erlaubt. Es wäre auch möglich, einen Toleranzbereich zu wählen, der nicht

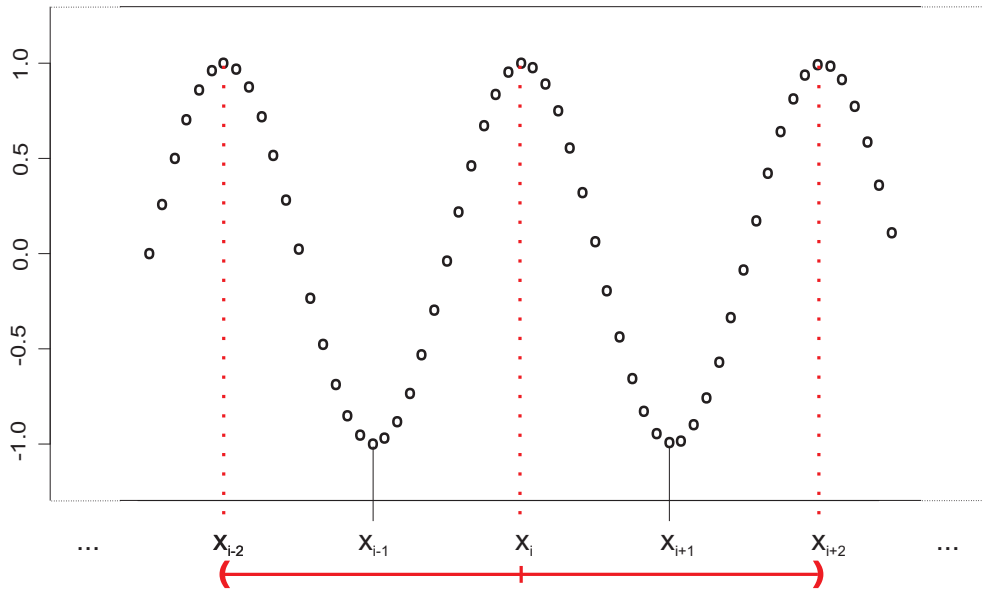


Abbildung 4.1: Erlaubtes Toleranzintervall um das Maximum x_i

symmetrisch um die Extremstelle liegt. Für eine Anwendung, bei der eine Abweichung der Extremstellen in eine Richtung nicht zu groß werden darf, in die andere Richtung aber größere Abweichungen erlaubt sind, kann ein nicht symmetrischer Toleranzbereich sinnvoll sein. In dieser Arbeit werden ausschließlich symmetrische Toleranzbereiche betrachtet.

Die Toleranz für verschiedene Extremstellen einer Funktion muss nicht konstant sein. Beispielsweise kann sie in einem Bereich, in dem wenige Extremstellen liegen, groß sein und in einem Bereich mit vielen, nah beieinanderliegenden Extremstellen klein.

Zur Untersuchung des Einflusses der Toleranz auf die Ergebnisse des Extremstellenverlustes werden insgesamt sechs verschiedene Toleranzen betrachtet.

Die kleinstmöglichen Toleranzen sind

$$\boldsymbol{\delta}^{\min} = (\delta_1^{\min}, \dots, \delta_k^{\min}) = (0, \dots, 0),$$

d.h. genau genommen gar keine Toleranzen. Eine Extremstelle kann nur identifiziert werden, wenn die Schätzung eine Extremstelle an genau derselben Stelle liefert.

Die Elemente $\delta_i^{\max}, i = 1, \dots, k$ des Vektors der nach Definition 4.1 maximalen Toleranzen $\boldsymbol{\delta}^{\max}$ können wie folgt bestimmt werden: Zunächst wird der Abstand einer

Extremstelle x_i zu den beiden benachbarten Extremstellen der gleichen Art, x_{i-2} und x_{i+2} , bestimmt. Der kleinere dieser Abstände wird als Toleranz verwendet:

$$\delta_i^{\max} = \min(|x_i - x_{i-2}|, |x_i - x_{i+2}|) - \varepsilon, \quad i = 3, \dots, k-2, \quad (4.1)$$

wobei

$$\begin{aligned} \delta_1^{\max} &= \min(x_1, |x_1 - x_3|) - \varepsilon, \\ \delta_2^{\max} &= \min(x_2, |x_2 - x_4|) - \varepsilon, \\ \delta_{k-1}^{\max} &= \min(|x_{k-1} - x_{k-3}|, |1 - x_{k-1}|) - \varepsilon, \\ \delta_k^{\max} &= \min(|x_k - x_{k-2}|, |1 - x_k|) - \varepsilon, \end{aligned} \quad \text{und } \varepsilon = \frac{1}{2n}. \quad (4.2)$$

Um zu verhindern, dass eine weitere Extremstelle im Toleranzbereich liegt, wird ε von den entsprechenden Abständen abgezogen. Für $0 < \varepsilon \leq \frac{1}{n}$ ist gewährleistet, dass die größtmögliche Anzahl an Designpunkten im Toleranzintervall liegt, daher wird hier $\varepsilon = \frac{1}{2n}$ festgelegt.

Davies, Gather und Weinert (2008) verwenden die Hälfte der in (4.1) und (4.2) gegebenen Toleranzen:

$$\delta^{(1)} = \frac{\delta^{\max}}{2}. \quad (4.3)$$

Dieser Wert wird noch einmal halbiert, um eine kleine Toleranz, die aber größer als δ^{\min} ist, in die Untersuchung mit einzubeziehen:

$$\delta^{(2)} = \frac{\delta^{(1)}}{2}. \quad (4.4)$$

Die so definierten Toleranzen δ^{\max} , $\delta^{(1)}$ und $\delta^{(2)}$ hängen sowohl von der Funktion als auch von der Anzahl der Designpunkte ab. Bei den meisten Funktionen, speziell bei den in Kapitel 2.1 vorgestellten Testbed-Funktionen, sind diese Toleranzen für verschiedene Extremstellen auch verschieden groß. In Tabelle 4.1 ist die erlaubte Abweichung bei maximaler Toleranz für 9 Testbed-Funktionen aus Kapitel 2 angegeben, die zehnte Funktion ist eine Konstante und hat keine Extremstellen. Bei der Doppler-Funktion mit 256 Designpunkten, d.h. $\frac{1}{n} = 0,0039$, ist beispielsweise die erlaubte Abweichung für die erste Extremstelle und δ^{\max} 0,0073 und damit ein Designpunkt in jede Richtung. Da die Extremstellen der Doppler-Funktion verschieden weit auseinander liegen, verändert sich auch die jeweilige Anzahl Designpunkte, die eine geschätzte Extremstelle abweichen

darf. Bei der Extremstelle Nr. 13 sind schon vier Designpunkte Abweichung erlaubt; bei den letzten beiden Extremstellen, Nr. 26 und Nr. 27, darf die Abweichung 56 bzw. 58 Designpunkte und damit relativ groß sein. Auch für andere Funktionen verändert sich die größtmögliche Toleranz, je nachdem für welche Extremstelle sie berechnet wird. Bei der Bumps-Funktion liegt die erlaubte Abweichung zwischen 4 und 35 ($n = 256$) bzw. 19 und 132 ($n = 1024$) Designpunkten. Bei anderen Funktionen aus Kapitel 2.1 bleibt die Toleranz für eine Stichprobengröße außer an den Rändern gleich: Bei der Sine-Funktion sind 4 ($n = 256$) bzw. 19 ($n = 1024$) Designpunkte Abweichung erlaubt. Bei der Wave-Funktion schwankt die erlaubte Abweichung um 20 ($n = 256$) bzw. 80 ($n = 1024$).

Es liegt nahe, statt dieser verschiedenen Toleranzen für alle Extremstellen einer Funktion dieselbe Toleranz zu verwenden. Eine solche feste Toleranz darf höchstens so groß sein wie das Minimum der maximalen Toleranzen gebildet über die Extremstellen. Beispielsweise wäre dies für die Doppler-Funktion $\delta_1^{\max} = 1$, für die Heavisine-Funktion $\delta_3^{\max} = 24$ bzw. $\delta_3^{\max} = 97$ (vgl. Tabelle 4.1). Der Vektor $\boldsymbol{\delta}^{\text{minimax}}$ dieser Toleranzen hat die Elemente

$$\delta_i^{\text{minimax}} = \min(\delta_1^{\max}, \dots, \delta_k^{\max}), i = 1, \dots, k.$$

Je nach Funktion kann die erlaubte Abweichung für $\boldsymbol{\delta}^{\text{minimax}}$ also sehr klein sein. Daher wird in dieser Arbeit als weitere mögliche Toleranz noch der Mittelwert der erlaubten maximalen Abweichungen genutzt. Für jede Extremstelle findet das Minimum von diesem Wert und das der maximalen Toleranz Verwendung:

$$\delta_i^{\text{mean}} = \min(\delta_i^{\max}, \bar{\delta}^{\max}), i = 1, \dots, k, \quad (4.5)$$

wobei $\bar{\delta}^{\max} = \frac{1}{k} \sum_{i=1}^k \delta_i^{\max}$. Der Vektor $\boldsymbol{\delta}^{\text{mean}}$ enthält die Toleranzen $\delta_i^{\text{mean}}, i = 1, \dots, k$. Tabelle 4.1 liefert auch die erlaubten Abweichungen für diese letzten beiden Toleranzen.

Diese sechs vorgestellten Toleranzen werden im Folgenden verglichen. Weitere Definitionen der Toleranz sind denkbar. Besonders bei Betrachtung einer einzelnen Testbed-Funktion sind häufig weitere sinnvolle Toleranzen möglich. Beispielsweise bei der Step-Funktion könnte nur der ganze konstante Abschnitt um den Extremwert herum als Toleranz dienen. Hier soll aber der Einfluss der Toleranz allgemein betrachtet werden, daher bezieht sich die Untersuchung nur auf die genannten Toleranzen.

Funktion		δ^{\max}						δ^{mean}	δ^{minimax}
Doppler $n = 256$	Extremwert Nr.	1	2	3	...	26	27		
	erlaubte Abw.	1	3	2	...	55	57	11	1
Doppler $n = 1024$	Extremwert Nr.	1	2	...	37	38	39		
	erlaubte Abw.	1	2	...	146	222	231	30	1
Bumps $n = 256$	Extremwert Nr.	1	2	3	...	20	21		
	erlaubte Abw.	6	6	4	...	6	6	11	4
Bumps $n = 1024$	Extremwert Nr.	1	2	...	19	20	21		
	erlaubte Abw.	30	26	...	20	30	29	46	19
Heavisine $n = 256$	Extremwert Nr.	1	2	3	4	5	6		
	erlaubte Abw.	31	87	24	39	24	31	40	24
Heavisine $n = 1024$	Extremwert Nr.	1	2	3	4	5	6		
	erlaubte Abw.	127	352	97	158	97	127	160	97
Blocks $n = 256$	Extremwert Nr.	1	2	3	...	8	9		
	erlaubte Abw.	18	24	18	...	58	22	30	18
Blocks $n = 1024$	Extremwert Nr.	1	2	3	...	8	9		
	erlaubte Abw.	75	101	75	...	234	91	121	75
Sine $n = 256$	Extremwert Nr.	1	2	3	...	78	79		
	erlaubte Abw.	3	5	4	...	4	5	4	3
Sine $n = 1024$	Extremwert Nr.	1	2	3	...	79	80		
	erlaubte Abw.	4	14	19	...	20	19	19	4
Step $n = 256$	Extremwert Nr.	1							
	erlaubte Abw.	116						116	116
Step $n = 1024$	Extremwert Nr.	1							
	erlaubte Abw.	468						468	468
Blip $n = 256$	Extremwert Nr.	1	2	3	4				
	erlaubte Abw.	78	84	51	50				66
Blip $n = 1024$	Extremwert Nr.	1	2	3	4				
	erlaubte Abw.	317	340	204	203				266
Wave $n = 256$	Extremwert Nr.	1	2	3	...	22	23		
	erlaubte Abw.	10	19	20	...	19	10	19	10
Wave $n = 1024$	Extremwert Nr.	1	2	3	...	22	23		
	erlaubte Abw.	44	80	82	...	80	44	79	44
Spikes $n = 256$	Extremwert Nr.	1	2	3	...	8	9		
	erlaubte Abw.	24	33	24	...	45	34	34	24
Spikes $n = 1024$	Extremwert Nr.	1	2	3	...	8	9		
	erlaubte Abw.	101	135	101	...	181	142	139	101

Tabelle 4.1: Erlaubte Abweichung in Designpunkten

Um den Einfluss genauer zu untersuchen, wurde eine Simulationsstudie durchgeführt. Sie soll die Frage beantworten, ob unterschiedliche Toleranzen zu gänzlich verschiedenen Simulationsergebnissen führen können. Dazu wird der Extremstellenverlust ESV für die verschiedenen Toleranzen und möglichst viele Regressionsszenarien bestimmt. Es werden die in Kapitel 2.1 vorgestellten Testbed-Funktionen, die Extremwerte enthalten, und alle Regressionsmethoden herangezogen. Für jede Testbed-Funktion erfolgt nach dem Modell (2.1) $y(t_i) = f(t_i) + \varepsilon(t_i)$, $i = 1, \dots, n$ die Erzeugung von jeweils 1000 Datensätzen mit den Stichprobengrößen $n = 256$ und $n = 1024$ durch Überlagerung mit standardnormalverteiltem Rauschen. Jede Regressionsmethode wird auf diese Datensätze angewendet. Nun kann für jede der sechs vorgestellten Toleranzen und jede dieser Simulationssituationen der ESV berechnet werden.

Beispiel 4.2 stellt zunächst für zwei Funktionen den Einfluss der Toleranz dar.

Beispiel 4.2 (Einfluss der Toleranz)

Die Abbildungen 4.2 und 4.3 zeigen für die Doppler-Funktion mit $n = 1024$ Designpunkten und die Heavisine-Funktion mit $n = 256$ Designpunkten, welchen Einfluss die vorgestellten Toleranzen auf die Ergebnisse des Extremstellenverlustes in einer Simulation haben. Dargestellt sind Boxplots der ESV-Ergebnisse aus 1000 Simulationen für fünf Regressionsmethoden. Die Anzahl der abgebildeten Methoden wurde beschränkt, damit die Ergebnisse für je sechs Toleranzen übersichtlich bleiben. Die Auswahl der Methoden WH, MDL, AWSg, LPI und TS ist beispielhaft.

Wird zur Beurteilung der Güte einer Methode der Median des ESV – ergänzt durch den gesamten Boxplot – herangezogen, so lassen sich die Regressionsmethoden in eine Reihenfolge bezüglich der Güte bringen. Die Methode mit dem kleinsten absoluten Wert des Median schneidet im Mittel am besten ab. Ein konzentrierter Boxplot mit kleinen Werten bedeutet, dass die zugehörige Methode besser ist als die eines Boxplots mit mehr Streuung und größeren positiven und negativen Werten.

Im ersten Graph von Abbildung 4.2 sind die Ergebnisse für δ^{\max} dargestellt. WH, AWSg und TS liefern mit Median $-0,2$ gute Ergebnisse, MDL und LPI mit einem Median um $0,4$ schlechtere. Durch Halbieren der Toleranz zu $\delta^{(1)}$ ändert sich diese Reihenfolge kaum

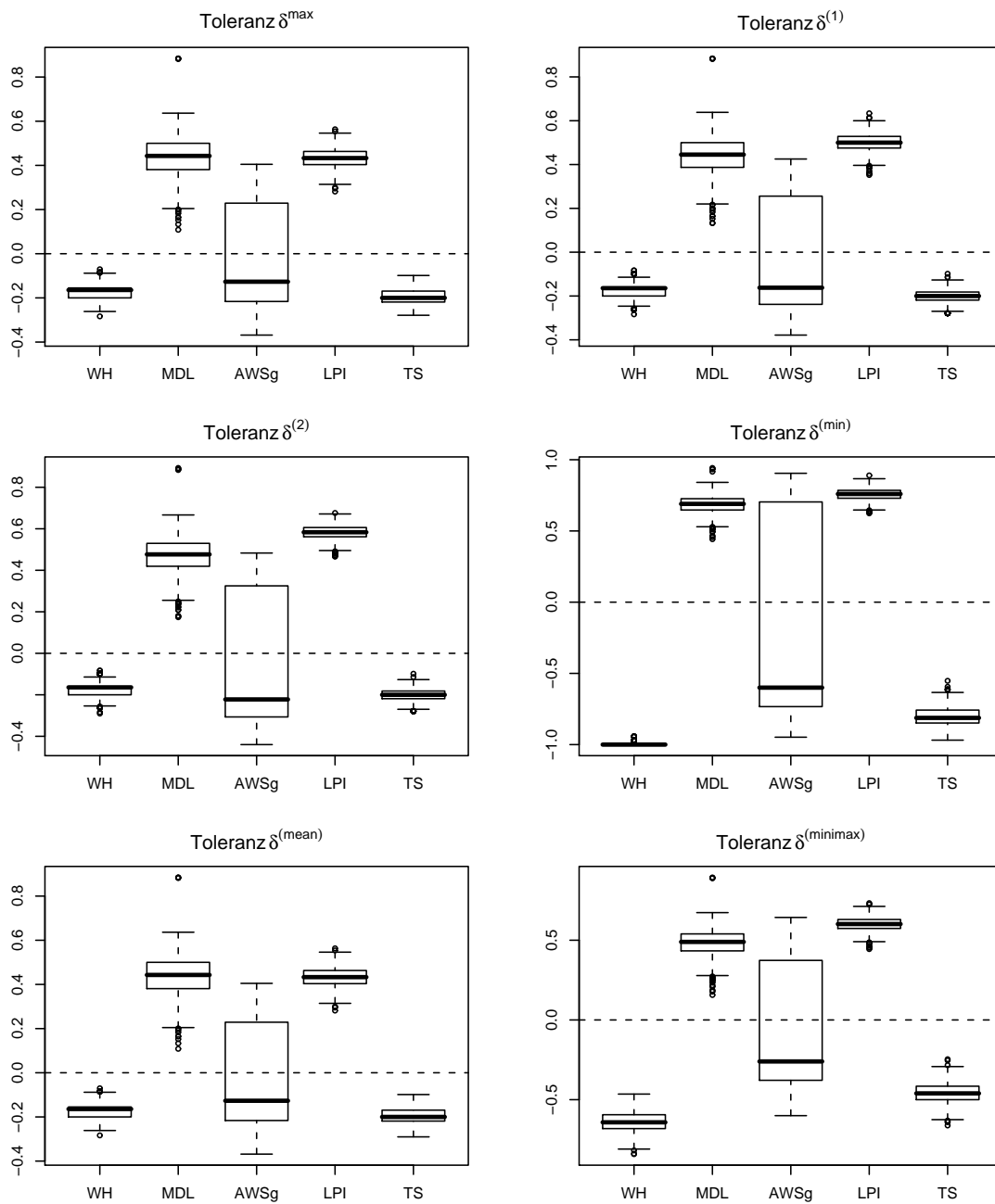


Abbildung 4.2: Extremstellenverluste von 1000 Simulationen der Doppler-Funktion mit $n = 1024$ für 6 verschiedene Toleranzen

und auch die ESV-Werte bleiben gleich. Erkennbar ist, dass MDL mit einem Median von etwa 0,4 etwas besser ist als LPI mit etwa 0,5. Auch bei weiterem Verkleinern der Toleranz zu $\delta^{(2)}$ ändert sich die Reihenfolge nicht. Die Mediane für die Methoden MDL und LPI werden größer, die der anderen Methoden bleiben mit -0,2 gleich. Die minimale Toleranz δ^{\min} lässt deutliche Veränderungen erkennen: der Median von AWSg liegt bei etwa -0,6 und ist damit der betragsmäßig kleinste. Mit einem Wert von etwa 0,65 ist der Median von MDL der betragsmäßig zweitkleinste. Danach folgen TS und LPI mit -0,75 bzw. 0,75 und den schlechtesten Median von -1 liefert WH. Bei Verwendung des Mittelwertes der Toleranzen ist das Muster der Boxplots gleich dem der maximalen Toleranz, und auch die ESV-Werte sind vergleichbar. Die Toleranz δ^{\minimax} ergibt die gleiche Reihenfolge wie δ^{\min} . Die Ergebnisse sind aber besser als bei δ^{\min} , betragsmäßig liegt kein Median über 0,6.

Im ersten Graphen von Abbildung 4.3 sind die Ergebnisse für δ^{\max} dargestellt. WH und TS schneiden hier als beste Methoden ab, der Boxplot ist nahezu komplett bei 0 konzentriert. Die zweitbeste Methode ist AWSg mit Median 0. LPI mit einem Median von 0,15 und MDL mit 0,55 liefern schlechtere Ergebnisse. Bei Verwendung von $\delta^{(1)}$ verändert sich die Reihenfolge der Methoden: AWSg mit Median 0 bleibt am besten, LPI mit 0,1 ist am zweitbesten, WH und TS sind mit einem Median von -0,2 gleich gut, und MDL ist mit 0,5 am schlechtesten. Ähnlich sind auch die Ergebnisse bei $\delta^{(2)}$. Nur die Methode WH liefert mit Median -0,4 ein schlechteres Ergebnis. Für die minimale Toleranz δ^{\min} sind die meisten Boxplots viel weniger konzentriert. WH, AWSg und LPI liefern sogar ESV-Werte über die ganze Bandbreite von -1 bis 1. Dabei ist AWSg mit Median 0 die Methode, die das beste Ergebnis aufweist. MDL mit Median 0,9 ist etwas besser als WH und TS. Die Toleranz δ^{mean} ergibt dieselben Ergebnisse wie δ^{\max} . Die Toleranz δ^{\minimax} zeigt noch einmal andere Ergebnisse, AWSg und WH haben Median 0, für LPI ist der Median 0,15, für TS -0,2 und für MDL 0,55.

Das Beispiel zeigt, dass je nach Situation die Wahl der Toleranz durchaus Einfluss auf das Ergebnis einer Simulation haben kann, aber auch, dass dieser

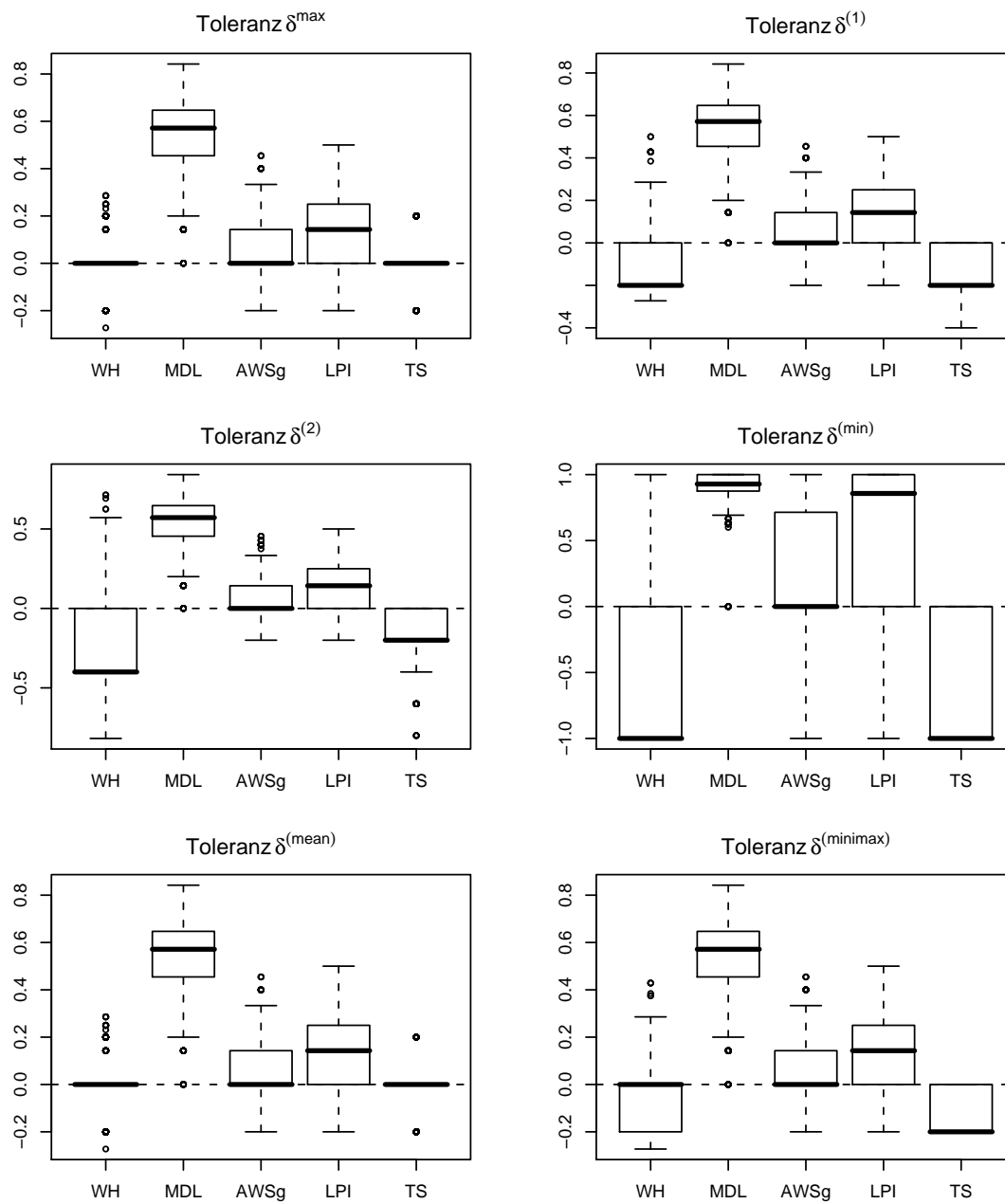


Abbildung 4.3: Extremstellenverluste von 1000 Simulationen der Heavisine-Funktion mit $n = 256$ für 6 verschiedene Toleranzen

Einfluss von der zugrunde liegenden Testbed-Funktion abhängt. Um die Ergebnisse der gesamten Simulation beurteilen zu können, wird der Mittelwert und der Median des Betrags von ESV über die jeweils 1000 Simulationen einer Testbed-Funktion gebildet. Von Interesse ist, wie alle 10 Methoden aus Kapitel 2.2 bei verschiedenen Toleranzen abschneiden. Dazu werden für jede Testbed-Funktion Ränge über die Mittelwerte und Mediane der $|\text{ESV}|$ -Werte gebildet. Rang 1 erhält die beste Methode, Rang 10 die schlechteste. Wenn eine Methode bei unterschiedlichen Toleranzen starke Differenzen in den Rängen aufweist, so zeigt das, dass die Wahl der Toleranz großen Einfluss hat. Treten nur geringe Differenzen, etwa von gar keinem oder nur einem Rang Unterschied auf, so hat die Wahl der Toleranz geringen Einfluss. Für jede Methode lässt sich die größte Abweichung in Rängen bestimmen. In Tabelle 4.2 sind diese Abweichungen für alle Testbed-Funktionen wie folgt zusammengefasst: Jeder Eintrag gibt diejenige Abweichung an, die in der zugehörigen Situation am größten ist. Ein Wert von 3 bedeutet also, dass es eine Methode gibt, bei der sich der Rang bei zwei verschiedenen Toleranzen um 3 unterscheidet, d.h. diese Methode hat zum Beispiel mit $\delta^{(1)}$ berechnet den Rang 2, mit δ^{\max} berechnet den Rang 5. Da insgesamt 10 Methoden betrachtet werden, ist die größtmögliche Abweichung für eine einzelne 9. Da sich herausgestellt hat, dass zum Teil nur die kleinste Toleranz δ^{\min} andere Ergebnisse liefert, sind die Rangabweichungen in Tabelle 4.2 zusätzlich ohne diese Toleranz dargestellt.

Deutlich wird, dass die Abweichungen für $n = 256$ etwas größer sind als für $n = 1024$. Außerdem ist ersichtlich, dass die minimale Toleranz häufig ein anderes Ergebnis als die anderen Toleranzen ergibt.

Tabelle 4.2 zeigt, dass bei der Doppler-Funktion der Einfluss der Toleranz sehr groß zu sein scheint. Die größte Abweichung beträgt 9 Ränge. Beispiel 4.2 hat aber gezeigt, dass bei differenzierterer Betrachtung der Ergebnisse mittels Boxplots die Unterschiede nicht so erheblich sind, wie der Wert der Abweichung vermuten lässt. Der Grund dafür liegt darin, dass für die Doppler-Funktion ebenso bei der Toleranz δ^{minimax} sehr kleine Werte auftreten (vgl. Tabelle 4.1). Wird auch diese Toleranz ausgelassen, so ergeben sich als größte Rangabweichungen beim Mittelwert 5 ($n = 256$) und 1 ($n = 1024$) und beim Median 5 ($n = 256$) und 1 ($n = 1024$). Die große Abweichung bei $n = 256$ lässt sich damit erklären, dass hier die Ergebnisse dicht beieinander liegen und ein kleiner Güteunterschied schon einen deutlich schlechteren Rang bedeuten kann.

Funktion	n	größte Abweichung in Rängen			
		Mittelwert		Median	
		gesamt	ohne δ^{\min}	gesamt	ohne δ^{\min}
Doppler	256	6	6	9	1
	1024	8	8	8	8
Bumps	256	2	2	2	0
	1024	1	0	2	0
Heavisine	256	5	5	7	5
	1024	1	0	2	0
Blocks	256	4	1	4	0
	1024	3	1	3	1
Sine	256	5	1	3	2
	1024	5	4	4	2
Step	256	4	2	5	5
	1024	5	1	3	0
Blip	256	1	0	1	0
	1024	0	0	1	0
Wave	256	4	3	8	2
	1024	1	0	1	0
Spikes	256	2	0	3	0
	1024	3	1	5	1

Tabelle 4.2: Größte Abweichung einer Methode in Rängen

Große Abweichungen von 5 oder mehr Rängen sind auch bei anderen Funktionen zu erkennen, beispielsweise Heavisine oder Wave jeweils mit $n = 256$ und anderen Szenarien. Bei differenzierterer Betrachtung der Ergebnisse anhand von Boxplots zeigt sich, dass dieser Unterschied häufig nur durch eine Methode, die mal gut und mal schlecht abschneidet, zustande kommt. Dadurch, dass die Ergebnisse aller Methoden verhältnismäßig dicht beieinander liegen, kann ein geringer Unterschied im Ergebnis schon zu einem anderen Rang führen.

Eine geringe Abweichung von bis zu zwei Rängen tritt in zahlreichen Situationen auf, insbesondere, wenn die minimale Toleranz nicht in die Betrachtung aufgenommen wird. Für die Ränge des Mittelwerts der Simulationen sind die Abweichungen für 13 der 18 Simulationsmodelle gering, bei 6 Modellen sind die Abweichungen sogar 0. Beim Median stellen sich 15 von 18 Rangabweichungen als gering dar und davon sind 9 gleich 0. Als

Fazit lässt sich feststellen, dass die Wahl der Toleranz die Abweichung einer Methode in Rängen nur schwach beeinflusst, solange die Toleranz nicht zu klein gewählt wird. In den Fällen, in denen Tabelle 4.2 größere Abweichungen zeigt, stellt sich bei differenzierterer Betrachtung der Ergebnisse heraus, dass sie alle sehr dicht beieinander liegen oder dass der Unterschied durch die Toleranz $\delta^{minimax}$ zustande kommt.

Bei Betrachtung aller simulierten Ergebnisse fällt auf, dass sich die ESV-Werte für δ^{\max} und δ^{mean} häufig gleichen. So sind die Ränge der Methoden in 18 von 20 Simulationsmodellen für beide Toleranzen gleich.

Durch die Untersuchung verschiedener Toleranzen ist es möglich, Aussagen über die Genauigkeit zu machen, mit der eine Methode die wahren Extremstellen schätzt. Bei Betrachtung von Abbildung 4.3 fällt auf, dass der Median der AWSg-Methode unabhängig von der Toleranz bei 0 liegt. Der Boxplot unterscheidet sich nur bei der minimalen Toleranz δ^{\min} wesentlich von den anderen. Das bedeutet, dass die AWSg-Methode die Extremstellen der Heavisine-Funktion in den meisten Fällen sehr dicht an der wahren Stelle schätzt. Im Gegensatz dazu lässt sich ableiten, dass die TS-Methode die Extremstellen nicht so genau schätzt, denn nur bei Verwendung der Toleranzen δ^{\max} und δ^{mean} ergibt sich ein Median von 0, sonst liegt der Wert etwa bei -0,2 oder bei der kleinsten Toleranz δ^{\min} sogar bei -1.

Für eine Simulationsstudie, in der verschiedene Methoden in Bezug auf das Erkennen von Extremstellen beurteilt werden, sollte eine Toleranz ausgewählt werden, die einen leichten Vergleich ermöglicht. Sie kann auch abhängig vom Ziel einer solchen Simulationsstudie bestimmt werden. Wenn für eine Anwendung die genaue Position einer Extremstelle wichtig ist, sollte eine kleinere Toleranz gewählt werden, als es zur Erkennung der richtigen Anzahl Extremstellen bedarf. In dieser Arbeit steht die Untersuchung des ESV im Vordergrund, die keinen sachlogischen Grund für den Einsatz einer bestimmten Toleranz liefert. Da nur durch Verwendung der minimalen Toleranz δ^{\min} das Ergebnis einer Simulationsstudie stark verändert wird, kann hier im Weiteren eine feste Toleranz verwendet werden. Diese wird auf die größtmögliche Toleranz δ^{\max} festgelegt, damit für die geschätzten Extremstellen auch große Abweichungen von den wahren Extremstellen erlaubt sind. Nach Definition 3.4 wird dennoch diejenige geschätzte Extremstelle mit dem kleinsten Abstand zur wahren als identifizierte Stelle gezählt.

EXTREMWERTVERLUST IN Y -RICHTUNG

In der nichtparametrischen Regression ist nicht nur von Interesse, ob Extremwerte an der richtigen Stelle geschätzt werden, sondern auch, ob ein Extremwert den richtigen Funktionswert hat. Beispielsweise kann eine Schätzfunktion zwar einen Extremwert an der richtigen Stelle aber mit falscher Höhe bestimmen. Als zusätzliche Information zum Extremstellenverlust ist es daher sinnvoll, einen weiteren Verlust in Y -Richtung zu betrachten.

In diesem Kapitel wird ein Extremwertverlust in Y -Richtung definiert, der als Zusatzinformation zum ESV einbezogen werden kann. Daher dienen für den Verlust diejenigen Extremwerte als Grundlage, die zu bereits identifizierten Extremstellen gehören (vgl. Kapitel 3). Die Anzahl identifizierter Extremstellen ist nur für $n_{id} > 0$ sinnvoll. Für Situationen mit $n_{id} = 0$ wird ausschließlich der Extremstellenverlust betrachtet. Jeder wahren identifizierten Extremstelle lässt sich genau eine geschätzte zuordnen. Durch die Benutzung der Toleranz δ in Definition 3.4 liegen identifizierte wahre und zugehörige geschätzte Extremstellen nicht zwangsläufig an demselben Designpunkt. Der Vektor \mathbf{x}_{true} mit Elementen $x_{i,true}, i = 1, \dots, n_{id}$ enthalte die identifizierten Extremstellen der wahren Funktion, der Vektor \mathbf{x}_{est} mit $x_{i,est}, i = 1, \dots, n_{id}$ die zugehörigen der geschätzten Funktion.

Im Gegensatz zu den Extremstellen gibt es für Extremwerte in Y -Richtung prinzipiell nur einen Fall, der die Schätzung verschlechtert, da nur die Abweichung vom wahren Funktionswert von Interesse ist. Diese Abweichung lässt sich bestimmen als

$$d_i^Y = |f(x_{i,true}) - \hat{f}(x_{i,est})|, \quad i = 1, \dots, n_{id}. \quad (5.1)$$

Um einen Verlust herzuleiten, gibt es nun zwei grundlegende Möglichkeiten.

- Es kann definiert werden, wann d_i^Y als akzeptabel gilt, d.h. wann der Abstand zwischen dem wahren und dem geschätzten Extremwert *klein genug* ist. Dann kann die Anzahl der entsprechenden Extremwerte für einen Verlust verwendet werden.
- Es ist möglich, d_i^Y über alle identifizierten Extremstellen zusammenzufassen und gegebenenfalls zu normieren.

Im Folgenden werden drei Verlustfunktionen für Extremwerte vorgestellt.

Toleranzabhängiger Extremwertverlust

Analog zum Extremstellenverlust können für die Y -Richtung die geschätzten Funktionswerte an den identifizierten Extremstellen mit den wahren verglichen werden, wobei auch hier eine Toleranz γ verwendet werden soll.

Ein geschätzter Extremwert $x_{i,est}$ hat *die richtige Höhe*, wenn gilt:

$$|f(x_{i,true}) - \hat{f}(x_{i,est})| \leq \gamma_i \quad (5.2)$$

Die Anzahl der Extremwerte mit richtiger Höhe sei mit n_{id}^Y bezeichnet.

Ein Extremwertverlust in Y -Richtung ist:

$$loss_1^Y = 1 - \frac{n_{id}^Y}{n_{id}} \quad (5.3)$$

Durch das Abziehen von 1 entsteht ein Verlust, der als besten Wert 0 annimmt, wenn alle Extremwerte die richtige Höhe haben. Dann stimmt n_{id}^Y mit n_{id} überein. Da $n_{id}^Y \leq n_{id}$, liegt $loss_1^Y$ zwischen 0 und 1. Hat kein Extremwert die richtige Höhe, nimmt $loss_1^Y$ den schlechtesten Wert 1 an. Wegen $n_{id} > 0$ liegen bei $loss_1^Y$ keine Definitionslücken vor.

Für die Untersuchung der Eigenschaften von $loss_1^Y$ werden fünf Arten von Toleranzen herangezogen:

- Es ist sinnvoll, die Toleranz in Y -Richtung von der Variation der wahren Funktion abhängig zu machen. Dazu kann der *Signal-to-Noise-Ratio (SNR)* einer Funktion verwendet werden. Die in dieser Arbeit verwendeten Testbed-Funktionen sind so normiert, dass $SNR = 7$ gilt (vgl. Kapitel 2.1).
- Anstatt des SNR der Funktion kann die Variation in den Extremwerten Verwendung finden. Dazu werden getrennt nach Art des Extremwerts die jeweiligen Standardabweichungen bestimmt, wobei $x_i^{\min}, i = 1, \dots, k^{\min}$, die Minima und $x_i^{\max}, i = 1, \dots, k^{\max}$, die Maxima bezeichnen:

$$\sigma^{\min} = \sqrt{\text{var}(f(x_1^{\min}), \dots, f(x_{k^{\min}}^{\min}))} \quad (5.4)$$

$$\sigma^{\max} = \sqrt{\text{var}(f(x_1^{\max}), \dots, f(x_{k^{\max}}^{\max}))} \quad (5.5)$$

σ^{\min} und σ^{\max} werden dann jeweils als Toleranz für Minima bzw. Maxima verwendet.

- Als weitere Möglichkeit kann die Differenz von größtem und kleinstem Funktionswert berechnet werden:

$$r = \max_i(f(x_{i,true})) - \min_i(f(x_{i,true})).$$

Dieser Wert wird hier kurz *Spannweite* genannt. Von der Spannweite dienen die festen Anteile 10%, 5% und 1% als Toleranz.

- Die minimale Toleranz ist auch in Y -Richtung 0. Da nun aber keine Designpunkte betrachtet werden, ist die Skala möglicher Abweichungen viel feiner. In Vorsimulationen zeigte sich, dass die Toleranz 0 dazu führt, dass keine Extremwerte die richtige Höhe haben.
- Die Toleranz kann außerdem abhängig von der Größe des wahren Extremwerts sein, d.h. von $|f(x_{i,true})|$. Von diesem Wert können wiederum Anteile als Toleranz dienen, beispielsweise geteilt durch den SNR, $\frac{|f(x_{i,true})|}{7}$, durch die Variation der Minima bzw. Maxima $\frac{|f(x_{i,true}^\ell)|}{\sigma^\ell}$, $\ell \in \{\min, \max\}$, oder durch die Spannweite $\frac{|f(x_{i,true})|}{r}$, $i = 1, \dots, n$.

n	SNR	Standardabweichung der				Spannweite		
		Minima		Maxima		10 %	5 %	1 %
		256	1024	256	1024			
Doppler	7	3.56	3.302	3.182	3.308	2.4	1.199	0.24
Bumps	7	2.08	2.02	13.13	10.48	5.32	2.66	0.53
Heavisine	7	6.60	6.54	3.26	3.17	2.36	1.18	0.24
Blocks	7	6.60	6.60	4.47	4.47	2.64	1.32	0.26
Konstante		keine Extremwerte vorhanden						
Sine	7	0.06	0.003	0.06	0.004	0.2	0.1	0.02
Step	7	nur ein Maximum				1.42	0.71	0.14
Blip	7	10.56	10.56	0.12	0.08	2.51	1.08	0.25
Wave	7	6.54	6.54	6.22	6.22	2.54	1.27	0.25
Spikes	7	1.34	1.34	14.24	14.28	4.39	2.19	0.44

Tabelle 5.1: Verschiedene Toleranzen für $loss_1^Y$

Tabelle 5.1 fasst die Toleranzen für die in Kapitel 2.1 vorgestellten Testbed-Funktionen zusammen. Angegeben sind dabei die Größen aller festen Toleranzen. Die Toleranzen, die sich für jeden Extremwert einer Funktion verändern, werden nicht gezeigt.

Beispiel 5.1 (Einfluss der Toleranzen auf $loss_1^Y$)

In Abbildung 5.1 sind die Ergebnisse von $loss_1^Y$ für die festen Toleranzen und fünf Regressionsmethoden bei der Funktion *Bumps* mit $n = 256$ dargestellt. Abbildung 5.2 zeigt die Ergebnisse bei der Funktion *Blocks* mit $n = 1024$ für die Toleranzen, die abhängig von der Größe des wahren Extremwerts variieren. Gezeigt werden Boxplots über 1000 Simulationen des Modells $y(t_i) = f(t_i) + \varepsilon(t_i)$, $i = 1, \dots, n$ mit standardnormalverteiltem Rauschen.

Der erste Graph stellt das Ergebnis für die Verwendung der Standardabweichung der *Minima* und *Maxima* dar. Die Ergebnisse sind hier mit Werten weitgehend zwischen

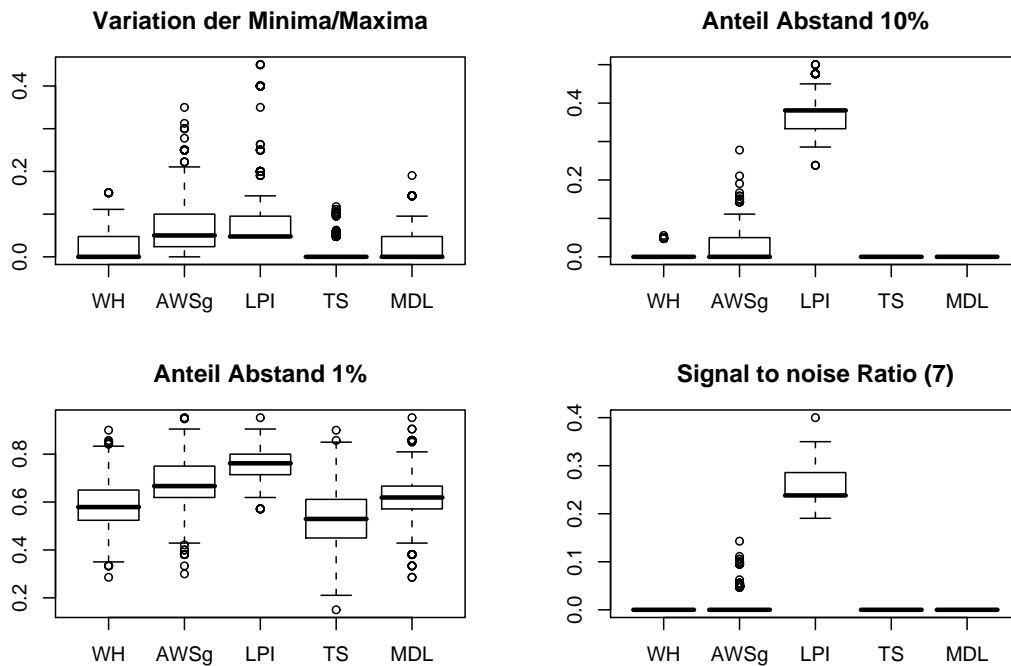


Abbildung 5.1: Verluste $loss_1^Y$ von 1000 Simulationen der Bumps-Funktion mit $n = 256$ und verschiedenen Toleranzen

0 und 0,1 gut. Die verschiedenen Methoden lassen sich bezüglich ihrer Güte im Schätzen der richtigen Höhe von Extremwerten sortieren: TS schneidet hier am besten ab, gefolgt von WH und MDL, dann AWSg und schließlich LPI. Im zweiten Graphen ist die Toleranz 10% der Spannweite der Bumps-Funktion. Dies führt dazu, dass drei Methoden (WH, TS und MDL) immer die richtige Höhe der Extremwerte schätzen. AWS schneidet etwas schlechter ab, LPI viel schlechter. Der dritte Graph zeigt die Ergebnisse für 1% der Spannweite, also eine sehr viel kleinere Toleranz als im zweiten Graphen. Die Methoden erzielen schlechtere Verluste und die Verlustwerte streuen stärker über die Simulationen. Im letzten Graphen sind die Ergebnisse für den SNR von 7 dargestellt, also eine größere Toleranz als im zweiten und dritten Graphen. Die Methoden schneiden hier daher wieder sehr viel besser ab.

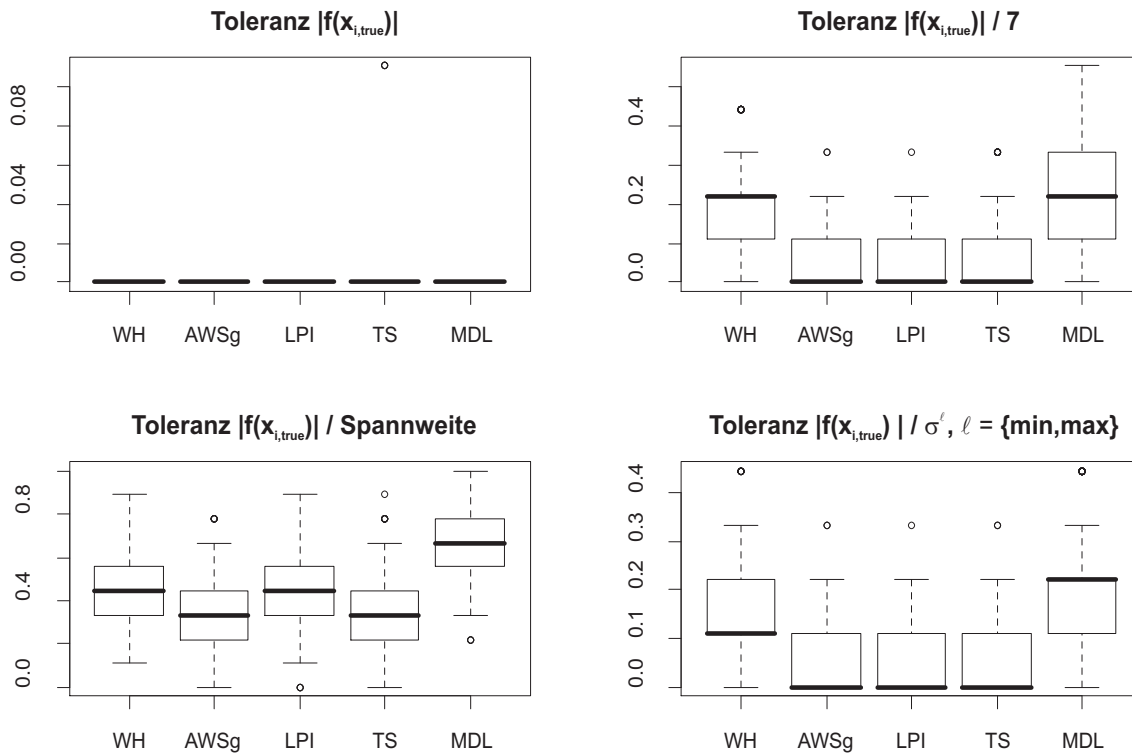


Abbildung 5.2: Verluste $loss_Y^1$ von 1000 Simulationen der Blocks-Funktion mit $n = 1024$ und verschiedenen Toleranzen

Abbildung 5.2 zeigt Ergebnisse für Toleranzen, die abhängig von der Höhe der einzelnen Extremwerte variieren. Die Toleranzen sind hier also für die einzelnen Extremwerte verschieden groß. Wird nur der Betrag des Extremwerts verwendet, so erkennen alle Methoden die Extremwerte in nahezu allen Fällen richtig. Bei Teilen durch den SNR von 7 sind Unterschiede in den Methoden zu erkennen. Wird durch die Spannweite, die bei Blocks 26,4 beträgt, geteilt, so ergibt sich ein ganz anderes Bild als für die anderen Toleranzen. Bei Teilen durch die Variation der jeweiligen Art der Extremwerte werden ähnliche Ergebnisse wie im zweiten Graphen erzielt, also wie beim Teilen durch den SNR.

Beispiel 5.1 macht deutlich, dass die Güte der verschiedenen Methoden für die gezeigten Situationen sehr stark von der verwendeten Toleranz abhängt. Dieses Ergebnis lässt sich auf andere Signalfunktionen übertragen. Im Gegensatz zur Toleranz für Extrem-

stellen hat die Toleranz hier also einen viel größeren Einfluss auf die Ergebnisse einer Simulation. Daher erscheint es sinnvoll, weitere mögliche Verluste zu entwickeln und zu untersuchen.

Verluste über den Abstand wahrer und geschätzter Extremwerte

Über das arithmetische Mittel der in Gleichung (5.1) eingeführten Abweichungen $d_i^Y = |f(x_{i,true}) - \hat{f}(x_{i,est})|$, $i = 1, \dots, n_{id}$ lässt sich ein *Verlust pro Extremwert* definieren:

$$loss_2^Y = \frac{\sum_{i=1}^{n_{id}} d_i^Y}{n_{id}}. \quad (5.6)$$

Bemerkung 5.2 (Eigenschaften von $loss_2^Y$)

- $loss_2^Y$ ist proportional zum absoluten Abstand von wahren und geschätzten Extremwerten und damit nicht normiert.
- Der Verlust ist 0, wenn die geschätzten Extremstellen genau die richtige Höhe haben und wird immer größer, je mehr die Y -Werte der geschätzten Extremstellen von denen der wahren abweichen. Die Division durch n_{id} führt zu einem durchschnittlichen Verlust pro identifizierter Extremstelle.
- $loss_2^Y$ ist translationsinvariant, ändert sich also nicht, wenn zu jedem wahren und geschätzten Extremwert ein konstanter Faktor addiert wird.
- Die absolute Abweichung des geschätzten Extremwerts vom wahren zählt in beide Richtungen gleich. Das Ergebnis ist also gleich, egal ob der geschätzte Extremwert vom wahren nach oben oder nach unten abweicht.
- $loss_2^Y$ hat keine zusätzlichen Parameter wie beispielsweise eine Toleranz.

Um einen normierten Verlust zu erhalten, kann der Abstand d_i^Y auf den jeweils größeren der eingehenden Werte bezogen werden. Auf diese Weise entsteht ein relativer Verlust

für jeden Extremwert. Für eine Schätzfunktion wird das arithmetische Mittel über die relativen Verluste der identifizierten Extremwerte verwendet:

$$\text{loss}_3^Y = \frac{1}{n_{id}} \sum_{i=1}^{n_{id}} \frac{|f(x_{i,true}) - \hat{f}(x_{i,est})|}{\max\{|f(x_{i,true})|, |\hat{f}(x_{i,est})|\}}, \quad (5.7)$$

wobei $\text{sign}(f(x_{i,true})) = \text{sign}(\hat{f}(x_{i,est}))$ für alle $i = 1, \dots, n_{id}$, gelten muss, damit loss_3^Y normiert ist.

Bemerkung 5.3 (Eigenschaften von loss_3^Y)

- loss_3^Y bezieht den Abstand von wahren und geschätzten Extremwerten auf den jeweils größeren der beiden Werte und ist dadurch normiert. Der Definitionsbereich ist $[0; 1]$.
- Der Verlust ist 0, wenn die geschätzten Extremstellen genau die richtige Höhe haben. Der Definitionsbereich wird in der Praxis nicht ausgeschöpft, da wahrer und geschätzter Extremwert in den meisten Fällen nah beieinander liegen. Der Verlust ist nahe 1, wenn die Werte stark voneinander abweichen, beispielsweise also einer 0 oder nahezu unendlich groß ist. Da die meisten Schätzmethoden Y-Werte ähnlich den Daten liefern, ist der zweite Fall bei normalverteiltem Rauschen irrelevant. Wird zu einer Testbed-Funktion beispielsweise Cauchy-Rauschen addiert, können extrem große Unterschiede zwischen wahren und geschätztem Extremwert auftreten.
- loss_3^Y ist skaleninvariant aber nicht translationsinvariant. Der Verlust ändert sich also nicht, wenn jeder wahre und geschätzte Extremwert mit einem konstanten Faktor multipliziert wird.
- Eine absolute Abweichung des geschätzten Extremwerts in Richtung des Mittelwerts der wahren Funktion wird durch den Verlust mit einem größeren Wert, d.h. schlechter beurteilt als die gleiche Abweichung in die andere Richtung. Das liegt daran, dass die Größe der Extremwerte wegen der Vorzeichenregelung in die

eine Richtung durch 0 beschränkt ist, in die andere Richtung aber nicht, also die Extremwerte betragsmäßig unendlich groß sein könnten.

- $loss_3^Y$ hat keine zusätzlichen Parameter wie beispielsweise eine Toleranz.

Qualitativer Verlust für die Glattheit

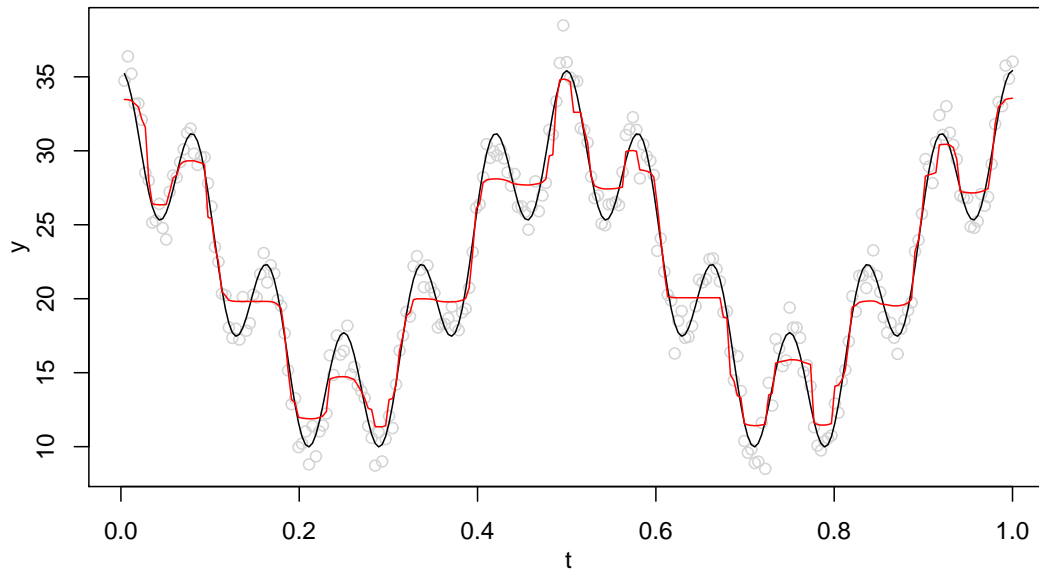


Abbildung 5.3: Schätzung der Wave-Funktion durch die STS-Methode, $n = 256$

Oft ist hauptsächlich von Interesse, ob eine Methode die Daten zu stark glättet, so dass Minima zu groß und Maxima zu klein geschätzt werden. Da im Allgemeinen Regressionsmethoden die *Daten* glätten, um das weiße Rauschen zu entfernen, sind einige überglättete Extremwerte in einer Schätzfunktion zu erwarten. Abbildung 5.3 zeigt dagegen ein Beispiel für eine zu stark glättende Schätzung. Grau sind aus der Wave-Funktion simulierte Daten dargestellt, schwarz die Wave-Funktion und rot die zugehörige Schätzung mittels der STS-Methode. Offensichtlich glättet die Methode die Extremwerte der Wave-Funktion zu stark. Um ein solches Verhalten in einen Verlust zu fassen, wird zunächst für jeden Extremwert einzeln erfasst, ob er zu stark geglättet

wurde oder nicht:

$$g_i = \begin{cases} 0 & f(x_{i,true}^{\max}) < \hat{f}(x_{i,est}^{\max}) \\ 1 & f(x_{i,true}^{\max}) > \hat{f}(x_{i,est}^{\max}) \\ 0 & f(x_{i,true}^{\min}) > \hat{f}(x_{i,est}^{\min}) \\ 1 & f(x_{i,true}^{\min}) < \hat{f}(x_{i,est}^{\min}) \end{cases} \quad (5.8)$$

Als *Glattheitsverlust* wird die Summe über alle identifizierten Extremwerte einer Funktion gebildet und durch deren Anzahl geteilt:

$$loss_g^Y = \frac{\sum_{i=1}^{n_{id}} g_i}{n_{id}} \quad (5.9)$$

In diesen Glattheitsverlust geht die absolute Abweichung wahrer und geschätzter Extremwerte nicht ein. Der Verlust zeigt die Häufigkeit an, mit der Extremwerte überglättet werden. Ein kleiner Wert kann sowohl durch eine gute als auch durch eine zu wenig glättende Schätzung zustande kommen.

Beispiel 5.4 zeigt für die Wave-Funktion die Ergebnisse von $loss_2^Y$, $loss_3^Y$ und $loss_g^Y$ in einer Simulation.

Beispiel 5.4 (Verluste der Wave-Funktion)

In Abbildung 5.4 sind die Ergebnisse der drei Verlustfunktionen $loss_2^Y$, $loss_3^Y$ und $loss_g^Y$ für die 10 Regressionsmethoden über 1000 Simulationen der Wave-Funktion mit $n = 256$ und standardnormalverteiltem Rauschen in Boxplots dargestellt. Zusätzlich ist für jede Methode der durchschnittliche $|ESV|$ -Wert gegeben, da die Verluste nur mit diesem zusammen betrachtet werden sollten.

Die Abbildung macht deutlich, dass $loss_2^Y$ und $loss_3^Y$ die Methoden bezüglich ihrer Güte ähnlich beurteilen. Gemessen mit beiden Verlusten zeigen LPI, AWSg und MDL die kleinsten Abweichungen zwischen wahren und geschätzten Extremwerten, STS und POT die größten.

Auffällig ist dabei, dass die Methoden mit kleinen $loss_2^Y$ - und $loss_3^Y$ -Werten große ESV-Werte haben. Offensichtlich liefern diese Methoden zu viele Extremwerte in der

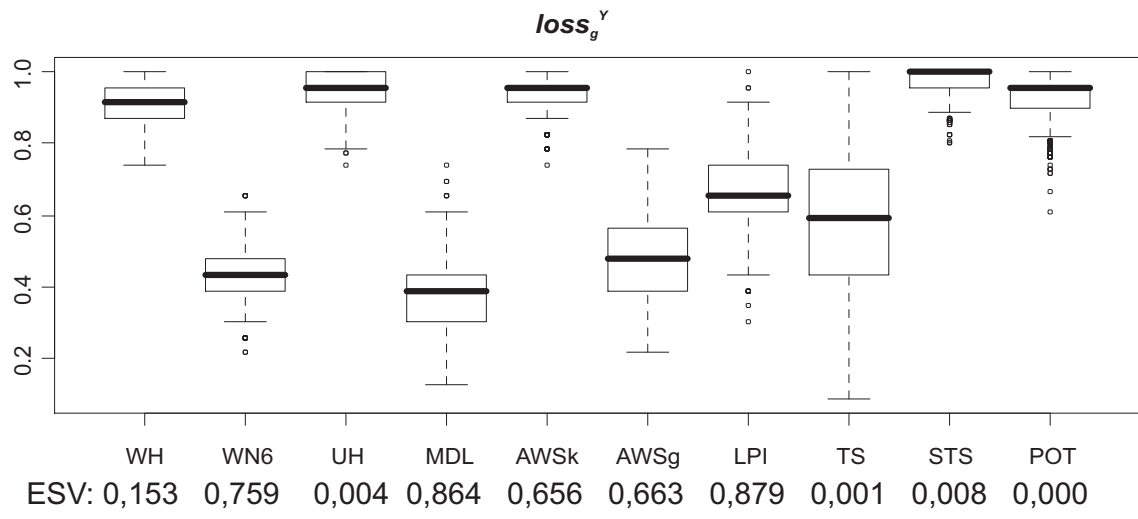
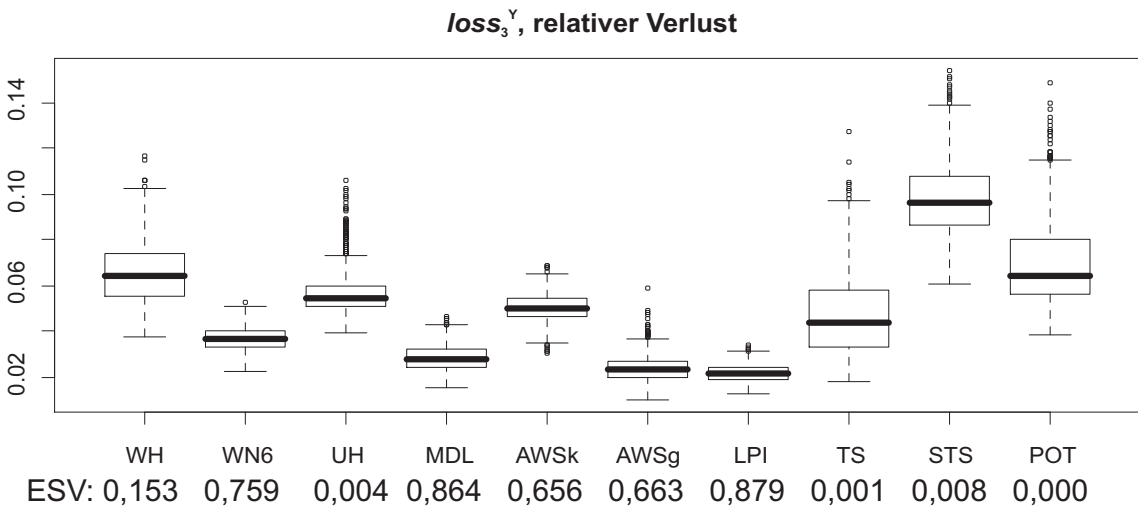
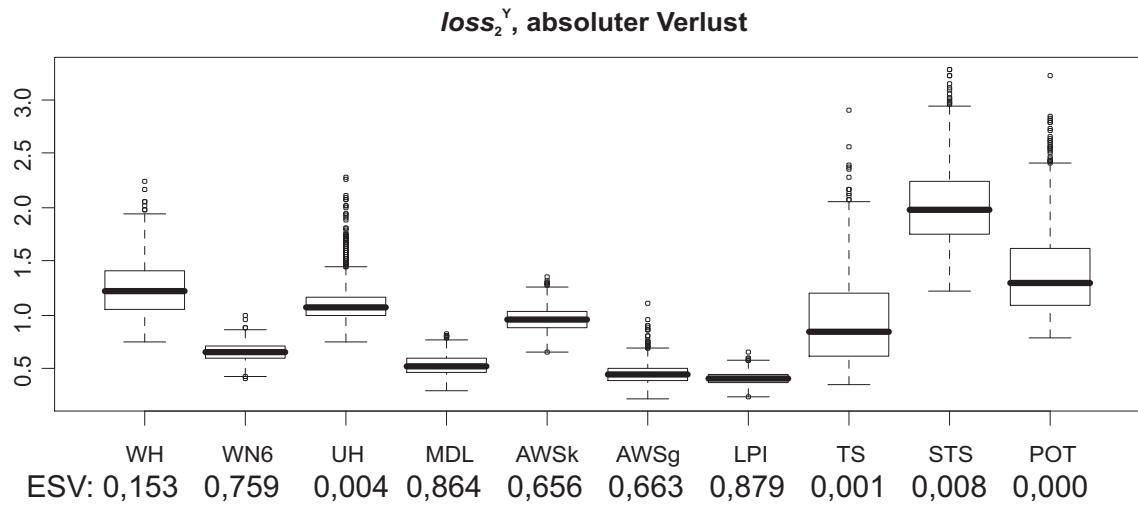


Abbildung 5.4: $loss_2^Y$, $loss_3^Y$ und $loss_g^Y$ von 1000 Simulationen der Wave-Funktion mit $n = 256$.

Schätzfunktion. Die richtig geschätzten Extremwerte haben dabei die richtige Höhe. Die Methoden UH, TS, STS und POT haben einen kleinen ESV-Wert. Durch die Verluste $loss_2^Y$ und $loss_3^Y$ wird deutlich, dass die Methoden UH und TS dabei die Höhe der Extremwerte besser erfassen, als STS und POT.

Die Boxplots von $loss_9^Y$ zeigen, dass neben STS und POT auch WH, UH und AWSk in nahezu allen Simulationsläufen zu glatte Schätzungen liefern. Mit Ausnahme von STS sind das stückweise konstante Schätzungen. Die einzige solche Methode, für die das nicht der Fall ist, ist TS.

Insgesamt schneiden daher UH und TS am besten ab, da sie einen sehr geringen ESV-Wert aufweisen und bezüglich der Höhe im mittleren Bereich liegen. Dabei zeigt UH eine stärkere Überglättung als TS.

Die ähnlichen Ergebnisse von $loss_2^Y$ und $loss_3^Y$ bezüglich der Beurteilung der Methoden aus dem Beispiel lassen sich auf die meisten Testbed-Funktionen übertragen. Eine Betrachtung beider Verluste liefert also keine zusätzlichen Erkenntnisse. Ergebnisse für alle Testbed-Funktionen werden daher in Kapitel 7 für $loss_2^Y = \ell^Y$ gegeben. Auch Ergebnisse bezüglich der Überglättung werden anhand von $loss_9^Y$ gezeigt.

ALTERNATIVE HERANGEHENSWEISE: DER SOFTLOSS

Dieses Kapitel stellt eine gänzlich andere Herangehensweise an ein Gütekriterium bezüglich des Erkennens von Extremstellen und -werten dar. Ausgehend vom Vergleich zweier Funktionen mittels L_p -Normen (vgl. Kapitel 2.3) sollen auch hier die Funktionen mit einer L_1 -Norm verglichen werden, jedoch in Bezug auf ihre Extremwerte. Dafür werden die Funktionen zunächst *auf ihre Extremwerte reduziert*, gefaltet und daraufhin mittels der L_1 -Norm verglichen.

Dieser neue Verlust wird *Softloss* genannt. Der Name resultiert daraus, dass der Verlust im Gegensatz zum Extremstellenverlust nicht so streng einzelne Extremstellen der Schätzfunktion als *identifiziert oder nicht* beurteilt.

Die Abbildungen 6.1 und 6.2 stellen zunächst das Vorgehen für den Softloss anhand der Doppler- und der Wave-Funktion dar. Der Graph zeigt jeweils die Funktion f und die *auf ihre Extremstellen und -werte* reduzierte Kurve, bezeichnet mit g_{Ext}^f . Diese Kurve enthält die Extremstellen und -werte der Funktion f und ist sonst 0. Der dritte Graph beschreibt schließlich die Funktion g , die aus der Faltung von g_{Ext}^f mit einem Gausskern

$$\phi = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

mit Mittelwert 0 und Standardabweichung σ entsteht. Abgebildet ist sie schwarz für $\sigma = 1$ und rot für $\sigma = 3$.

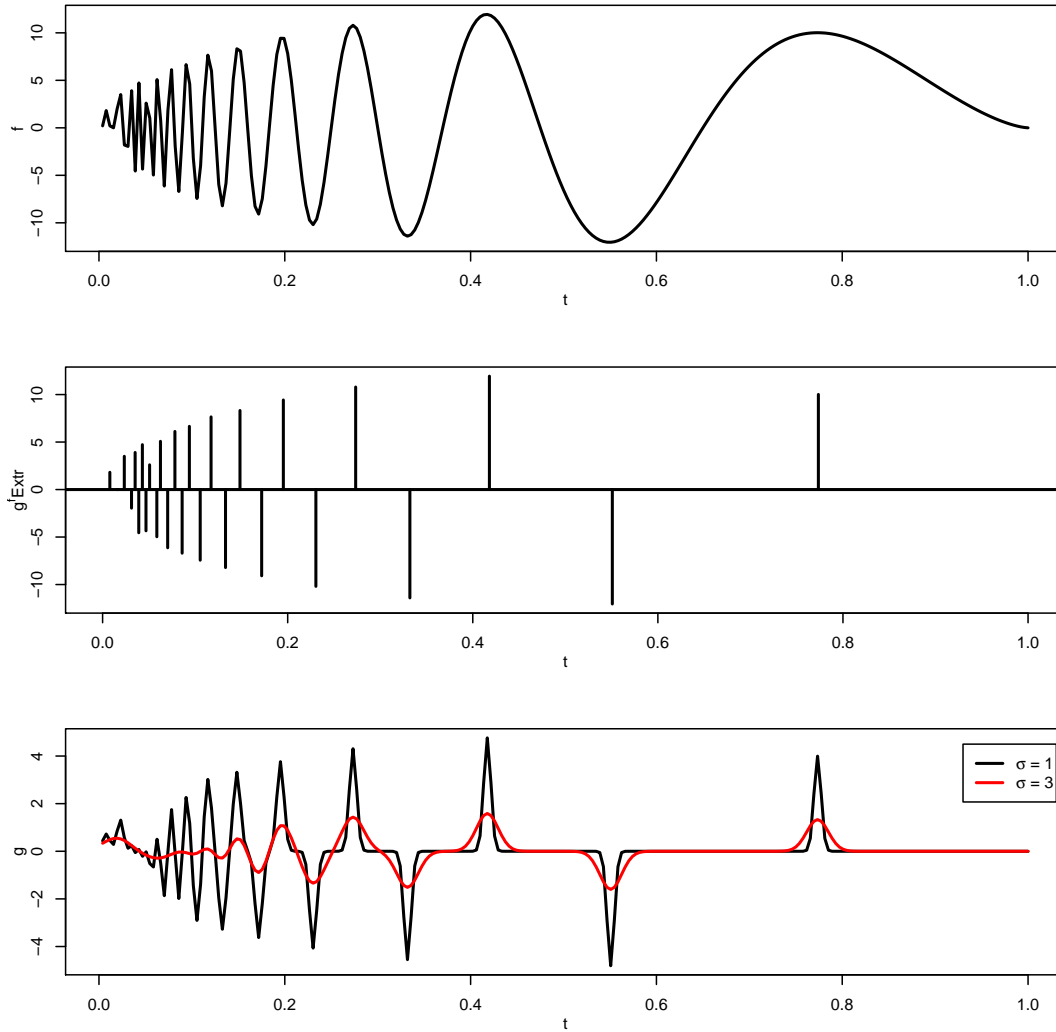


Abbildung 6.1: Doppler-Funktion, zugehöriges g^f_{Extr} und Kurve gefalteter Extremwerte

Seien x_1, \dots, x_k die Extremstellen einer Funktion f . Dann lässt sich g^f_{Extr} bestimmen durch

$$g^f_{Extr}(x) = \sum_{i=1}^k f(x_i) \delta_{x_i}(x),$$

wobei δ_{x_i} eine Punktmasse (Dirac-Impuls) an der Stelle des i -ten Extremwerts bezeichnet. g^f_{Extr} ist sozusagen das Gerüst der Funktion f , bezogen auf deren Extremwerte und keine Funktion im eigentlichen Sinne. Durch die Faltung mit einem Gausskern entsteht

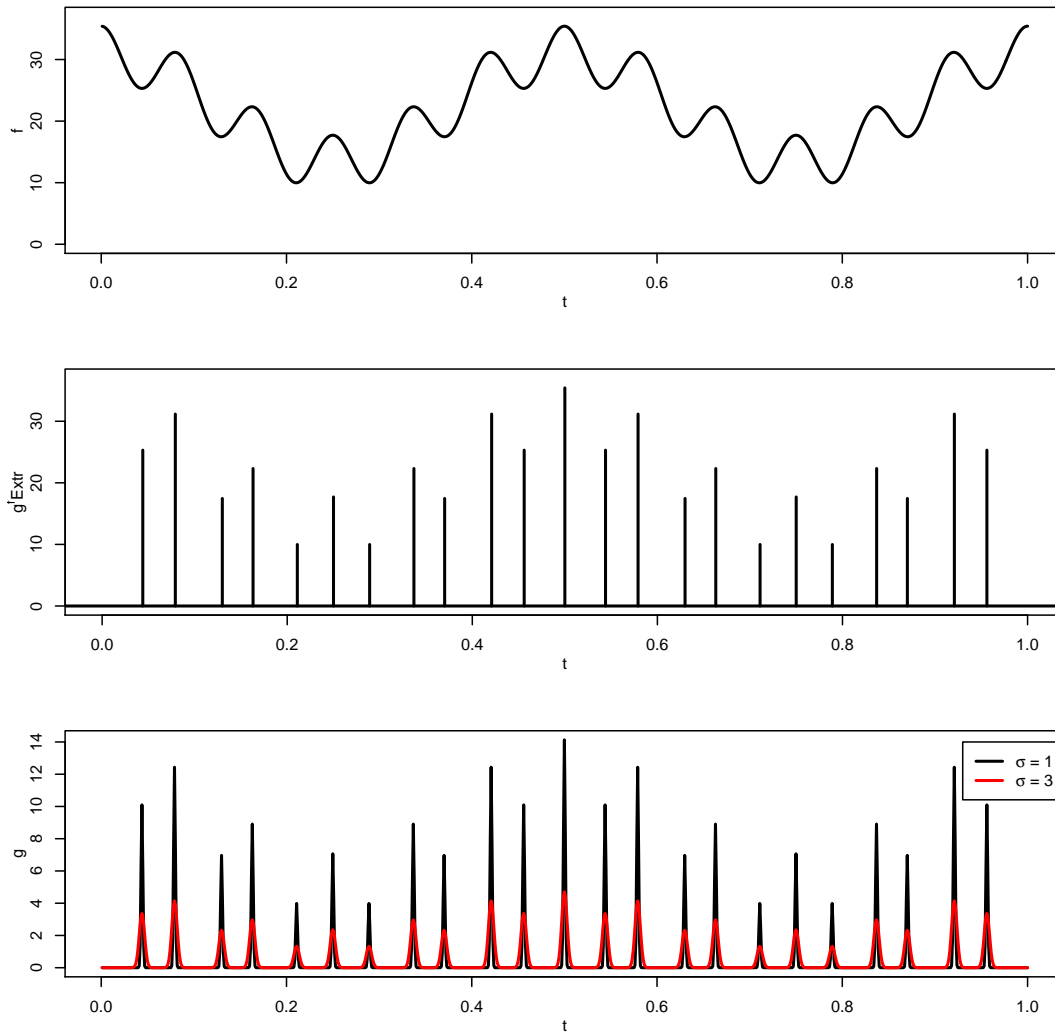


Abbildung 6.2: Wave-Funktion, zugehöriges g_{Extr}^f und Kurve gefalteter Extremwerte

wieder eine Funktion:

$$g = \phi * g_{Extr}^f(x) = \sum_{i=1}^k f(x_i) \phi(x - x_i). \quad (6.1)$$

Für eine Schätzfunktion \hat{f} werden entsprechend $g_{Extr}^{\hat{f}}$ und \hat{g} mit demselben Wert für σ unter Verwendung der Extremstellen von \hat{f} bestimmt. Anstatt eines Vergleichs von f und \hat{f} können nun g und \hat{g} verglichen werden. Diese Vorgehensweise ermöglicht die Beurteilung der geschätzten Extremwerte, ohne die Extremstellen in *identifizierte* und

nicht identifizierte einzuteilen.

Abbildung 6.1 verdeutlicht die Wirkung der Faltung auf die auf ihre Extremwerte reduzierte Doppler-Funktion. Bei verwendeter Standardabweichung von $\sigma = 3$, dargestellt in rot, heben sich Minima und Maxima der Doppler-Funktion im Bereich von $t = 0$ bis etwa $t = 0,1$ gegenseitig auf. Die folgenden Extremwerte werden in ihrer Höhe deutlich reduziert. Bei kleinerer Standardabweichung von $\sigma = 1$ sind diese beiden Effekte nicht so stark. Die Wave-Funktion in Abbildung 6.2 hingegen hat deutlich voneinander getrennte Extremstellen, die alle positive Werte aufweisen. Die durch Faltung mit Standardabweichung $\sigma = 3$ und $\sigma = 1$ entstehenden Kurven enthalten alle Extremwerte der Wave-Funktion, diese haben nur eine geringere Höhe.

Beispiel 6.1 (Wirkung der Faltung auf Testbed- und Schätzfunktion)

Beispiele für aus Schätzfunktionen gewonnene Kurven \hat{g} finden sich in den Abbildungen 6.3 und 6.4. In den oberen Graphen von Abbildung 6.3 ist die Step-Funktion schwarz dargestellt und rot eine Schätzung durch die WN6- bzw. die AWSg-Methode. Die unteren Graphen zeigen dementsprechend die Kurve g schwarz und die zugehörigen Kurven \hat{g} rot. Die Standardabweichung der verwendeten Gaussfunktion ist $\sigma = 1$. In den Abbildungen ist zu erkennen, dass die Abweichungen zwischen f und \hat{f} durch g und \hat{g} wiedergegeben werden. Die zahlreichen überflüssigen Extremwerte in der Schätzung durch die WN6-Methode führen zu ebenso zahlreichen Extremwerten in der zugehörigen Funktion \hat{g} . Der gefaltete wahre Extremwert der Step-Funktion überschneidet sich dabei mit einem der geschätzten Extremwerte.

Die AWSg-Methode zeigt weniger überflüssige Extremwerte, allerdings überschneidet sich der gefaltete wahre Extremwert auch nicht mit einem geschätzten.

Ähnliche Beobachtungen lassen sich in Abbildung 6.4 machen. Dargestellt ist die Blip-Funktion und jeweils eine Schätzung durch WN6 und AWSg. Wiederum zeigt die WN6-Methode zahlreiche überflüssige Extremwerte, die auch im zugehörigen \hat{g} deutlich werden. Gleichzeitig ist ersichtlich, dass der erste Extremwert der Blip-Funktion bei etwa

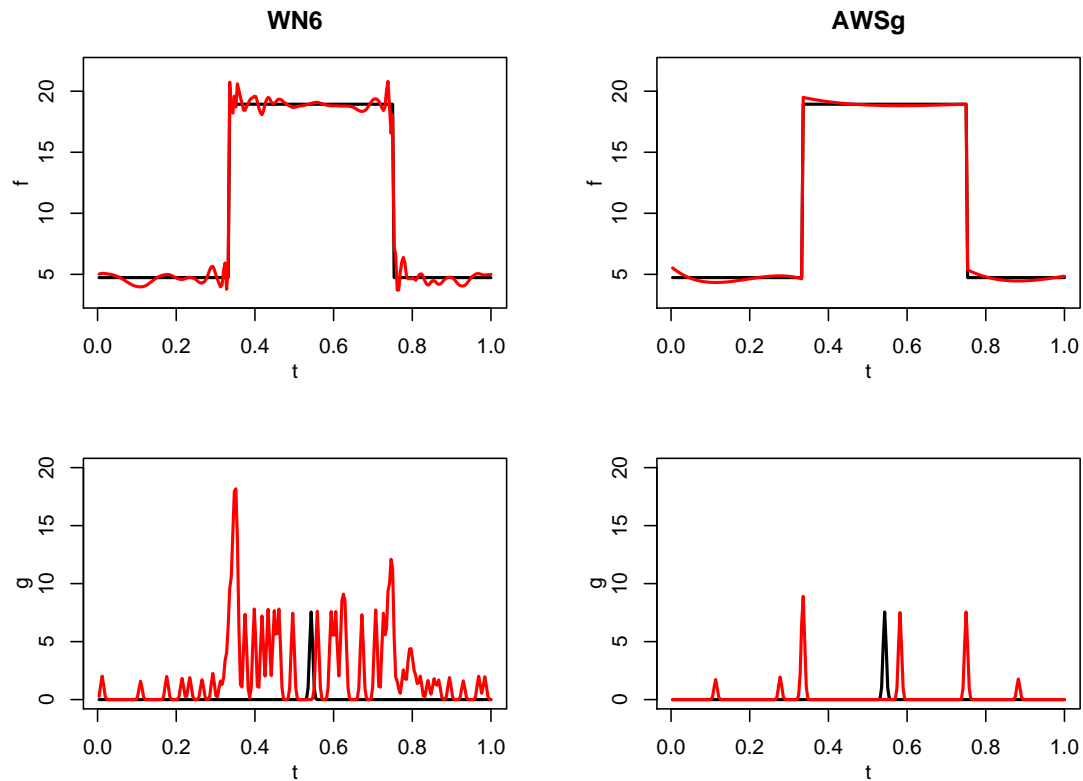


Abbildung 6.3: Step-Funktion und Schätzungen durch WN6 und AWSg und jeweils zugehörige Funktionen g (schwarz) und \hat{g} (rot)

$t = 0,3$ durch WN6 richtig geschätzt wird, da sich hier g und \hat{g} genau überschneiden. Der zweite Extremwert bei etwa $t = 0,45$ wird durch WN6 an einer leicht abweichenden Stelle geschätzt. Der dritte Extremwert bei etwa $t = 0,8$ wird in \hat{g} durch überflüssige Extremwerte überlagert.

Die AWSg-Methode liefert hingegen auf den ersten Blick keine überflüssigen Extremwerte. \hat{g} zeigt für den mittleren Extremwert jedoch eine Überlagerung von nah beieinander liegenden Extremwerten. Insgesamt weicht \hat{g} dadurch in der Höhe stark von g ab, während die anderen beiden Extremwerte gut getroffen werden.

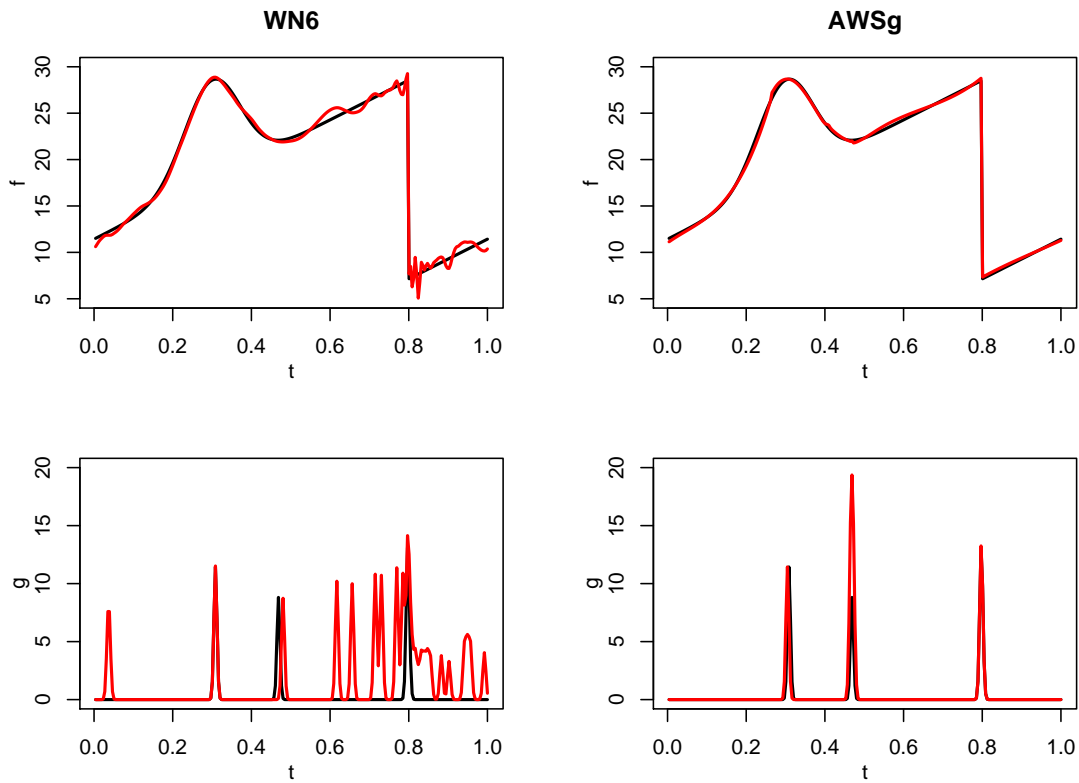


Abbildung 6.4: Blip-Funktion und Schätzungen durch WN6 und AWSg und jeweils zugehörige Funktionen g und \hat{g}

Um die Kurven g und \hat{g} zu vergleichen, dient die L_1 -Norm als Abstandsmaß. Ausgewertet werden die Funktionen jeweils an den Designpunkten $t_i, i = 1, \dots, n$. Damit ist der *Softloss* wie folgt definiert.

Definition 6.2 (Softloss)

Der *Softloss* ist definiert als

$$\text{Softloss} = \frac{1}{n} \sum_{i=1}^n |g(t_i) - \hat{g}(t_i)|.$$

	WN6	AWSg
Step (Abbildung 6.3)	2,225	0,362
Blip (Abbildung 6.4)	1,362	0,262

Tabelle 6.1: Beispielergebnisse des Softloss

Bemerkung 6.3 (Eigenschaften des Softloss)

- *Der Softloss verwendet die Extremstellen und -werte der wahren bzw. geschätzten Funktion simultan.*
- *Werden die Extremstellen und -werte richtig geschätzt, so gleichen sich g und \hat{g} und der Softloss ist 0.*
- *Der Verlust ist abhängig von dem bei der Faltung verwendeten σ . Da aber für g und \hat{g} dasselbe σ verwendet wird, ist der Einfluss gering.*
- *Überflüssige Extremwerte der Schätzfunktion zeigen sich auch in \hat{g} . Daher gehen sie stark in das Ergebnis des Softloss ein.*
- *Der Softloss ist als L_1 -Abstand nicht normiert.*
- *Ob der Verlust durch zu viele oder zu wenige Extremwerte zustande kommt, lässt sich am Ergebnis nicht erkennen.*

Die Ergebnisse des Softloss für die Situationen aus Beispiel 6.1 sind in Tabelle 6.1 eingetragen. Der Softloss gibt die aus den Abbildungen 6.3 und 6.4 gewonnenen Eindrücke im Ergebnis wieder. AWSg zeigt mit 0,362 und 0,262 jeweils bessere Ergebnisse als WN6 mit 2,225 und 1,362.

Der Softloss wird in Kapitel 7 zusammen mit den in früheren Kapiteln vorgestellten Verlustfunktionen angewendet und verglichen. Für die Berechnungen in der Simulationsstudie wird die Faltung von g_{Extr}^f und $g_{Extr}^{\hat{f}}$ jeweils mit einer Gaussfunktion mit Standardabweichung $\sigma = 1$ durchgeführt.

SIMULATIONSSTUDIE

In diesem Kapitel folgt die Anwendung des Extremstellenverlustes, des Abstands ℓ^Y , des Glattheitsverlustes $loss_g^Y$ und des Softloss in einer Simulationsstudie sowie eine Gegenüberstellung mit anderen Gütemaßen. Die Studie soll beurteilen, ob der Extremstellenverlust in der Lage ist, die Güte von Regressionsmethoden bezogen auf die Extremstellen zu bewerten. Genauso soll untersucht werden, wie sich der Softloss als Gütekriterium in einer Studie verhält. Das Studiendesign basiert auf dem nichtparametrischen Regressionsmodell (2.1) (vgl. Kapitel 2.1)

$$y(t_i) = f(t_i) + \varepsilon(t_i), t_i = \frac{i}{n}, i = 1, \dots, n.$$

Daten werden erzeugt, indem zunächst die 10 verschiedenen Testbed-Funktionen an $n = 256$ und $n = 1024$ Designpunkten bestimmt werden. In 1000 Simulationsläufen wird jeweils standardnormalverteiltes Rauschen $\varepsilon \sim N(0, 1)$ addiert. Auf die so simulierten Datensätze werden die 10 Regressionsmethoden aus Kapitel 2.2 angewendet. Von diesen Methoden liefern die Hälfte stückweise konstante Schätzungen, namentlich WH, UH, AWSk, TS und POT.

7.1 Andere Gütemaße

In Kapitel 2.3 wurden Gütemaße vorgestellt, die häufig verwendet werden. Welche Gütemaße in einer Simulation für die nichtparametrische Regression verwendet werden, sollte davon abhängen, welches Ziel die Regression verfolgt. Wenn das Ziel einer Regression zum Beispiel darin liegt, gute Vorhersagen zu machen, ist es im Allgemeinen

nicht so wichtig, ob die Schätzfunktion um die wahre Funktion *schwankt* und damit überflüssige Extremstellen liefert, wenn sie nur insgesamt nahe genug an der wahren Funktion liegt.

L_p-Normen

Die L_p -Normen basieren allein auf dem vertikalen Abstand von wahrer und geschätzter Funktion (vgl. Marron und Tsybakov, 1995). Kleine und damit gute Ergebnisse treten auf, wenn die geschätzte Funktion nahe an der wahren liegt. Ob sie dabei Artefakte in Form von überflüssigen Extremstellen liefert ist für das Ergebnis nicht relevant, wie bereits in Abbildung 3.2 in Kapitel 3 gezeigt wird. Die L_2 -Norm hat Optimalitätseigenschaften, wenn die Regression für Vorhersagen genutzt werden soll. Bezüglich des Erkennens von Extremwerten scheint sie hingegen weniger geeignet.

Die L_p -Normen sind nicht normiert, der durch sie gemessene Abstand zwischen Signal- und Schätzfunktion kann theoretisch beliebig groß werden. Ein Vergleich verschiedener Ergebnisse insbesondere bei verschiedenen Testbed-Funktionen wird dadurch erschwert.

Hausdorff-Abstand

Der Hausdorff-Abstand zwischen der Menge der wahren und der Menge der geschätzten Extremstellen kann als Gütekriterium eingesetzt werden. Damit hat der Hausdorff-Abstand als Grundlage dieselben Werte wie der ESV. Bei Betrachtung der Definition des Hausdorff-Abstands (vgl. Kapitel 2.3) wird deutlich, dass in diesen nur der Abstand zwischen zwei einzelnen Elementen dieser Mengen eingeht. Daher hat die Anzahl der Elemente in den zugrundeliegenden Mengen auf den Hausdorff-Abstand keinen Einfluss. Es macht also keinen Unterschied, ob beispielsweise nur an einer Stelle ein Extremwert zu viel geschätzt wird, oder an vielen Stellen, da immer nur der größte Abstand zählt. Die Werte des Hausdorff-Abstands sind nicht normiert.

Modalwerte

Um die Güte der Schätzmethoden bezüglich der Extremwerte zu messen, wird in Kovac (2007) lediglich die Anzahl der geschätzten Modalwerte, d.h. der Maxima, herangezogen. Der Median der Anzahl der Modalwerte in der Schätzfunktion über alle Simulationsläufe wird als Vergleichskriterium genutzt. Das alleinige Betrachten dieser Größe ist nicht aussagekräftig, sie muss mit der auch angegebenen wahren Anzahl Modalwerte verglichen werden. Es handelt sich bei diesem Wert daher nicht um ein Gütekriterium. In die mediane Anzahl der Modalwerte geht weder mit ein, ob diese Maxima an der richtigen Stelle geschätzt werden, noch ob sie die richtige Höhe haben und dieser Median ist nicht normiert.

Aus dem Ergebnis ist zwar ersichtlich, wenn zu viele Maxima geschätzt werden, jedoch nicht, ob *die richtigen* dabei sind, d.h. ob die Stellen identifiziert werden. Erkennbar ist auch, wenn zu wenige geschätzt werden, denn dann ist die Anzahl der geschätzten Modalwerte kleiner als die wahre. Wenn die Anzahl der geschätzten Modalwerte gleich der wahren ist, kann kaum eine Aussage getroffen werden, da nicht erfasst wird, wo die geschätzten Modalwerte liegen. Die Untersuchungen des ESV in Kapitel 4 haben gezeigt, dass durchaus nicht alle Regressionsmethoden die Extremwerte an der richtigen Stelle schätzen, und dass es damit sinnvoll ist, die Richtigkeit der Stellen mit einzubeziehen.

Da nur Maxima betrachtet werden, gehen in diesen Wert weniger Informationen ein als beispielsweise in den ESV.

Fryzlewiz (2007) gibt die Summe der Simulationsläufe an, in denen sowohl die richtige Anzahl Modalwerte, als auch einer mehr oder einer weniger gefunden wurde. Diese müssen miteinander und mit der der Simulationsläufe verglichen werden, und erschweren damit ein einfaches Erfassen des Ergebnisses einer Studie. Genauso wie der Median der Anzahl der Modalwerte sind auch die in Fryzlewiz (2007) gegebenen Anzahlen nicht normiert und beziehen nicht mit ein, ob die Modalwerte an der richtigen Stelle liegen. Zudem ist nicht zu erkennen, ob zu viele oder zu wenige Maximalstellen geschätzt werden, denn über die Stichproben, die nicht die richtige Anzahl Modalwerte, bzw. einen mehr oder einen weniger liefern, werden keine Angaben gemacht. In diese Betrachtung gehen ebenfalls keine Informationen zu den Minima ein.

7.2 Ergebnisse der Simulationsstudie

In der folgenden Simulationsstudie werden jeweils die L_1 - und die L_2 -Norm, der Hausdorff-Abstand, der Extremstellenverlust, ℓ^Y , $loss_g^Y$ und der Softloss berechnet und die Ergebnisse, soweit dies möglich ist, verglichen.

Um später die Güte bezüglich der Extremwerte zu beurteilen, werden zunächst wichtige die Extremwerte betreffende Eigenschaften der verwendeten Regressionsmethoden vorgestellt, die durch einen Verlust wiedergegeben werden sollen:

- Die Kernschätzer-Methode LPI versucht, den L_2 -Abstand zu minimieren und sollte damit gut geeignet sein, um Vorhersagen zu machen, d.h. sie sollte nicht stark von der Testbed-Funktion abweichen. Weniger wichtig ist dabei, ob die Schätzung um sie *oszilliert*, d.h. mal über und mal unter der Testbed-Funktion liegt und dabei überflüssige Extremwerte zeigt. Es ist zu erwarten, dass LPI zu viele Extremwerte liefert.
- Wenn die Hard Thresholding-Schranke bei Wavelet-Methoden zu niedrig ist, werden zu viele Koeffizienten zur Schätzung der Signalfunktion verwendet (vgl. Kapitel 2.2). Dies äußert sich in einzelnen überflüssigen Extremstellen, die sehr *spitz* und vergleichsweise hoch sind. Die Methoden, die Wavelets verwenden, neigen dazu, solche Artefakte zu liefern. Speziell die Methode MDL benutzt häufig zu viele der Waveletkoeffizienten zur Rekonstruktion der Funktion (vgl. Kapitel 2.2 und Davies, Gather, Weinert 2008). Es ist bekannt, dass die MDL-Methode bei einer Konstanten, wenn die Daten also aus weißem Rauschen bestehen, etwa 66% der Koeffizienten verwendet und damit extrem viele überflüssige Extremwerte liefert.
- Die Methode TS schätzt häufig die richtige Anzahl, tendenziell aber eher zu wenige als zu viele Extremwerte (vgl. Davies, Gather, Weinert 2008). Theoretisch sollten TS und STS dieselben Extremwerte schätzen. Aufgrund von Unterschieden in den beiden Implementierungen der Methoden im R-Package `ftnonpar` ist dies nicht der Fall (vgl. auch Kapitel 2.2). Die Implementierung der STS-Methode kann für einige der Stichproben der konstanten Funktion und der Sine-Funktion keine Ergebnisse liefern.
- POT ist als Methode zum Erkennen von Sprungstellen entwickelt worden (Wink-

ler, Liebscher 2002). Daher ist zu erwarten, dass diese Methode besonders bei stückweise konstanten Testbed-Funktionen gut abschneidet.

- Um eine Funktion glatt zu schätzen, wird häufig in Kauf genommen, dass die Schätzung Artefakte in Form von überflüssigen Extremwerten zeigt. Methoden, die Funktionen stückweise konstant schätzen, sollten daher seltener überflüssige Extremstellen liefern als glatte Schätzungen. Zudem sollte eine Methode wie STS, die das gleiche Monotonieverhalten hat wie die stückweise konstante Schätzung der TS-Methode, dahingehend bessere Ergebnisse liefern, als andere glatte Schätzungen.

Die Ergebnisse der Simulationsstudie werden wie folgt demonstriert. Für jede Funktion wird eine Stichprobengröße ausgewählt, und es werden zunächst beispielhaft sechs Schätzungen für einen Datensatz gezeigt und die zugehörigen Ergebnisse der Gütekriterien gegeben. Jeweils grau sind die Datenpunkte, schwarz die wahre Funktion und rot die Schätzfunktionen gezeichnet. Unter den Graphen sind die folgenden zugehörigen Verluste vermerkt: der L_2 -Abstand, der Hausdorff-Abstand (HD), der ESV, der Abstand ℓ^Y und der Softloss (SoLo).

Des Weiteren erfolgen verschiedenartige Zusammenfassungen der Simulationsergebnisse. Dazu werden Boxplots von allen Simulationsergebnissen für jeweils drei der sechs Kriterien geliefert. Eine Tabelle zeigt die Mittelwerte der Ergebnisse über 1000 Simulationen aller Kriterien.

Die Zeile *Tendenz* gibt an, ob eine Methode häufig einen positiven oder negativen ESV hat. Bestimmt wurde dafür zunächst das Vorzeichen bei jeder der 1000 Stichproben und dann die Summe darüber. War das Ergebnis kleiner als -100 , so erscheint in der Tabelle ein $(-)$, war es größer als 100 , erscheint ein $(+)$. (\circ) steht für die Ergebnisse zwischen -100 und 100 und gibt daher einerseits an, wenn sich positive und negative Vorzeichen aufheben, andererseits, wenn die Methode häufig einen ESV-Wert von 0 liefert.

Bei Betrachtung der mittleren Ergebnisse von $loss_g^Y$ ist zu beachten, dass nichtparametrische Regressionsmethoden im Allgemeinen das weiße Rauschen *aus den Daten glätten*. Dadurch kommt es häufiger vor, dass eine Methode eine Testbed-Funktion überglättet, als dass sie nicht stark genug glättet. Dennoch zeigt ein großer Mittelwert

von $loss_g^Y$ an, dass eine Methode zu stark glättet, da ein solcher Wert zustande kommt, wenn die meisten Maxima zu klein und Minima zu groß geschätzt werden.

Um ein leichteres Erfassen der Mittelwerte der Ergebnisse zu ermöglichen, wird jeweils zusätzlich eine Tabelle mit den Rängen über die Mittelwerte geliefert. Darin zeigt der kleinste Rang (1) die beste Methode an, der größte (10) die schlechteste. Die Ränge werden über mehr Nachkommastellen gebildet als in den Mittelwerttabellen angegeben, so dass zwei Methoden, die in der Mittelwerttabelle den selben Wert haben, verschiedene Ränge aufweisen können. Für den Abstand ℓ^Y und den Glattheitsverlust $loss_g^Y$ erübrigt sich die Bildung von Rängen, weil es nicht sinnvoll ist, jeweils alle Methoden mit diesem Verlust zu vergleichen, denn es gehen nicht alle Extremwerte ein. Des Weiteren unterbleibt in einigen Situationen das Aufzeigen von Rängen, wenn die Mittelwerte nur sehr wenig voneinander abweichen.

Diese Darstellung soll eine möglichst vollständige und gleichzeitig übersichtliche Wiedergabe der Ergebnisse der 1000 Simulationen für alle Testbed-Funktionen, Regressionsmethoden und Gütemaße liefern. Die Ergebnisse werden für jeweils eine Stichprobengröße dargestellt. In den meisten Fällen lassen sich die gezeigten Ergebnisse auf die jeweils andere Stichprobengröße übertragen, insbesondere da hier nicht Methoden sondern Gütemaße beurteilt werden. Liefert die andere Stichprobengröße abweichende Ergebnisse, so wird dies angegeben.

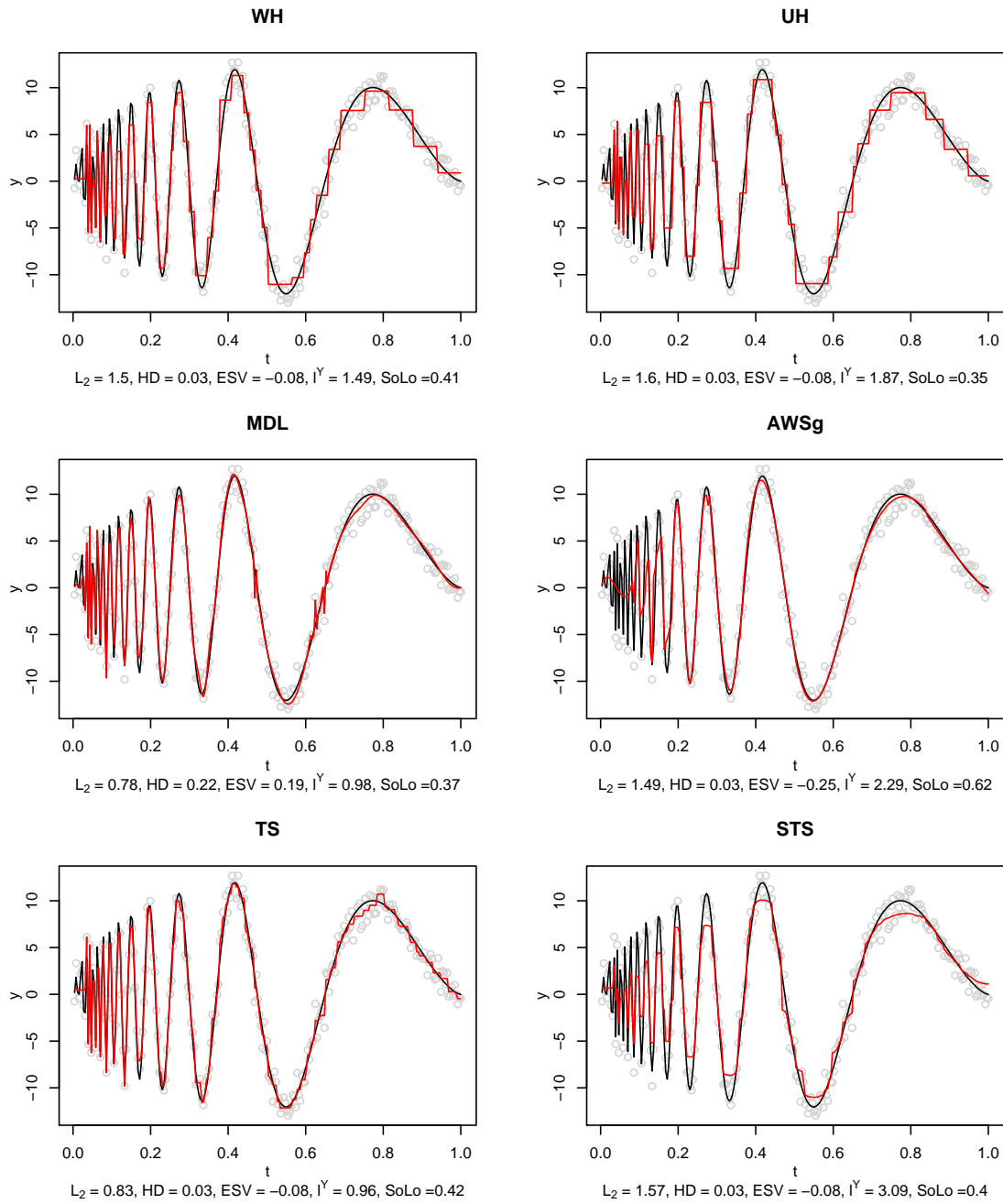
Die hier durchgeführten Berechnungen, speziell die Bestimmung der Extremstellen, verwenden eine Genauigkeit von fünf Nachkommastellen. Die Rechengenauigkeit kann je nach Ziel einer Regression angepasst werden und vor allem dafür genutzt werden, Einstellungen für die Genauigkeit, mit der ein Extremwert als solcher gilt, vorzunehmen.

Doppler-Funktion

- Abbildung 7.1 zeigt sechs Schätzungen der Doppler-Funktion. Die Schätzmethoden WH, UH, TS und STS haben alle einen ESV-Wert von -0,08. Bei diesen Methoden ist sichtbar, dass nur am Anfang des Intervalls $[0; 1]$ die wahren Extremstellen nicht geschätzt werden. Die Methode STS liefert auf den ersten Blick

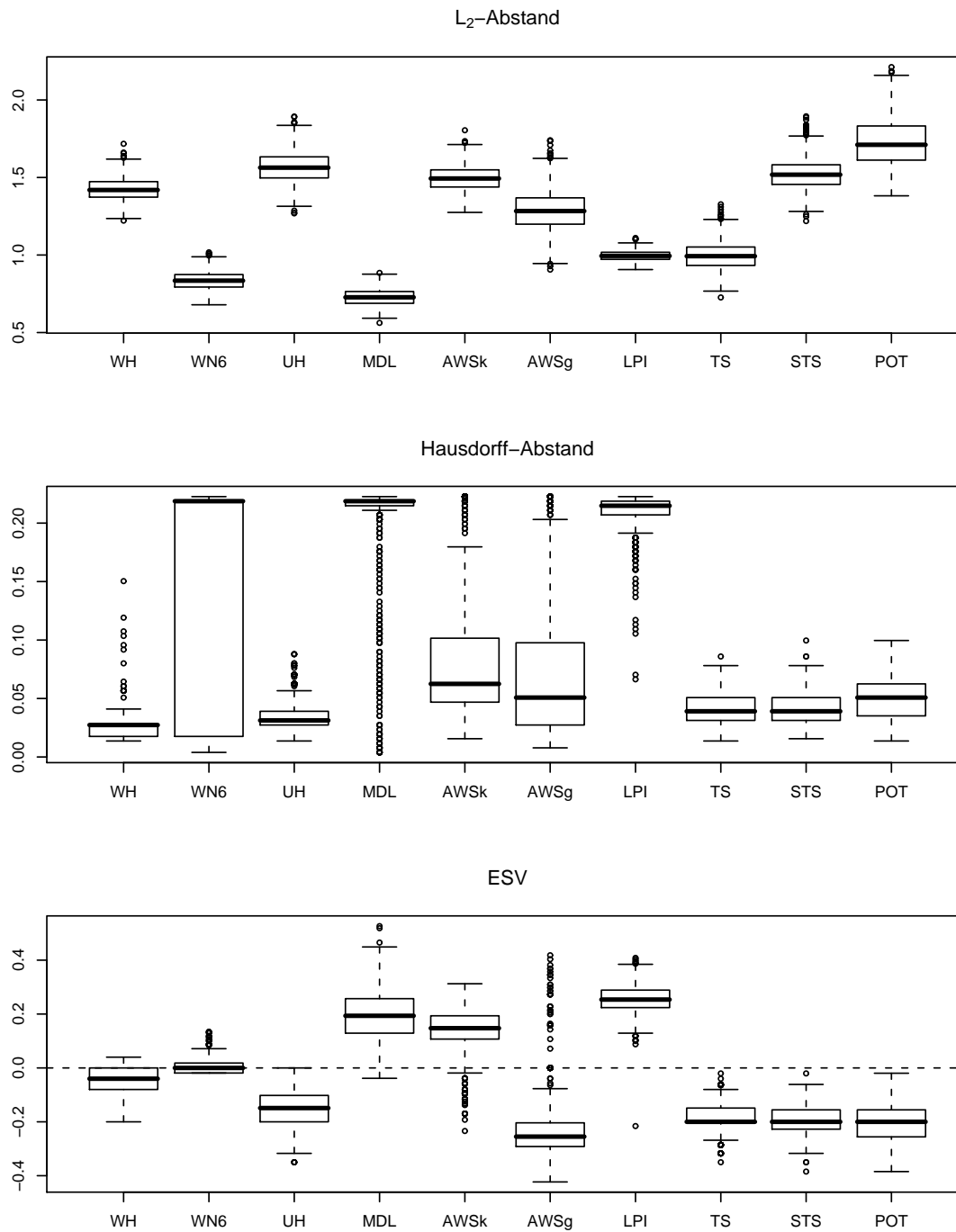
weniger geschätzte Extremstellen. Es zeigt sich jedoch, dass diese ein geringeres Ausmaß haben als die der anderen Methoden. Dies machen die Werte von ℓ^Y deutlich, mit 3,09 ist ℓ^Y bei STS erheblich größer als bei WH (1,49), UH (1,87) und TS (0,96). Der ℓ^Y -Wert gibt im Vergleich die Entfernung wahrer und geschätzter Extremwerte gut wieder: die Schätzung von TS liegt näher an der wahren Funktion als die der anderen Methoden.

- Bei Betrachtung der Schätzung durch die MDL-Methode fällt auf, dass zu viele Extremstellen vorhanden sind. Der ESV-Wert von 0,19 gibt diese Beobachtung wieder. Der ℓ^Y -Wert von 0,98 zeigt, dass die Extremwerte der identifizierten Extremstellen verglichen mit den anderen Methoden nah an den wahren liegen.
- Die AWSg-Methode schätzt am Anfang des Intervalls $[0; 1]$ offensichtlich zu wenige Extremstellen. Der ESV-Wert von -0,25 beschreibt diese Beobachtung. Der verglichen mit den anderen Ergebnissen relativ große ℓ^Y -Wert spiegelt wieder, dass die Extremwerte der identifizierten Stellen im Intervall $[0,1; 0,3]$ zu stark geglättet werden.
- Die Ergebnisse des L_2 -Abstands liefern einen anderen Eindruck der Güte der betrachteten Methoden: den kleinsten L_2 -Wert liefert MDL, gefolgt von TS. Die anderen Methoden haben ähnliche L_2 -Werte von 1,49 bis 1,6.
- Der Hausdorff-Abstand ist für fünf der Methoden mit dem Wert 0,03 gleich. Einzig MDL hat einen Hausdorff-Abstand von 0,22.
- Der Softloss liefert einen anderen Eindruck als die übrigen Verluste. Nur das Softloss-Ergebnis für die AWSg-Methode unterscheidet sich wesentlich von dem der anderen Methoden.
- Bei Betrachtung von Boxplots des L_2 -Abstands in Abbildung 7.2 wird deutlich, dass die Ergebnisse für jede Methode wenig variieren, da die Boxplots kompakt sind. Gemessen mit dieser Norm schneiden die Methoden, die stückweise konstante Schätzungen liefern, weitgehend schlecht ab. Unter den anderen Methoden liefern WN6 und MDL, also Methoden die glatte Wavelets verwenden, die besten Ergebnisse.

Abbildung 7.1: Beispiele für Schätzungen der Doppler-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	1,1	0,56	1,2	0,51	1,16	0,69	0,61	0,68	1,09	1,34
L_2 -Abstand	1,43	0,83	1,57	0,73	1,5	1,29	1	1	1,52	1,73
Hausdorff-Abstand	0,02	0,16	0,04	0,2	0,08	0,07	0,21	0,04	0,04	0,05
ESV (Betrag)	0,04	0,02	0,15	0,19	0,15	0,25	0,26	0,18	0,2	0,21
Tendenz	(-)	(+)	(-)	(+)	(+)	(-)	(+)	(-)	(-)	(-)
l^Y	1,23	0,95	1,77	0,79	1,58	1,61	1,81	0,87	2,83	1,85
$loss_g^Y$	0,73	0,54	0,88	0,48	0,87	0,73	0,85	0,52	1,00	0,90
Softloss	0,38	0,29	0,34	0,52	0,68	0,49	0,79	0,42	0,42	0,37

Tabelle 7.1: Mittelwert der Ergebnisse für die Doppler-Funktion, $n = 256$

Abbildung 7.2: Ergebnisse für die Doppler-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	7	2	9	1	8	5	3	4	6	10
L_2 -Abstand	6	2	9	1	7	5	3	4	8	10
Hausdorff-Abstand	1	8	2	9	7	6	10	3	4	5
ESV (Betrag)	2	1	3	6	4	9	10	5	7	8
Softloss	4	1	2	8	9	7	10	5	6	3

Tabelle 7.2: Ränge der mittleren Ergebnisse für die Doppler-Funktion, $n = 256$

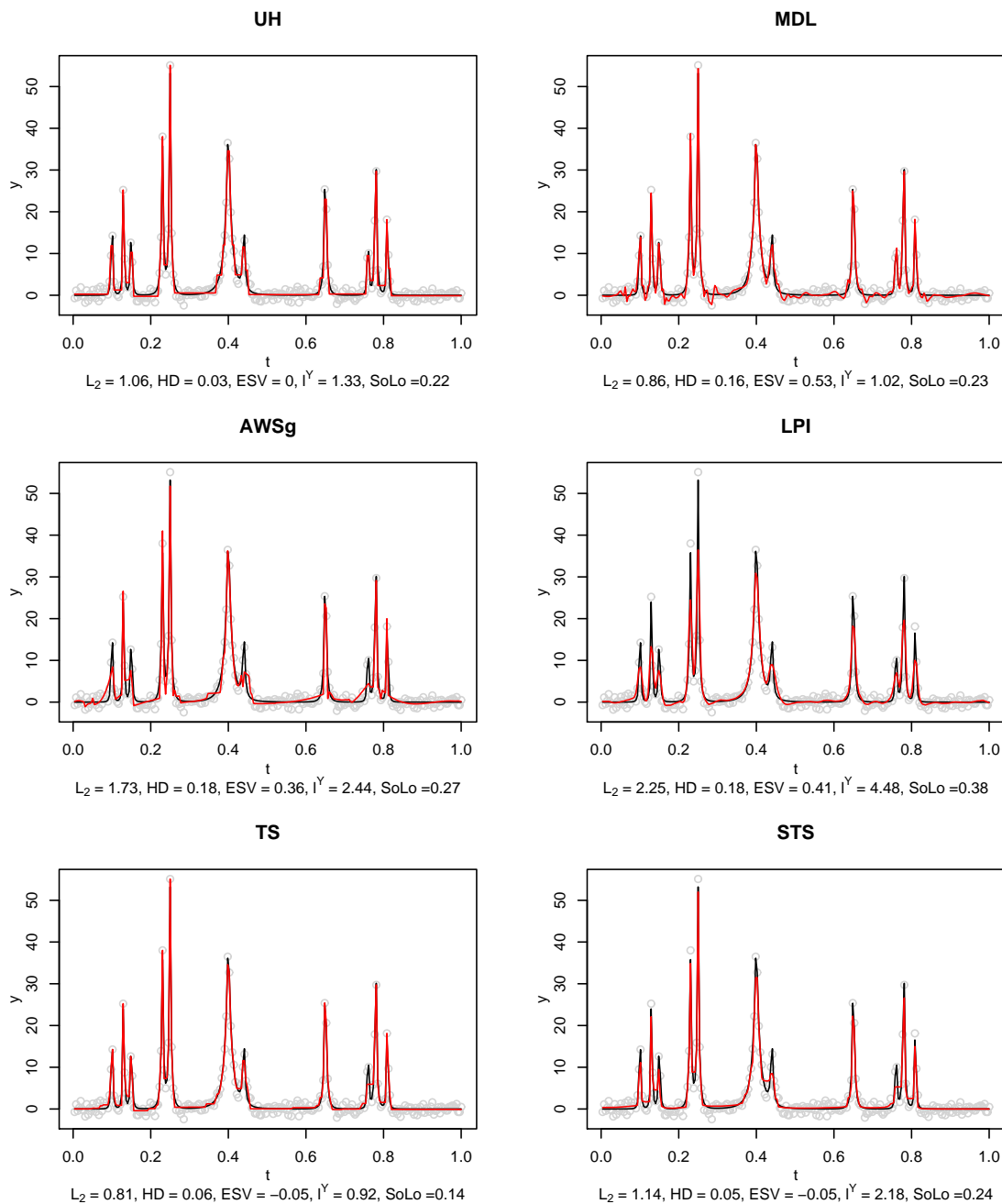
- Der Hausdorff-Abstand der wahren und der geschätzten Extremstellen führt bei den Methoden WN6, AWSk und AWSg zu nicht so konzentrierten Boxplots, das heißt, dass dieser Abstand für verschiedene Stichproben sehr unterschiedliche Ergebnisse liefern kann. Auffällig ist zudem, dass MDL und LPI zu viel schlechteren Ergebnissen führen als die anderen Methoden.
- Die Ergebnisse des ESV veranschaulichen, dass einige Methoden dazu neigen, eher zu wenige und andere eher zu viele Extremstellen der Doppler-Funktion zu schätzen. So zeigen MDL, AWSk und LPI eher zu viele Extremstellen, WH, UH, AWSg, TS, STS und POT eher zu wenige. Insgesamt liefert hier WN6 die besten Ergebnisse. Der ESV beurteilt die Methoden anders als der L_2 - oder Hausdorff-Abstand. Zwar ist WN6 für L_2 -Abstand und ESV ähnlich gut, weist aber ein verhältnismäßig schlechtes Ergebnis für den Hausdorff-Abstand auf. MDL schneidet beim L_2 -Abstand gut, aber beim Hausdorff-Abstand und ESV schlecht ab.
- Tabelle 7.1 zeigt die Mittelwerte der Ergebnisse. Deutlich wird, dass die Methoden im Mittel die Extremwerte der Doppler-Funktion gut erkennen, da die mittleren ESV-Werte nicht größer als 0,26 sind. Bestätigt wird dies durch die Boxplots in Abbildung 7.2, die belegen, dass die ESV-Werte betragsmäßig selten größer als 0,4 sind. Anhand der Tendenz bezüglich des Vorzeichens von ESV lässt sich wie schon im Boxplot erkennen, dass einige Methoden eher zu wenige Extremstellen der Doppler-Funktion schätzen. Zusammen mit der Betrachtung von Beispielstichproben wie in Abbildung 7.1 zeigt das, dass diese Methoden die Doppler-Funktion im Bereich $[0; 0,2]$, in dem sie stark oszilliert, nicht richtig schätzen und im weiteren Verlauf der Funktion keine überflüssigen Extremwerte produzieren. Auch Methoden mit positivem ESV erkennen im Anfangsbereich häufig nicht alle Extremstellen, schätzen jedoch im weiteren Verlauf überflüssige Extremstellen.
- Der ℓ^Y -Wert sollte nur zusammen mit dem ESV betrachtet werden. Beispielsweise liefert MDL beim ℓ^Y -Wert gute, d.h. kleine Werte, hat aber einen verhältnismäßig großen ESV. Für WN6 ist erkennbar, dass die Methode nicht nur die Extremstellen der Doppler-Funktion gut erkennt, sondern ebenso in der Lage ist, die richtigen Höhen zu schätzen. STS und POT sind relativ gut im Erkennen von Extremstellen, aber nicht von der zugehörigen Höhe. LPI schneidet in beidem

nicht gut ab: es ergeben sich sowohl recht große Verlustwerte für den ESV als auch große Werte beim ℓ^Y -Wert.

- Anhand des Mittelwerts von 1,00 beim Glattheitsverlust von STS ist erkennbar, dass diese Methode die Extremwerte immer zu stark glättet. Dieses Ergebnis deckt sich mit der Beobachtung aus Abbildung 7.1, da im Graphen offensichtlich ist, dass STS zu stark glättet. Auch die anderen Methoden weisen einen verhältnismäßig großen $loss_g^Y$ -Wert auf. Die Extremwerte der Doppler-Funktion werden zumeist überglättet.
- Bei Betrachtung der Ränge der Mittelwerte in Tabelle 7.2 wird auf den ersten Blick deutlich, dass die verschiedenen Verluste verschiedene Methoden als gut bewerten. Besonders augenfällig ist dies bei der MDL-Methode, die beim L_2 -Abstand am besten, bei Hausdorff-Abstand, ESV und Softloss viel schlechter abschneidet. Für $n = 1024$ hat sie sogar für den L_1 - und L_2 -Abstand den Rang 1 und für die anderen Kriterien den Rang 10.

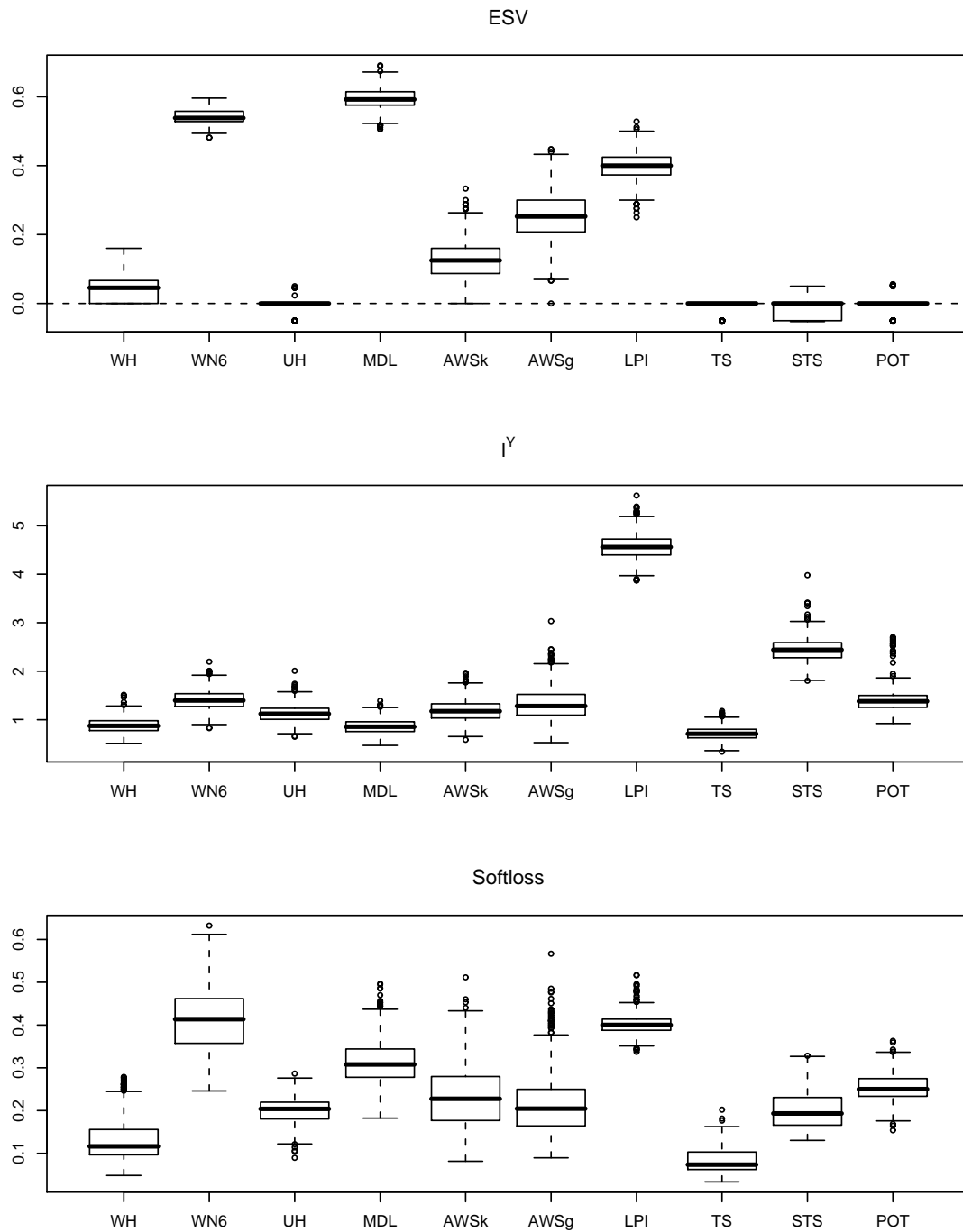
Bumps-Funktion

- Abbildung 7.3 zeigt, dass die UH-Methode die Extremstellen der Bumps-Funktion alle gut erkennt. Das gibt der ESV-Wert von 0 wieder. Von der TS und STS-Methode werden nur das Maximum und das Minimum kurz vor $t = 0,8$ nicht wahrgenommen und der ESV-Wert von $-0,05$ gibt das wieder. Gleichzeitig zeigt der ℓ^Y -Wert, dass die Methode TS die Höhe der Extremwerte am besten schätzt, STS liefert schlechtere Ergebnisse. Die anderen Methoden, MDL, AWSg und LPI schätzen überflüssige Extremwerte und haben einen größeren ESV-Wert. MDL schätzt dabei häufig die richtige Höhe der Extremwerte, was der ℓ^Y -Wert von 1,02 wiedergibt. LPI unterschätzt die Höhe der Maxima sichtbar. Dies spiegelt der ℓ^Y -Wert von 4,48 wider.
- Gemessen mit dem L_2 -Abstand liefern die Methoden MDL und TS die besten Ergebnisse. Auffällig ist, dass die MDL-Methode trotz überflüssiger Extremwerte einen vergleichsweise kleinen L_2 -Abstand hat.

Abbildung 7.3: Beispiele für Schätzungen der Bumps-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,61	0,88	0,59	0,66	0,69	0,83	1,04	0,43	0,7	0,75
L_2 -Abstand	0,98	1,23	0,98	0,88	1,15	1,41	2,26	0,66	1,17	1,27
Hausdorff-Abstand	0,05	0,17	0,03	0,17	0,07	0,14	0,17	0,05	0,03	0,03
ESV (Betrag)	0,04	0,54	0	0,59	0,12	0,25	0,4	0,01	0,01	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(-)	(-)	(o)
l^Y	0,88	1,41	1,13	0,86	1,19	1,33	4,57	0,72	2,44	1,39
$loss_g^Y$	0,57	0,46	0,81	0,39	0,73	0,61	0,82	0,52	0,99	0,86
Softloss	0,13	0,41	0,2	0,31	0,23	0,21	0,4	0,08	0,2	0,25

Tabelle 7.3: Mittelwert der Ergebnisse für die Bumps-Funktion, $n = 256$

Abbildung 7.4: Ergebnisse für die Bumps-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	3	9	2	4	5	8	10	1	6	7
L_2 -Abstand	3	7	4	2	5	9	10	1	6	8
Hausdorff-Abstand	4	10	3	9	6	7	8	5	1	2
ESV (Betrag)	5	9	1	10	6	7	8	3	4	2
Softloss	2	10	4	8	6	5	9	1	3	7

Tabelle 7.4: Ränge der mittleren Ergebnisse für die Bumps-Funktion, $n = 256$

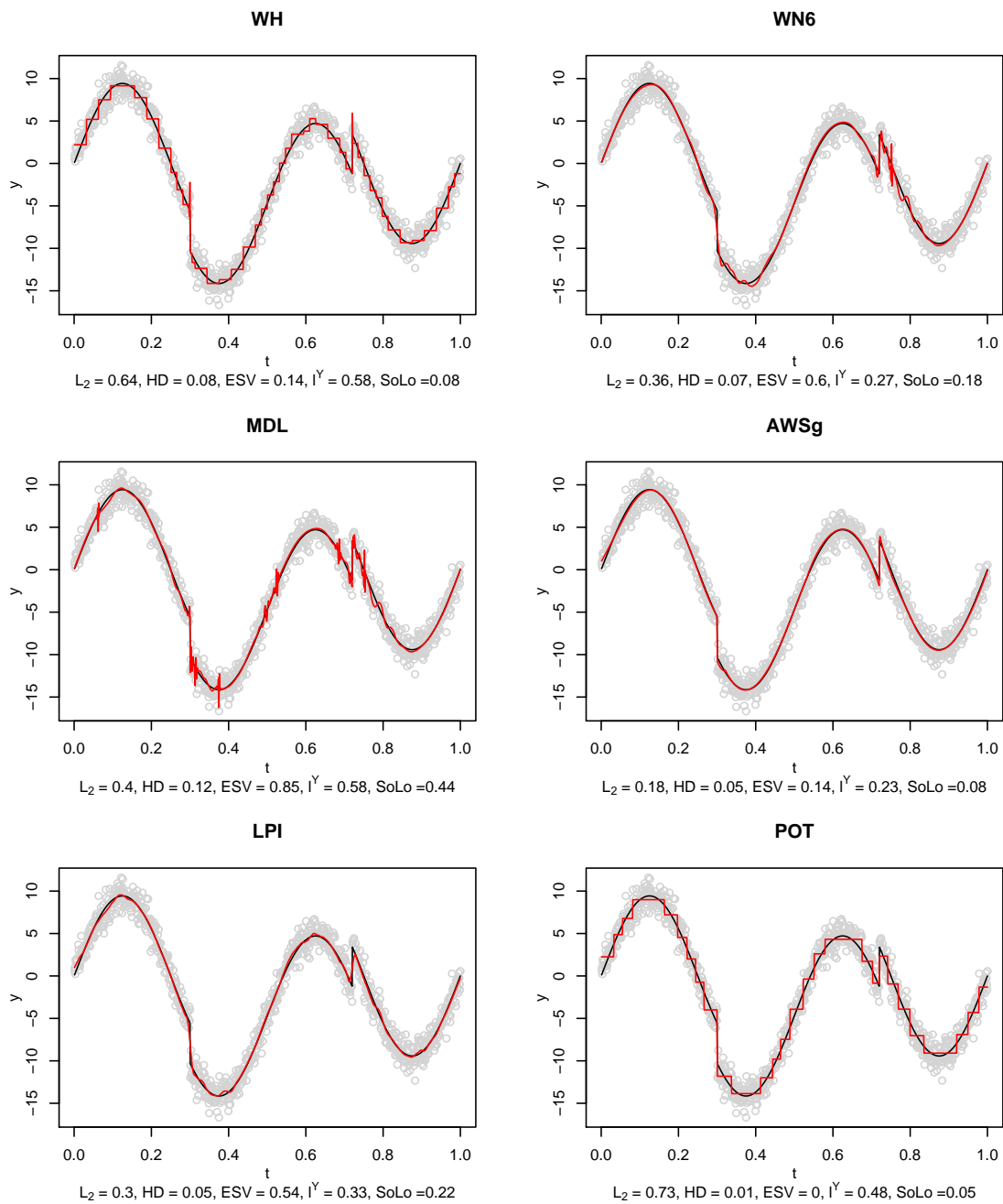
- Der Hausdorff-Abstand weist qualitativ ähnliche Ergebnisse wie der ESV auf.
- Der Softloss zeigt wenig Unterschiede zwischen den Methoden UH, MDL, AWSg und STS. Die Eindrücke aus den Abbildungen sind offensichtlich nicht verallgemeinerbar. TS hat mit 0,14 einen besseren Softloss als die anderen Methoden, LPI mit 0,38 einen schlechteren.
- Die Boxplots des ESV in Abbildung 7.4 verdeutlichen, dass die Methoden UH, TS und POT die Extremstellen der Bumps-Funktion nahezu immer richtig erkennen. WH und STS liefern bei 0 konzentrierte Boxplots, wobei WH eher zu viele und STS eher zu wenige Extremstellen schätzt. Die fünfte stückweise konstante Schätzung AWSk zeigt etwas größere ESV-Werte und zu viele Extremstellen. Offensichtlich sind die stückweise konstanten Schätzungen gut geeignet, die Extremstellen der Bumps-Funktion zu bestimmen. Die einzige der glatten Schätzungen mit ähnlich guten Ergebnissen wie die stückweise konstanten ist STS und diese Methode nutzt eine stückweise konstante Schätzung zur Bestimmung der Extremstellen. Von den übrigen Methoden schneidet AWSg am besten, LPI mittelmäßig gut und WN6 und MDL schlecht ab.
- Bei Betrachtung von ℓ^Y zeigt sich, dass die Methode TS auch die richtige Höhe der Extremwerte schätzt. Die anderen Methoden mit einem guten ESV weisen schlechtere Ergebnisse auf. Auffällig ist, dass LPI große Werte um 4,5 liefert. Abbildung 7.3 hat gezeigt, dass LPI die Kurve zu stark glättet. Das gute Abschneiden von STS beim ESV scheint auf Kosten des ℓ^Y -Werts zu gehen, was schon Abbildung 7.3 veranschaulicht.
- Die Boxplots des Softloss belegen sehr verschiedene Ergebnisse für die Methoden. Die niedrigen Softloss-Werte von TS und WH geben die Ergebnisse des ESV und ℓ^Y -Werts bei diesen Methoden wieder. Für die anderen Methoden fasst der Softloss zumindest tendenziell die Ergebnisse der anderen beiden Verluste zusammen.
- Die Mittelwerte des L_1 - und L_2 -Abstands in Tabelle 7.3 nennen mit 0,66 und 0,88 überraschend kleine Werte für die MDL-Methode. Das lässt sich damit erklären, dass die Schätzungen durch MDL zwar stark um die wahre Funktion schwanken und damit viele überflüssige Extremwerte liefern, der absolute Abstand zwischen

Bumps-Funktion und MDL-Schätzung jedoch nicht so groß ist (vgl. auch Abbildung 7.3). Bei TS ist auffällig, dass die Tendenz des Vorzeichens von ESV negativ ist. Obwohl der Boxplot von TS bei 0 konzentriert ist, ergibt sich für den ESV häufiger als 100 mal ein negatives Vorzeichen. Die einzelnen Ergebnisse der 1000 Stichproben zeigen, dass TS für 758 Stichproben einen ESV von 0 liefert, für 238 einen Wert von -0,05 und für drei den Wert -0,0526. Die 1000 ESV-Werte der POT-Methode sind nur in 84 Fällen ungleich 0 und zeigen dann einige positive und viele negative Ergebnisse.

- In den meisten Fällen können die Methoden die Maxima der Bumps-Funktion nicht hoch genug bestimmen. Das lässt sich an den hohen Ergebnissen von $loss_g^Y$ erkennen.
- Tabelle 7.4 zeigt, dass L_1 -, L_2 - und Hausdorff-Abstand andere Ergebnisse liefern als der ESV und der Softloss. Die Ränge zwischen L_1 - und L_2 -Abstand unterscheiden sich untereinander nur wenig, aber stark vom ESV. Beispielsweise hat die POT-Methode für die Normen den Rang 7 und 8 und für den ESV den Rang 2. Die Ränge des Hausdorff-Abstands differieren nicht so stark von den ESV-Werten, jedoch ist beispielsweise STS die beste Methode beim Hausdorff-Abstand und nur die viertbeste beim ESV.

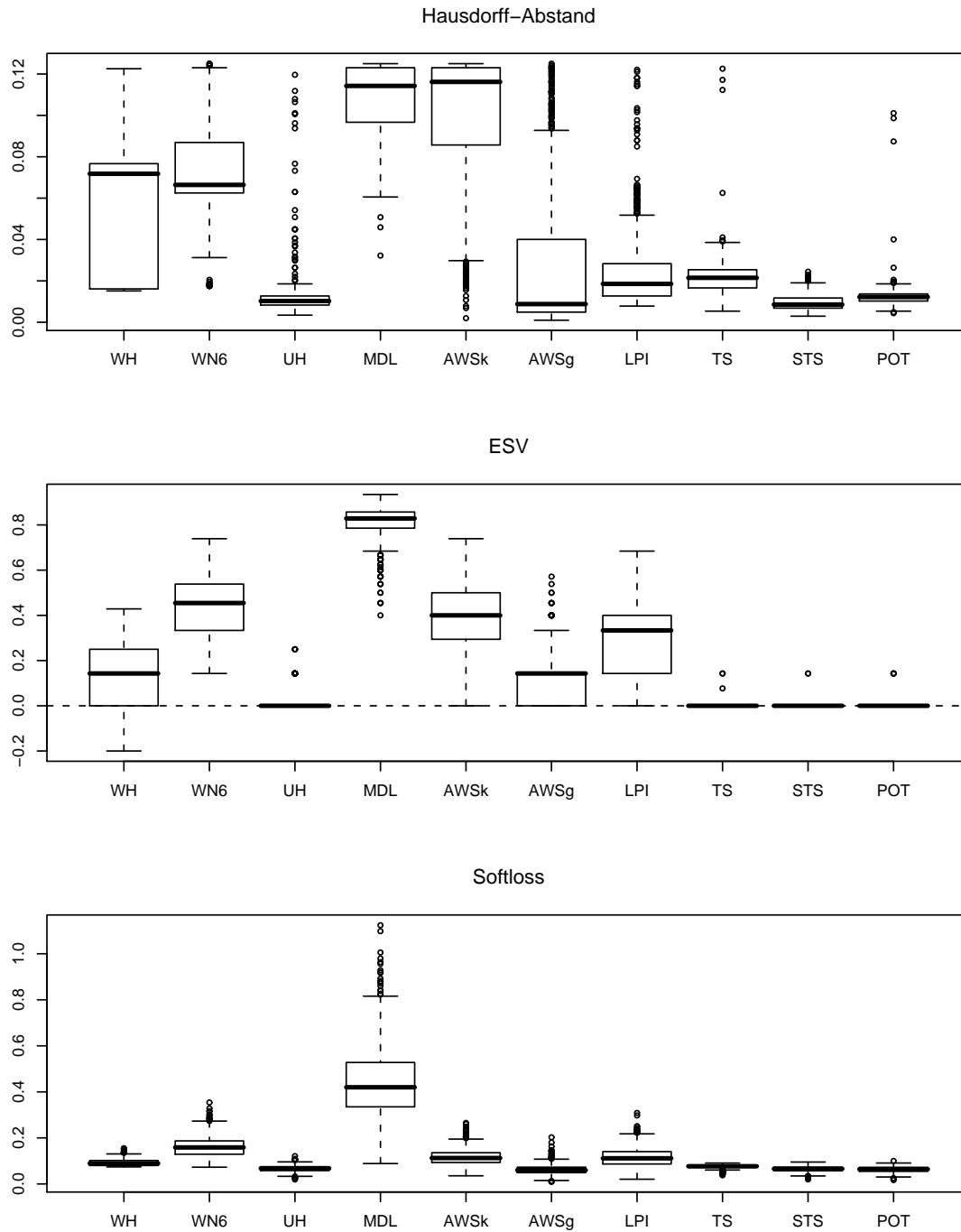
Heavisine-Funktion

- Abbildung 7.5 stellt Beispiele für Schätzungen der Heavisine-Funktion mit 1024 Designpunkten vor. Die WH-Methode zeigt wenige überflüssige Extremstellen und hat einen dazu passenden ESV von 0,14. Bei AWSg sind die überflüssigen Extremwerte nicht auf den ersten Blick erkennbar, bei genauerem Hinsehen lassen sie sich im Bereich zwischen $t = 0,3$ und $t = 0,4$ erkennen. Die überflüssigen Extremstellen, die durch WN6 und MDL produziert werden, sind hingegen offensichtlich. Die großen ESV-Werte von 0,6 bzw. 0,85 geben die Eindrücke aus den Abbildungen wieder. LPI zeigt ebenso überflüssige Extremstellen und hat einen ESV von 0,54. Die POT-Methode schätzt die Extremstellen der Heavisine-Funktion einwandfrei und hat ESV 0.

Abbildung 7.5: Beispiele für Schätzungen der Heavisine-Funktion, $n = 1024$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,47	0,26	0,48	0,25	0,42	0,13	0,19	0,28	0,34	0,58
L_2 -Abstand	0,61	0,39	0,6	0,4	0,53	0,19	0,31	0,36	0,45	0,71
Hausdorff-Abstand	0,05	0,07	0,01	0,11	0,1	0,03	0,02	0,02	0,01	0,01
ESV (Betrag)	0,12	0,45	0,01	0,81	0,39	0,12	0,3	0	0	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(o)	(o)	(o)
l^Y	0,53	0,5	0,45	0,39	0,4	0,19	0,46	0,4	0,72	0,59
$loss_g^Y$	0,82	0,56	0,91	0,47	0,87	0,50	0,68	0,41	0,99	0,95
Softloss	0,09	0,16	0,07	0,44	0,12	0,06	0,12	0,07	0,06	0,06

Tabelle 7.5: Mittelwert der Ergebnisse für die Heavisine-Funktion, $n = 1024$

Abbildung 7.6: Ergebnisse für die Heavisine-Funktion, $n = 1024$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	8	4	9	3	7	1	2	5	6	10
L_2 -Abstand	9	4	8	5	7	1	2	3	6	10
Hausdorff-Abstand	7	8	2	10	9	6	5	4	1	3
ESV (Betrag)	6	9	4	10	8	5	7	2	1	3
Softloss	6	9	4	10	8	1	7	5	3	2

Tabelle 7.6: Ränge der mittleren Ergebnisse für die Heavisine-Funktion, $n = 1024$

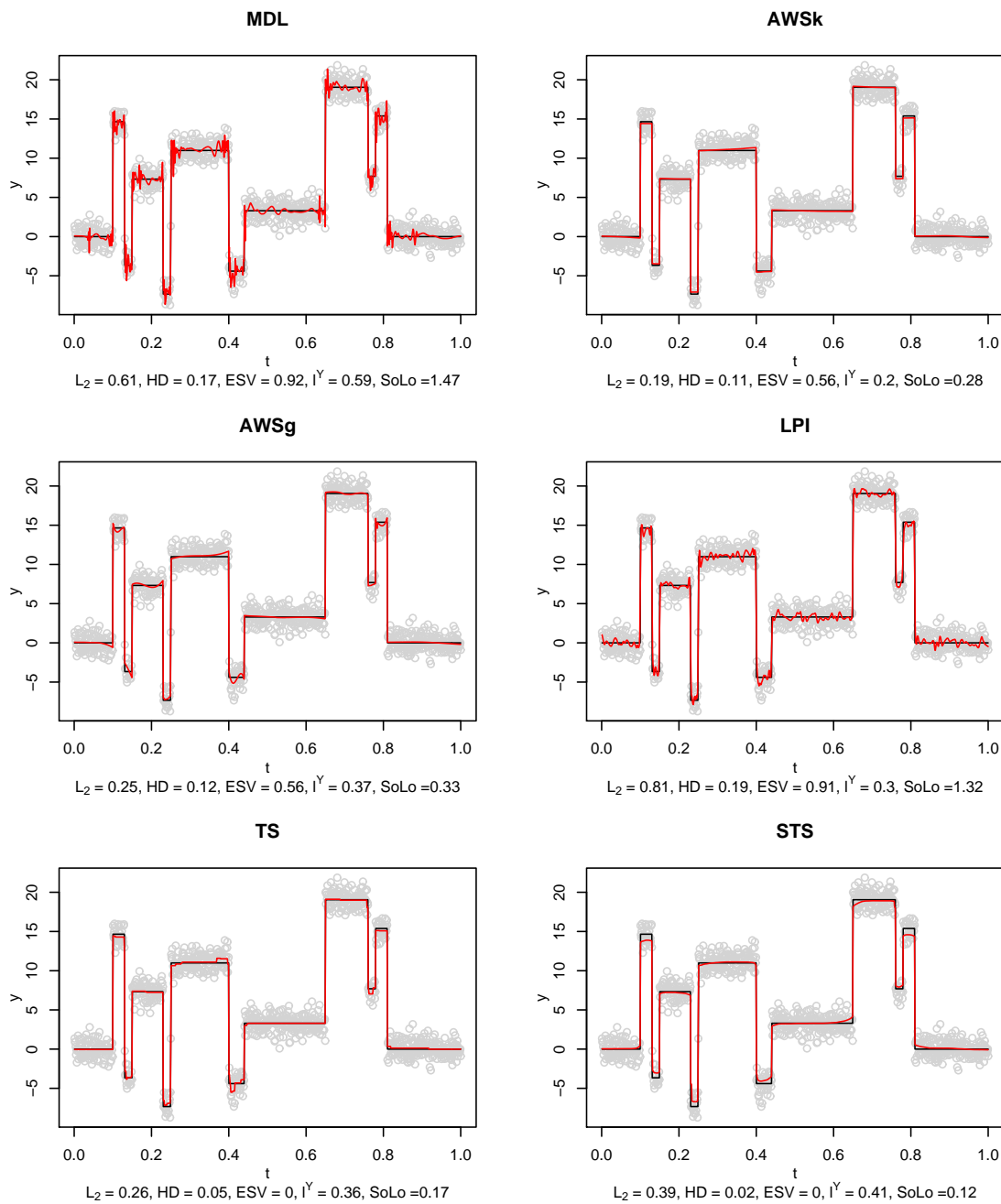
- Die Werte des Hausdorff-Abstands sind für alle Methoden relativ gering: sie variieren zwischen 0,01 bei POT und 0,12 bei MDL. Dieser Abstand erfasst die Güte im Erkennen von Extremstellen in diesem Fall nicht gut.
- ℓ^Y liefert große Werte bei WH und LPI. Für POT ist ℓ^Y mit 0,48 noch verhältnismäßig groß. WN6 und AWSg weisen kleinere ℓ^Y -Werte auf als LPI, was sich durch die Betrachtung der Abbildung nicht vermuten lässt.
- Die Softloss-Werte geben in diesem Fall die ESV-Werte tendenziell wieder.
- Verglichen mit dem L_2 -Abstand liefert AWSg das beste Ergebnis bei der Heavisine-Funktion, da der Abstand der Schätzung zur wahren Funktion insgesamt nicht so groß ist.
- Bei den Boxplots des Hausdorff-Abstands in Abbildung 7.6 fällt auf, dass die Ergebnisse insgesamt zwar mit Werten zwischen 0 und 0,12 keinen großen Abstand haben, die Variabilität über die verschiedenen Stichproben bei einigen Methoden aber recht groß ist. So sind die Boxplots von WH, WN6, MDL, AWSk und AWSg nicht konzentriert. UH, STS und POT zeigen niedrige Ergebnisse und konzentrierte Boxplots.
- UH, TS, STS und POT schätzen nahezu immer die Extremstellen richtig, was der ESV belegt. Die anderen Methoden schätzen eher zu viele Extremstellen. Am schlechtesten schneidet dabei MDL ab, gefolgt von AWSk, WN6, LPI.
- Der Softloss zeigt nur wenig Unterschiede zwischen den Methoden, nur MDL ist schlechter als die anderen. Nach diesem Verlust scheinen die Ergebnisse der übrigen Methoden alle gleich gut zu sein, anders als sich aus Abbildung 7.5 vermuten lässt.
- Tabelle 7.5 belegt, dass die Methoden STS und POT einen guten ESV-Durchschnitt haben, jedoch einen verhältnismäßig großen ℓ^Y -Wert. Durch den Softloss, der beide Richtungen messen sollte, wird das nicht wiedergegeben.
- Der Wert 0,41 von TS bei $loss_g^Y$ zeigt, dass TS in vielen Fällen die Extrema im Gegensatz zu anderen stückweise konstanten Schätzungen nicht zu stark glättet. Die anderen Methoden, die stückweise konstante Schätzungen liefern, WH, UH,

AWSk und POT haben $loss_g^Y$ -Werte über 0,8. Die STS-Methode weist wiederum einen sehr hohen Wert von 0,99 auf.

- Die durchschnittlichen Werte des L_1 -, L_2 - und Hausdorff-Abstands präsentieren kaum Unterschiede in der Beurteilung der Güte der Methoden. Das ist ein Hinweis darauf, dass die Heavisine-Funktion verhältnismäßig einfach zu schätzen ist.
- Die Bildung der Ränge in Tabelle 7.6 erfolgt daher zum Teil aufgrund von geringen Unterschieden. Deutlich wird hier wiederum, dass sich die Ergebnisse der beiden Normen untereinander kaum, von den anderen Kriterien aber stark unterscheiden.

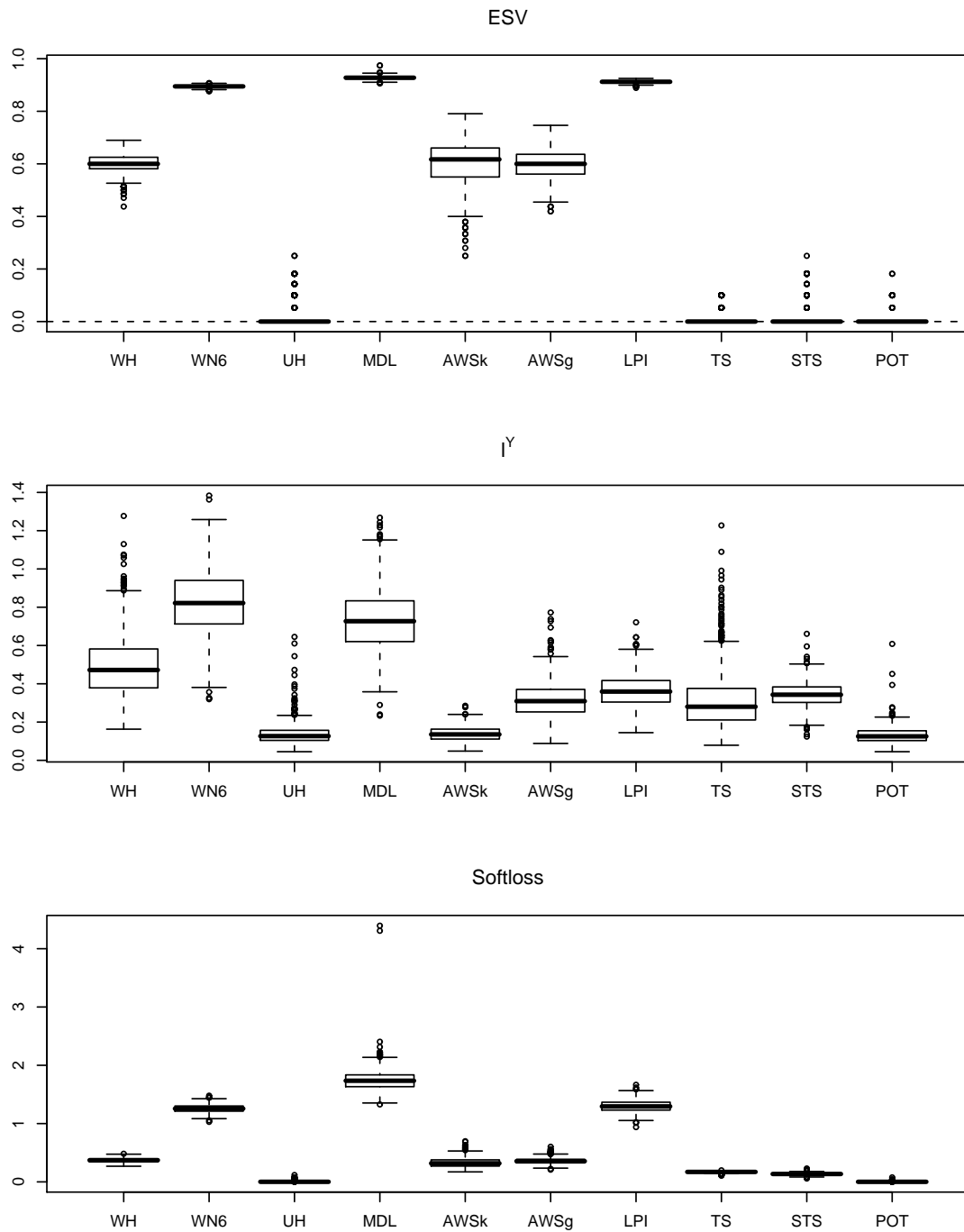
Blocks-Funktion

- In Abbildung 7.7 lässt sich erkennen, dass die Methoden MDL und LPI überflüssige Extremstellen erzeugen. Der ESV-Wert ist mit 0,92 und 0,91 entsprechend hoch. Der ESV-Wert 0,56 bei AWSk und AWSg deutet auf zu viele Extremstellen hin, die jedoch in den Graphen nicht sichtbar sind. Die Methoden TS und STS haben einen ESV von 0. Die Ergebnisse des Hausdorff-Abstands und des Softloss geben die Situation genauso wieder.
- Der Wert des L_2 -Abstands ist für die LPI-Methode mit 0,81 größer als für MDL mit 0,61. Die Betrachtung der Graphen lässt dies nicht erwarten.
- ℓ^Y macht deutlich, dass die Methode AWSk die Höhe der identifizierten Extremstellen am besten schätzt. MDL ist hier am schlechtesten. Die offensichtlichen Unterschiede von wahren und geschätzten Werten bei der STS-Methode werden durch den ℓ^Y -Wert nicht wiedergegeben.
- Die Boxplots in Abbildung 7.8 für den ESV sind für die Methoden UH, TS, STS und POT bei 0 konzentriert, diese Methoden schätzen die Extremstellen nahezu immer richtig. Die Methoden WN6, MDL und LPI zeigen häufig Artefakte wie in Abbildung 7.7 und liefern immer schlechte ESV-Werte.

Abbildung 7.7: Beispiele für Schätzungen der Blocks-Funktion, $n = 1024$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,19	0,54	0,08	0,46	0,1	0,17	0,4	0,15	0,24	0,08
L_2 -Abstand	0,35	0,78	0,12	0,65	0,19	0,26	0,8	0,24	0,4	0,11
Hausdorff-Abstand	0,11	0,2	0,01	0,19	0,16	0,16	0,2	0,05	0,04	0
ESV (Betrag)	0,6	0,89	0,01	0,93	0,6	0,6	0,91	0	0,02	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(o)	(+)	(o)
l^Y	0,49	0,82	0,14	0,73	0,14	0,32	0,36	0,31	0,34	0,13
$loss_g^Y$	0,15	0,10	0,49	0,12	0,43	0,15	0,23	0,23	0,86	0,49
Softloss	0,37	1,26	0	1,75	0,33	0,36	1,3	0,17	0,13	0

Tabelle 7.7: Mittelwert der Ergebnisse für die Blocks-Funktion, $n = 1024$

Abbildung 7.8: Ergebnisse für die Blocks-Funktion, $n = 1024$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	6	10	2	9	3	5	8	4	7	1
L_2 -Abstand	6	9	2	8	3	5	10	4	7	1
Hausdorff-Abstand	5	9	2	8	7	6	10	4	3	1
ESV (Betrag)	6	8	3	10	7	5	9	2	4	1
Softloss	7	8	2	10	5	6	9	4	3	1

Tabelle 7.8: Ränge der mittleren Ergebnisse für die Blocks-Funktion, $n = 1024$

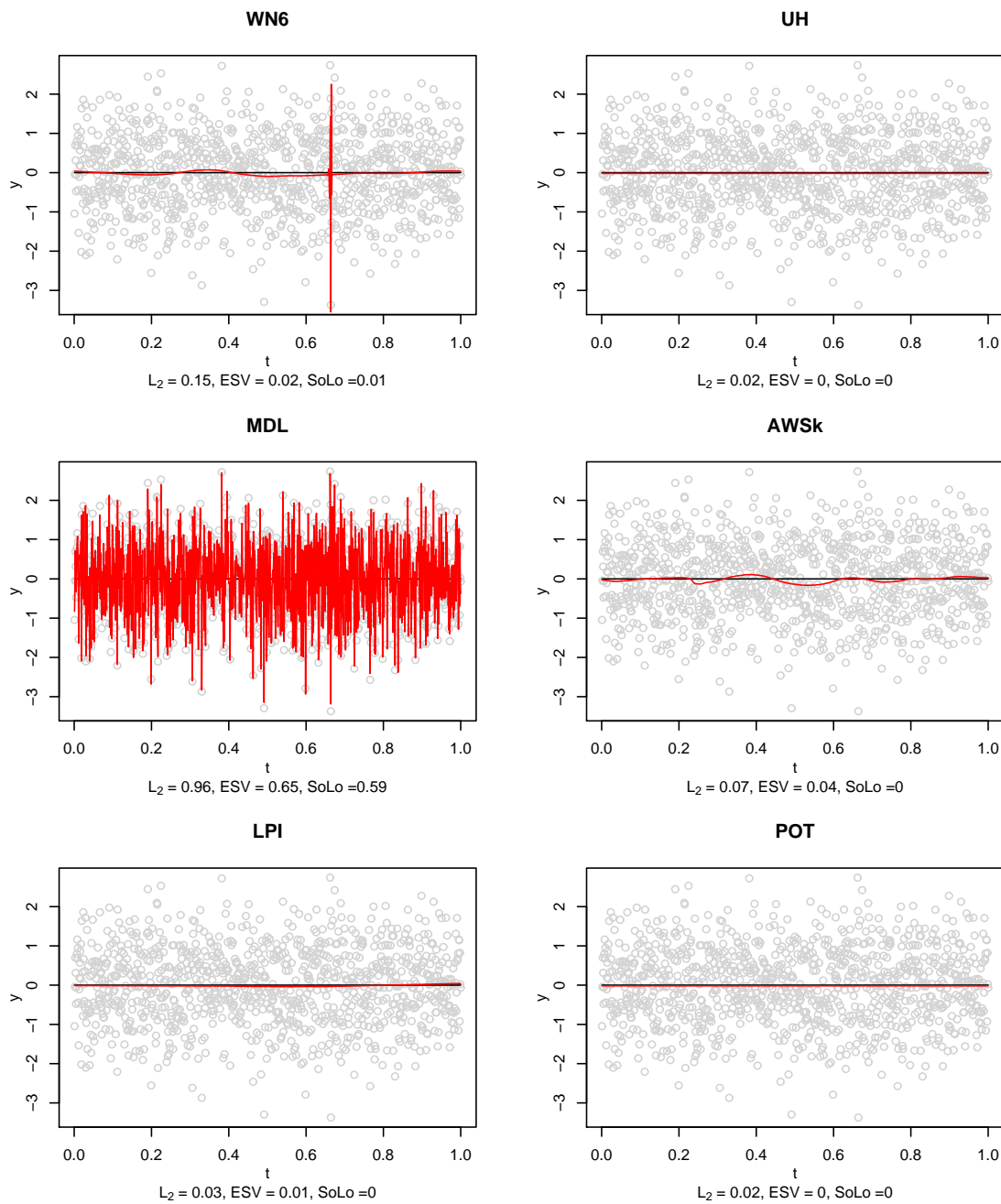
- Die ℓ^Y -Boxplots lassen erkennen, dass die UH und die POT-Methode auch die Höhe der Extremwerte richtig schätzen. AWSk liefert ebenfalls gute Ergebnisse. Die Boxplots veranschaulichen zudem, dass WN6 und MDL nicht nur einen schlechten Extremstellenverlust haben, sondern auch in Y -Richtung keine guten Schätzungen erbringen.
- Der Softloss spiegelt die Ergebnisse von ESV und ℓ^Y gemeinsam wieder: UH und POT schneiden am Besten ab, gefolgt von TS und STS, während WN6, MDL und LPI schlechtere Ergebnisse aufweisen.
- Die Mittelwerttabelle 7.7 zeigt, dass TS und POT auch im Mittel einen ESV von 0 haben. Die Blocks-Funktion wird selten überglättet, wie die kleinen $loss_g^Y$ -Werte verdeutlichen.
- Bei Betrachtung der Ränge für die Blocks-Funktion in Tabelle 7.8 fällt auf, dass POT unabhängig von der benutzten Gütefunktion die beste Methode ist. Insgesamt gibt es zwischen den Methoden keine so großen Unterschiede wie bei anderen Funktionen. Insbesondere unterscheiden sich die Ränge für Hausdorff-Abstand und ESV nur wenig.

Konstante Funktion

- Die konstante Funktion hat keine Extremstellen und stellt damit einen Sonderfall dar. Der ESV ist für diesen Fall definiert als Anteil der geschätzten Extremstellen an der Anzahl Designpunkte (vgl. Definition 3.11). Hausdorff-Abstand, ℓ^Y und $loss_g^Y$ sind wegen $n_{true} = 0$ nicht definiert oder konstant. Der Algorithmus der STS-Methode liefert für diese Daten häufig kein Ergebnis. Daher basieren die hier vorgestellten Ergebnisse nur auf 371 Stichproben. Bei $n = 256$ wurden für 417 der 1000 Stichproben Ergebnisse berechnet.
- In Abbildung 7.9 fällt auf, dass die MDL-Methode weißes Rauschen nicht als solches erkennt, sondern ein Signal schätzt. Die Methode benutzt etwa 65% der Waveletkoeffizienten von weißem Rauschen für die Schätzung (vgl. Davies, Gather, Weinert 2008). Der ESV-Wert ist entsprechend hoch. Die Werte des L_2 -Abstands

und des Softloss sind verglichen mit den anderen Methoden ebenso groß.

- Die Methoden UH und POT schätzen eine konstante Funktion, die gering von der wahren abweicht, was sich im Wert des L_2 -Abstands von jeweils 0,02, nicht jedoch im ESV- und Softloss-Wert von jeweils 0 zeigt. Die Methode LPI schätzt nahezu eine Konstante, weicht aber von der wahren Funktion ab und schneidet diese an einer Stelle. Der ESV zeigt, dass es überflüssige Extremstellen gibt, der Softloss hingegen nicht.
- Die WN6-Methode schätzt keine Konstante und liefert sogar große überflüssige Extremstellen. Der ESV-Wert ist 0,02. Die AWSk-Methode zeigt keine Konstante sondern eine *Wellenlinie* um die wahre Funktion. Der ESV-Wert ist mit 0,04 größer als bei WN6. Dies macht deutlich, dass der ESV die Größe von Extremstellen nicht mit einbezieht. Beim Softloss hingegen ist WN6 mit 0,01 etwas schlechter als AWSk mit 0. Die Größe der Extremstellen der AWSk-Schätzung ist so gering, dass der Softloss sie nicht anzeigt.
- In Abbildung 7.2 wird der Boxplot von MDL nicht dargestellt, weil er viel größere Werte enthält als die Boxplots der anderen Methoden. Die Ergebnisse der verschiedenen Methoden zeigen wenige Unterschiede gemessen an dem L_2 -Abstand. UH und POT beschreiben bessere Ergebnisse als die anderen Methoden.
- Die Boxplots des ESV verdeutlichen, dass UH, TS, STS und POT nur äußerst selten überflüssige Extremstellen schätzen. AWSk und LPI zeigen hingegen häufiger als die anderen Methoden überflüssige Extremstellen. Insgesamt liegen alle ESV-Werte unter 0,08 und damit liefern die Methoden ein gutes Ergebnis.
- Die Tabelle 7.2 der Mittelwerte bringt keine neuen Erkenntnisse, MDL zeigt schlechte Ergebnisse, die anderen Methoden gute. Anhand der Tendenz des Vorzeichens lässt sich erkennen, dass POT in weniger als 100 der 1000 Stichproben Extremstellen schätzt, die anderen Methoden in mehr. STS schätzt in weniger als 100 von nur 371 berechneten Stichproben Extremstellen. Da die Ergebnisse dicht beieinander liegen, ist es nicht sinnvoll, Ränge über diese Mittelwerte zu bilden.

Abbildung 7.9: Beispiele für Schätzungen der Konstanten, $n = 1024$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS*	POT
L_1 -Abstand	0,07	0,07	0,03	0,78	0,08	0,07	0,06	0,05	0,06	0,03
L_2 -Abstand	0,1	0,1	0,04	0,98	0,1	0,09	0,07	0,07	0,07	0,03
ESV (Betrag)	0	0,01	0	0,65	0,03	0,01	0,02	0	0	0
Tendenz	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(o)	(o)
Softloss	0	0	0	0,6	0	0	0	0	0	0

Tabelle 7.9: Mittelwert der Ergebnisse für die Konstante, $n = 1024$

* Für die STS-Methode nur aus 371 Stichproben bestimmt

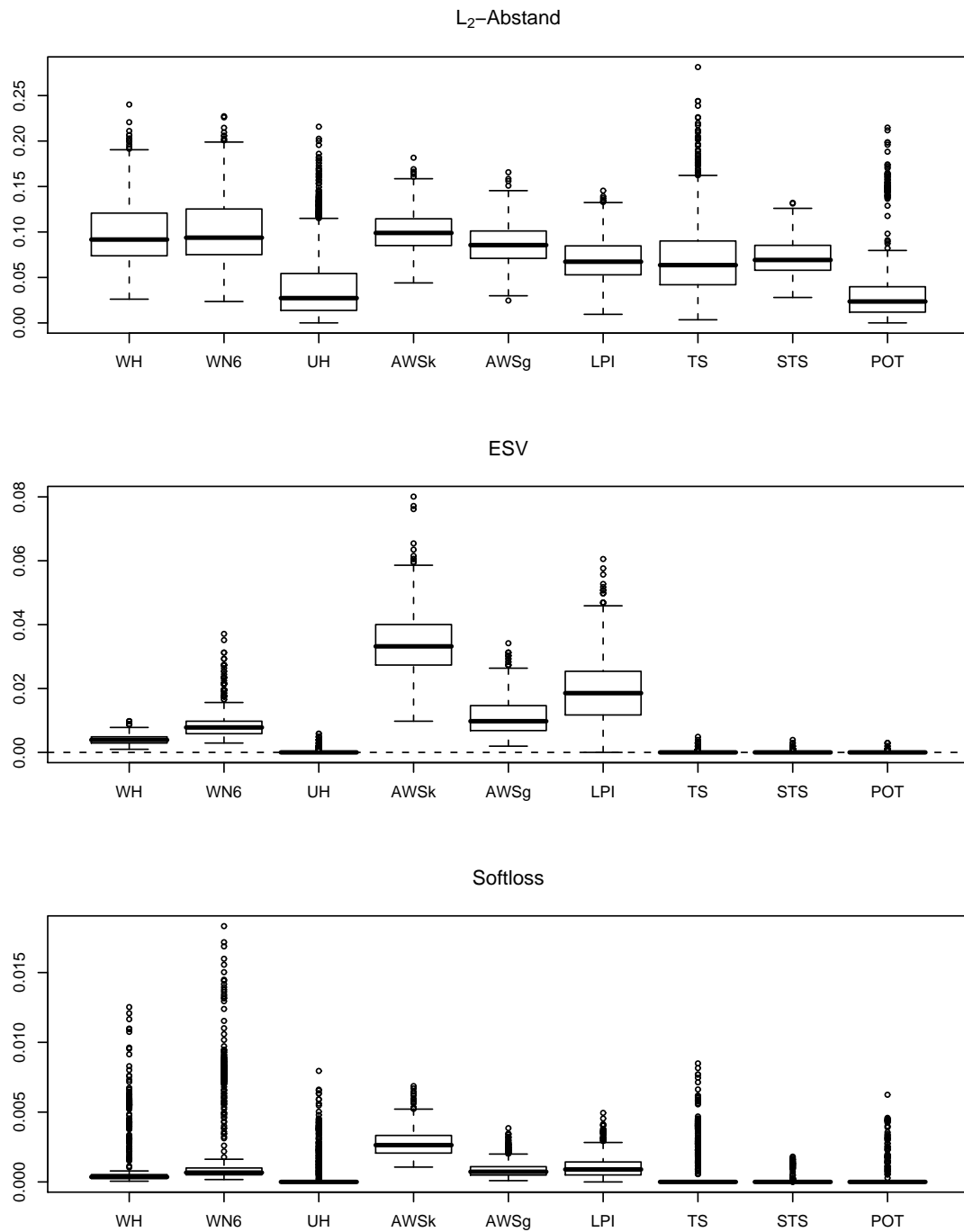


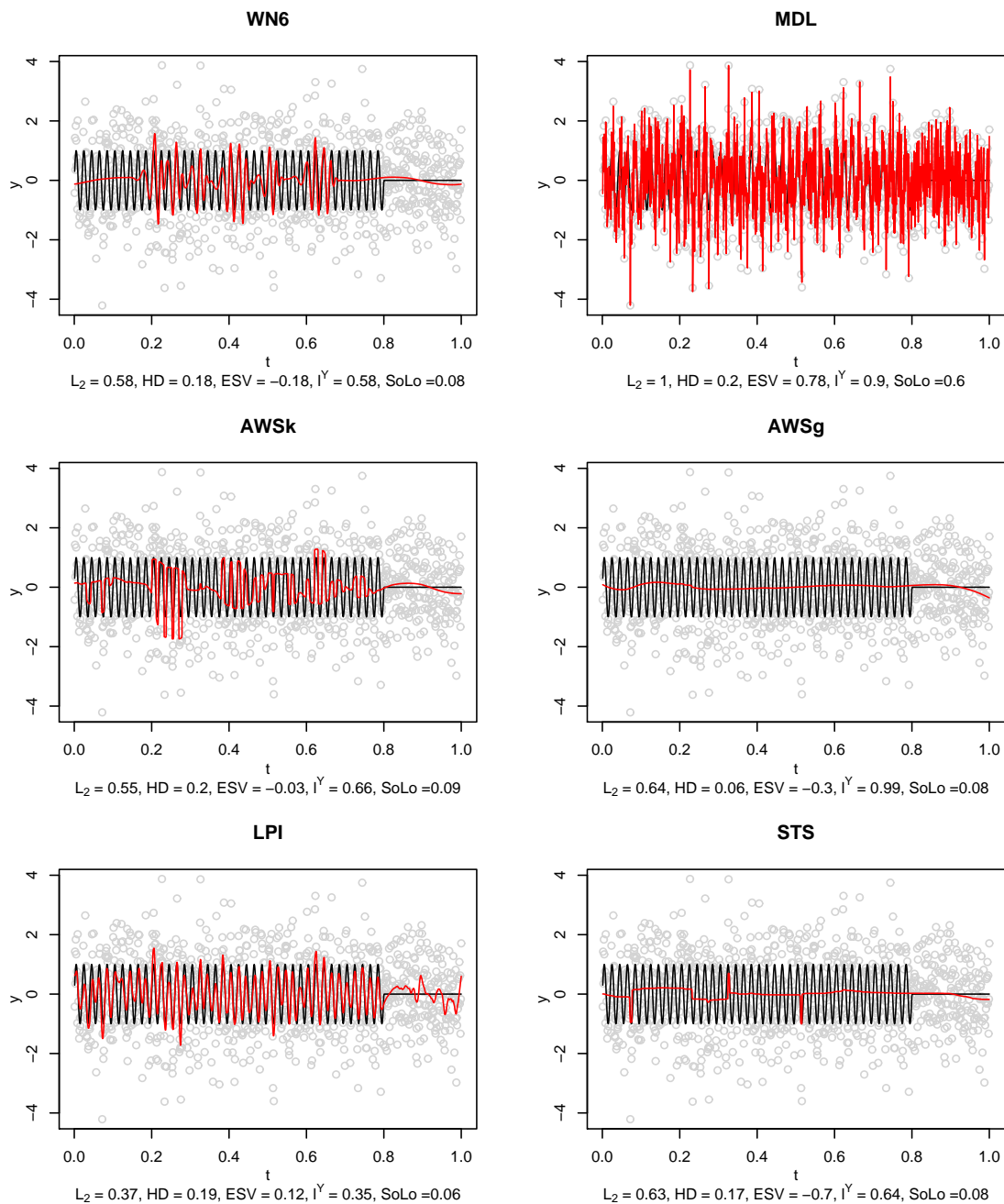
Abbildung 7.10: Ergebnisse für die Konstante, $n = 1024$
 Für die STS-Methode nur aus 371 Stichproben bestimmt

Sine-Funktion

- Die Sine-Funktion ist ein Sonderfall, da sie viele Extremstellen hat und sich nicht deutlich von weißem Rauschen unterscheidet. Bei $n = 256$ fällt auf, dass keine der Methoden das Signal erkennen kann. Auch bei dieser Funktion kann die Software der STS-Methode nicht für alle Stichproben ein Ergebnis liefern. Die gezeigten Berechnungen für $n = 1024$ basieren auf 950 Stichproben, bei $n = 256$ wurden 434 Stichproben berechnet.
- Das Signal der Sine-Funktion in Abbildung 7.11 ist für die verschiedenen Methoden offensichtlich nicht leicht zu erkennen. Die Methoden WN6 und AWSk erfassen einige der Extremstellen der Sine-Funktion. Auffällig ist der geringe Wert des ESV von -0,03 bei der AWSk-Methode. Offensichtlich hat die Methode Extremstellen richtig geschätzt, die aber so geringe Höhe haben, dass sie nicht erkennbar sind. Gleiches gilt für die AWSg-Methode, die einen niedrigen Wert beim ESV zeigt.
- Die Methode LPI schätzt zwar die Extremstellen der Funktion gut, liefert aber einige überflüssige Extremwerte im konstanten Bereich der Sine-Funktion. Der ESV gibt dies durch den Wert 0,12 mit positivem Vorzeichen wieder. Die Höhe der Extremwerte wird gut durch LPI geschätzt, der ℓ^Y -Wert beträgt 0,35.
- Die MDL-Methode liefert zahlreiche überflüssige Extremwerte. Die richtig geschätzten entsprechen in der Höhe nicht denen der Sine-Funktion. Beides wird durch ESV und ℓ^Y wiedergegeben.
- Der Hausdorff-Abstand gibt die Unterschiede zwischen den Methoden gar nicht wieder, die Werte liegen fast alle bei etwa 0,2. Der Softloss zeigt wenig Unterschiede und kleine Werte außer bei der Methode MDL.
- Bei Betrachtung der Boxplots in Abbildung 7.2 des L_2 -Abstands fällt auf, dass sie alle konzentriert sind. Die meisten liegen bei etwa 0,6. Nur die Methoden MDL und LPI liefern schlechtere bzw. bessere Ergebnisse.
- Der ESV differenziert mehr zwischen den Methoden. MDL und LPI schätzen eher zu viele Extremstellen, während die anderen Methoden eher zu wenige anzeigen.

Insgesamt präsentieren WN6 und AWSk häufig richtige Ergebnisse. WN6 schätzt dabei mal zu viele und mal zu wenige Extremstellen.

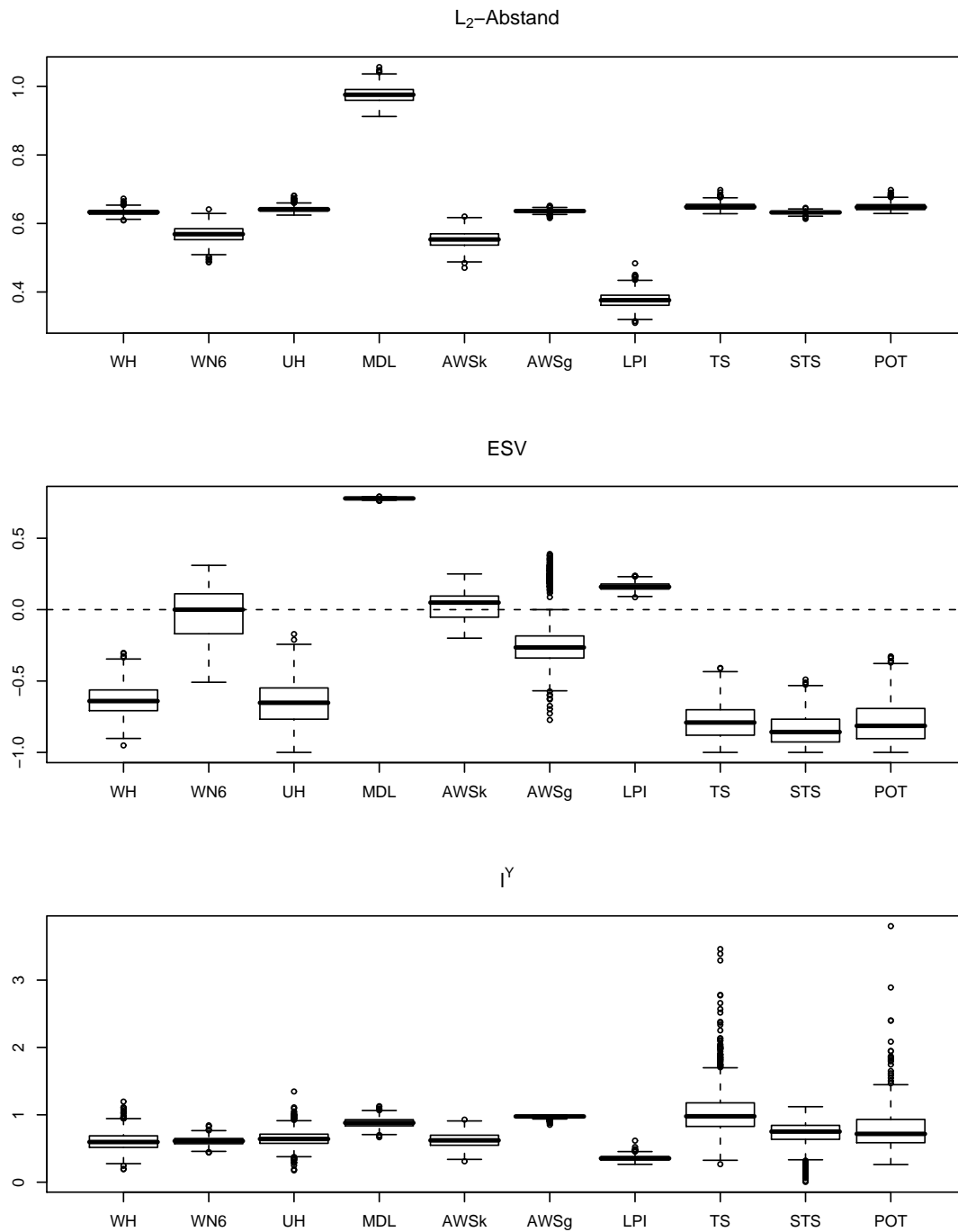
- ℓ^Y variiert wenig zwischen den Methoden. Zu beachten ist, dass die absolute Höhe der Extremwerte der Sine-Funktion etwa 1 ist und somit eine Schätzung nahe 0 zu einem ℓ^Y -Wert von 1 führt. Bei Betrachtung der durchschnittlichen ℓ^Y -Werte zeigt sich, dass dies für die Methoden AWSg und TS gilt.
- Aus dem durchschnittlichen ESV-Wert von 0,29 für AWSg in Tabelle 7.2 lässt sich erkennen, dass das Beispiel aus Abbildung 7.11 nicht typisch ist. Die Methode AWSk zeigt hingegen wie im Bild so auch im Durchschnitt einen kleinen ESV-Wert von 0,08.
- Bei Betrachtung von $loss_g^Y$ fällt auf, dass AWSk mit einem Wert von 1 im Durchschnitt die Sine-Funktion immer zu stark glättet, anders als aus Abbildung 7.11 zu vermuten ist. Insbesondere bei der Sine-Funktion ist zu beachten, dass in diesen Verlust nur diejenigen Extremwerte eingehen, deren zugehörige Stellen auch identifiziert wurden.
- L_1 -, L_2 - und Hausdorff-Abstand zeigen nur geringe Unterschiede zwischen den Methoden, ebenso der Softloss. Einzig die MDL-Methode liefert bei letzterem einen schlechteren Wert als die übrigen Methoden, bei denen die Unterschiede so gering sind, dass es nicht sinnvoll ist, Ränge zu bilden.
- Bei den Rängen in Tabelle 7.11 beurteilen Hausdorff-Abstand und ESV die Güte der Methoden ähnlich.
- Die Sine-Funktion weist große Unterschiede zwischen den Stichprobengrößen auf. Bei 256 Datenpunkten werden viel weniger Extremstellen erkannt. Die durchschnittlichen ESV-Werte sind bis auf die bei MDL größer als 0,5. Die POT-Methode hat sogar einen Durchschnitt von 1. Insgesamt differenzieren die verschiedenen Verluste weniger zwischen den Methoden als bei $n = 1024$.

Abbildung 7.11: Beispiele für Schätzungen der Sine-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS*	POT
L_1 -Abstand	0,52	0,45	0,52	0,77	0,43	0,52	0,3	0,53	0,52	0,52
L_2 -Abstand	0,63	0,57	0,64	0,98	0,55	0,64	0,38	0,65	0,63	0,65
Hausdorff-Abstand	0,16	0,17	0,2	0,2	0,15	0,13	0,2	0,27	0,33	0,31
ESV (Betrag)	0,64	0,14	0,65	0,78	0,08	0,29	0,16	0,78	0,84	0,79
Tendenz	(-)	(o)	(-)	(+)	(+)	(-)	(+)	(-)	(-)	(-)
l^Y	0,61	0,61	0,65	0,88	0,62	0,97	0,36	1,05	0,72	0,77
$loss_g^Y$	0,65	0,73	0,54	0,14	0,91	1,00	0,78	0,27	0,90	0,31
Softloss	0,08	0,08	0,08	0,59	0,09	0,08	0,06	0,08	0,08	0,08

Tabelle 7.10: Mittelwert der Ergebnisse für die Sine-Funktion, $n = 1024$

* Für die STS-Methode nur aus 950 Stichproben bestimmt

Abbildung 7.12: Ergebnisse für die Sine-Funktion, $n = 1024$

Für die STS-Methode aus 950 Stichproben bestimmt

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	4	3	6	10	2	7	1	9	5	8
L_2 -Abstand	5	3	7	10	2	6	1	9	4	8
Hausdorff-Abstand	3	4	6	7	2	1	5	8	10	9
ESV (Betrag)	5	2	6	7	1	4	3	8	10	9

Tabelle 7.11: Ränge der mittleren Ergebnisse für die Sine-Funktion, $n = 1024$

Step-Funktion

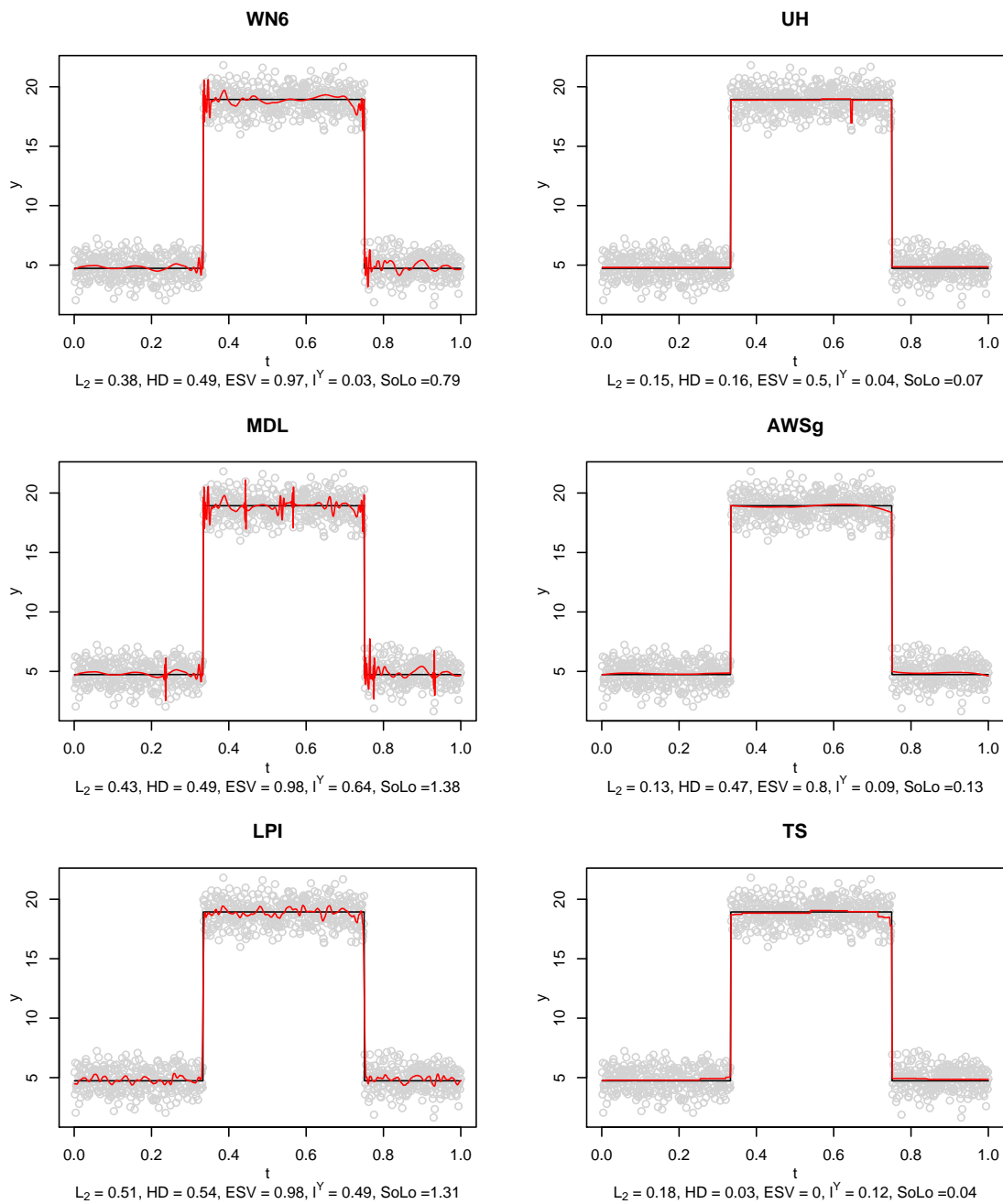
- Die Step-Funktion hat nur eine Extremstelle. In Abbildung 7.13 scheinen UH, AWSg und TS die Funktion gut zu erkennen. Der ESV-Wert von TS ist 0, diese Methode schätzt keine überflüssigen Extremstellen. UH und AWS weisen größere ESV-Werte auf. Da nur eine wahre Extremstelle existiert, hat die Anzahl geschätzter Extremstellen relativ großen Einfluss auf das ESV-Ergebnis. Wenn $n_{id} = n_{true} = 1$ gilt beispielsweise $ESV=0,8$ für $n_{est} = 9$, $ESV=0,9$ für $n_{est} = 19$ und $ESV=0,99$ für $n_{est} = 199$. Die anderen abgebildeten Methoden, WN6, MDL und LPI, schätzen viele überflüssige Extremstellen. Die hohen ESV-Werte von nahezu 1 geben das wieder.
- Der Hausdorff-Abstand liefert für die Methoden WN6, MDL, AWSg und LPI Werte um 0,5, beurteilt damit also die AWSg-Methode genauso wie die anderen. Gemessen an dem L_2 -Abstand fallen die Abweichungen der LPI-Schätzung am meisten ins Gewicht, gefolgt von MDL und WN6. Von den übrigen Methoden hat TS den größten Wert. ℓ^Y liefert in diesem Fall nur den Abstand zwischen dem einen wahren Maximum und dem zugehörigen geschätzten. Die Ergebnisse veranschaulichen, dass die TS-Schätzung zu diesem Maximum einen größeren Abstand hat als WN6, UH und AWSg.
- Die Boxplots in Abbildung 7.14 zeigen, dass sich die L_2 -Ergebnisse bei den verschiedenen Methoden stark unterscheiden. Für eine einzelne Methode sind sie sehr konzentriert. Die Methoden WN6, MDL und LPI schneiden schlechter ab als die anderen.
- Bei Betrachtung der ESV-Boxplots ist ersichtlich, dass die Extremstelle der Step-Funktion entweder gut erkannt wird, oder viele überflüssige Extremwerte geliefert werden. Die Methoden UH, TS, STS und POT haben in den meisten Fällen einen ESV von 0, die anderen Methoden liegen nahe 1.
- Der Softloss differenziert wenig zwischen den Methoden. Nur MDL und LPI haben große Werte.
- Die Tabelle 7.12 der Mittelwerte und die der zugehörigen Ränge 7.13 zeigen, dass sich die Güte der Methoden gemessen mit L_1 - und L_2 -Abstand nur we-

nig unterscheiden. Außerdem fällt auf, dass die POT-Methode wie schon bei der Blocks-Funktion bei allen Gütemaßen am besten abschneidet. Ebenso ist die Variation in den Rängen bei verschiedenen Gütemaßen für eine Methode insgesamt nicht so groß wie bei anderen Funktionen.

- Wie bereits bei der Blocks-Funktion sind auch bei der Step-Funktion die $loss_g^Y$ -Werte verhältnismäßig klein. Der Extremwert wird also selten unterschätzt.

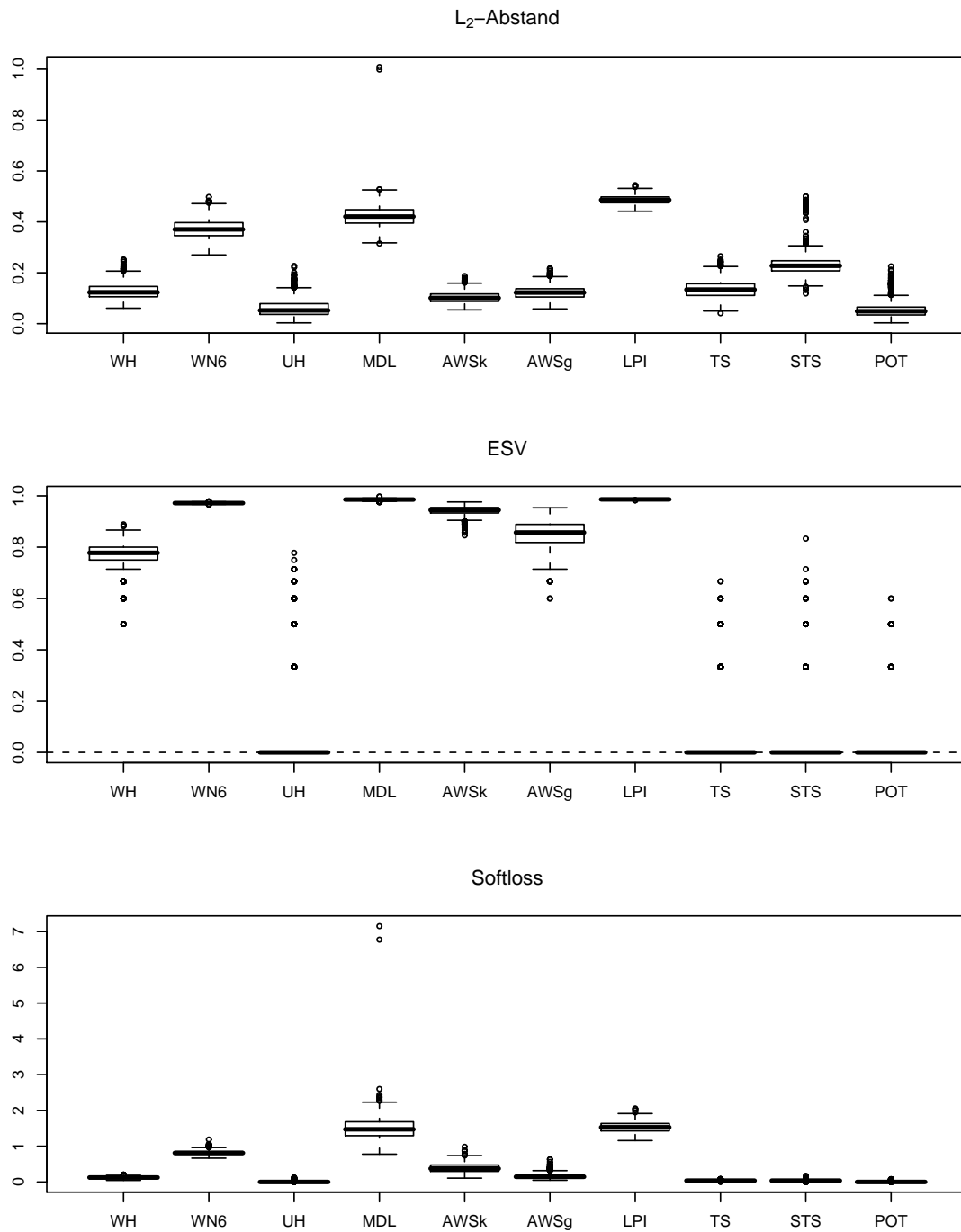
Blip-Funktion

- Von den in Abbildung 7.15 gezeigten Methoden liefern WH, TS, STS und POT die richtige Anzahl Extremstellen, was durch den ESV-Wert von jeweils 0 wiedergegeben wird. WN6 und LPI erzeugen zu viele Extremstellen und der ESV-Wert ist entsprechend groß. Der Hausdorff-Abstand zeigt ähnliche Ergebnisse.
- Die STS-Methode schätzt die Funktion zu glatt. Es ist deutlich, dass die Maxima zu klein und die Minima zu groß geschätzt werden. Für die POT-Methode gilt das gleiche, jedoch nicht so stark ausgeprägt. Die anderen Methoden neigen eher dazu, das Maximum bei 0,8 zu überschätzen und das folgende Minimum zu unterschätzen. Durch die Verlustfunktionen werden diese Ergebnisse nicht wiedergegeben.
- ℓ^Y ist für TS, STS und POT verhältnismäßig groß und gibt den Eindruck wieder, der bei Betrachtung der Abweichung der geschätzten Graphen von den wahren entsteht.
- Die Boxplots des Hausdorff-Abstands in Abbildung 7.16 zeigen, dass die Ergebnisse der beiden AWS-Methoden viel stärker variieren als die der anderen Methoden. Verglichen mit den Boxplots des ESV wird deutlich, dass beide Gütekriterien die gleichen Methoden als gut oder schlecht beurteilen.
- Der Softloss beurteilt WN6, MDL und LPI schlechter als die anderen Methoden, zwischen denen durch diesen Verlust kaum differenziert wird.

Abbildung 7.13: Beispiele für Schätzungen der Step-Funktion, $n = 1024$

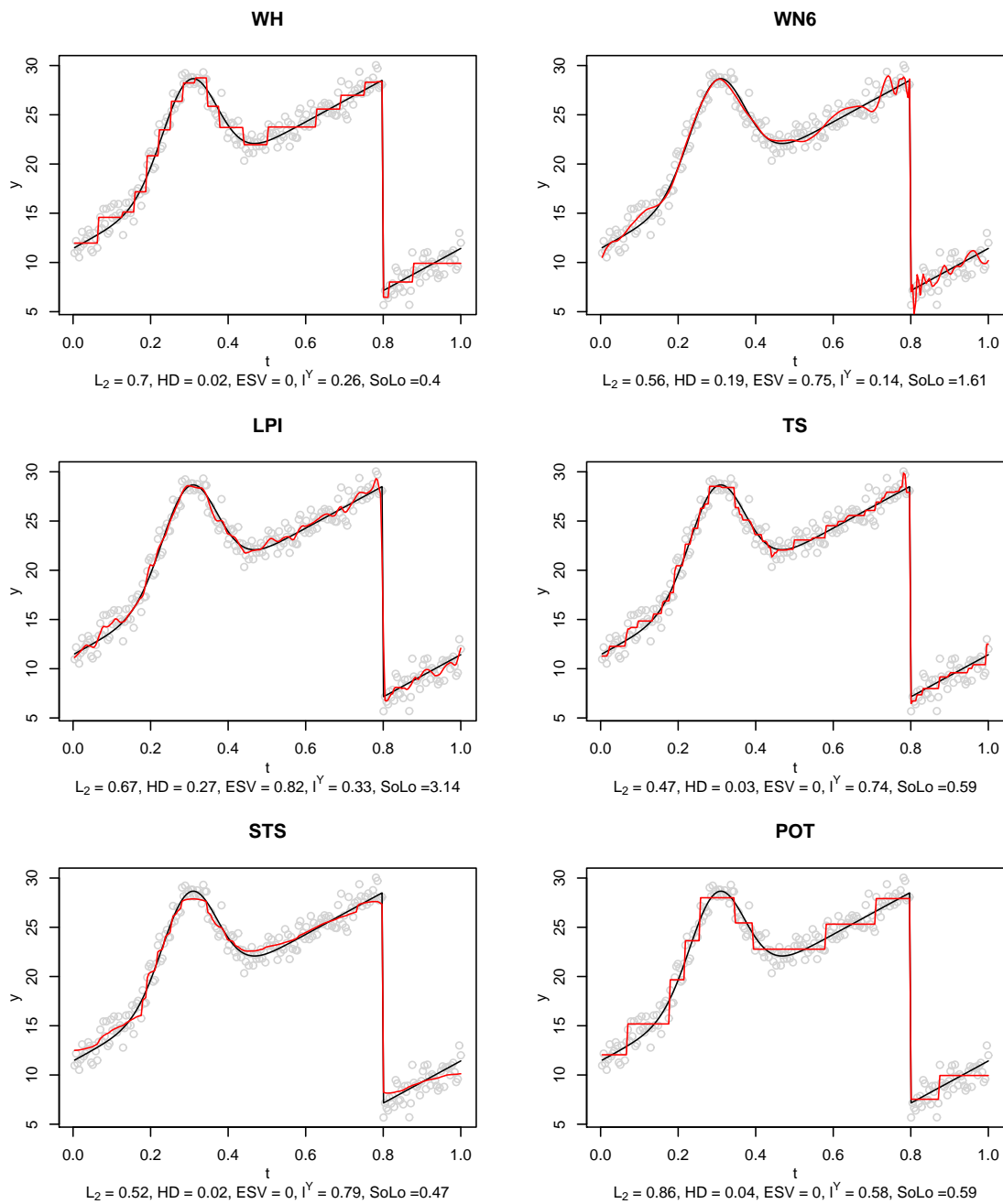
	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,08	0,24	0,05	0,27	0,08	0,09	0,25	0,08	0,13	0,04
L_2 -Abstand	0,13	0,37	0,06	0,42	0,1	0,12	0,49	0,14	0,24	0,05
Hausdorff-Abstand	0,33	0,51	0,05	0,52	0,49	0,46	0,53	0,13	0,07	0,01
ESV (Betrag)	0,77	0,97	0,07	0,99	0,94	0,85	0,99	0,04	0,04	0,01
Tendenz	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(o)	(o)
l^Y	0,15	0,19	0,1	0,35	0,09	0,1	0,25	0,34	0,08	0,08
$loss_g^Y$	0,14	0,15	0,48	0,13	0,33	0,20	0,24	0,11	0,20	0,50
Softloss	0,12	0,82	0,01	1,51	0,39	0,16	1,53	0,04	0,04	0

Tabelle 7.12: Mittelwert der Ergebnisse für die Step-Funktion, $n = 1024$

Abbildung 7.14: Ergebnisse für die Step-Funktion, $n = 1024$

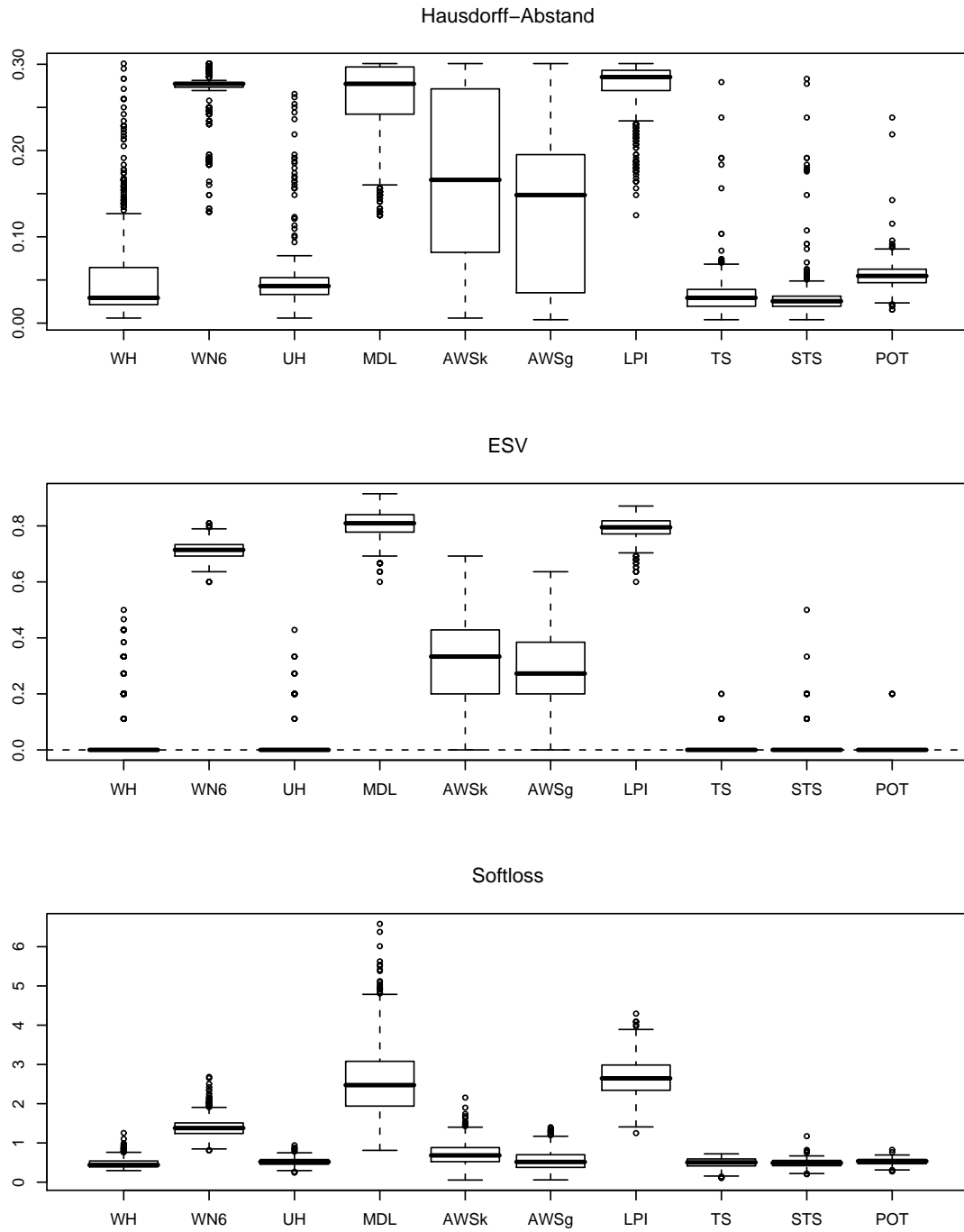
	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	4	8	2	10	3	6	9	5	7	1
L_2 -Abstand	5	8	2	9	3	4	10	6	7	1
Hausdorff-Abstand	5	8	2	9	7	6	10	4	3	1
ESV (Betrag)	5	8	4	9	7	6	10	2	3	1
Softloss	5	8	2	9	7	6	10	3	4	1

Tabelle 7.13: Ränge der mittleren Ergebnisse für die Step-Funktion, $n = 1024$

Abbildung 7.15: Beispiele für Schätzungen der Blip-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,56	0,42	0,55	0,4	0,53	0,23	0,36	0,35	0,42	0,67
L_2 -Abstand	0,7	0,57	0,68	0,56	0,67	0,3	0,68	0,44	0,54	0,83
Hausdorff-Abstand	0,05	0,27	0,05	0,26	0,17	0,14	0,28	0,03	0,03	0,05
ESV (Betrag)	0,06	0,71	0,01	0,81	0,31	0,26	0,79	0	0	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(o)	(o)	(o)
l^Y	0,39	1	0,56	1,3	0,59	0,34	0,84	0,49	0,91	0,77
$loss_g^Y$	0,61	0,49	0,87	0,35	0,92	0,54	0,39	0,40	0,99	0,95
Softloss	0,47	1,4	0,52	2,6	0,72	0,55	2,67	0,49	0,49	0,53

Tabelle 7.14: Mittelwert der Ergebnisse für die Blip-Funktion, $n = 256$

Abbildung 7.16: Ergebnisse für die Blip-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	9	5	8	4	7	1	3	2	6	10
L_2 -Abstand	9	5	8	4	6	1	7	2	3	10
Hausdorff-Abstand	4	9	3	8	7	6	10	2	1	5
ESV (Betrag)	5	8	4	10	7	6	9	1	3	2
Softloss	1	8	4	9	7	6	10	3	2	5

Tabelle 7.15: Ränge der mittleren Ergebnisse für die Blip-Funktion, $n = 256$

- Die Mittelwerte von $loss_g^Y$ in Tabelle 7.14 zeigen erneut, dass fast alle stückweise konstanten Schätzungen die Funktion zu stark glätten. Nur TS hat einen kleineren $loss_g^Y$ -Wert von 0,4. Von den stetigen Schätzungen zeigt STS auch bei dieser Funktion einen großen $loss_g^Y$ -Wert.
- Anhand der Ränge in Tabelle 7.15 wird deutlich, dass der L_1 - und der L_2 -Abstand andere Methoden als gut beurteilen als die anderen Verluste. Die Ergebnisse von Hausdorff-Abstand und ESV ähneln sich stark.

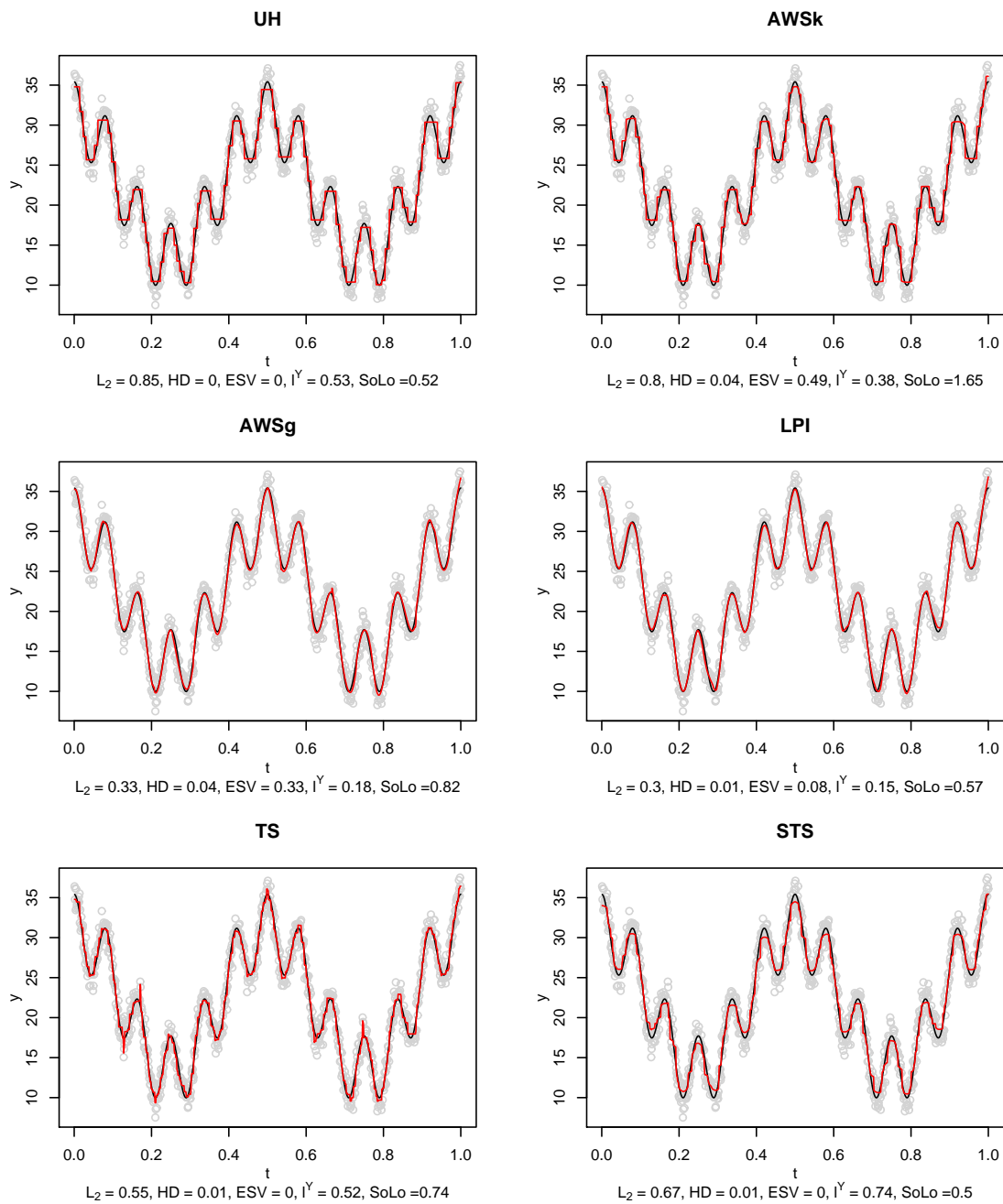
Wave-Funktion

- Abbildung 7.17 belegt, dass die Methoden die Wave-Funktion recht gut schätzen. Auf den ersten Blick scheinen alle Verfahren die Extremwerte richtig zu erkennen. Die ESV-Werte zeigen, dass die beiden AWS-Methoden überflüssige Extremstellen liefern, die im Graphen nicht zu sehen sind. Die anderen Verfahren haben einen kleinen ESV-Wert. Die LPI-Methode schätzt dabei auch die Höhe der Extremwerte besser als die anderen Methoden, der ℓ^Y -Wert ist mit 0,15 der kleinste.
- Der Hausdorff-Abstand ist bei allen Methoden auffällig klein. Die Werte des L_2 -Abstands zeigen große Abstände von wahrer und geschätzter Funktion für die Methoden UH und AWSk, mittelgroße für TS und STS und kleine für AWSg und LPI.
- Der Softloss liefert kleine Ergebnisse für UH, LPI und STS. Das gilt, obwohl UH und STS sichtbare Abweichungen von der wahren Funktion aufweisen.
- Die Boxplots für den L_2 -Abstand in Abbildung 7.18 zeigen, dass die Beispiele aus Abbildung 7.17 weitgehend typisch sind. AWSg und LPI haben kleine Werte, UH und AWSk große. STS liefert schlechtere Ergebnisse als TS.
- Die ESV-Boxplots verdeutlichen, dass es bei der Wave-Funktion einige Methoden gibt, die die Extremstellen der Funktion nahezu immer richtig schätzen. Die Boxplots von WH, WN6, UH, TS, STS und POT sind bei 0 konzentriert. Die anderen Verfahren, besonders MDL, schätzen zu viele Extremstellen.

- Mittels der Boxplots von ℓ^Y wird ersichtlich, dass einige der Methoden, die die Extremstellen richtig schätzen, die zugehörigen Extremwerte nicht so gut bestimmen. STS und POT zeigen größere Abweichungen von den wahren Werten, TS und WN6 kleinere. Die Methoden, die einen schlechteren ESV-Wert haben, liefern dagegen kleinere ℓ^Y -Werte.
- Die Tabelle 7.16 der Mittelwerte bestätigt den Eindruck, der für den Hausdorff-Abstand aus Abbildung 7.17 gewonnen wurde. Die Mittelwerte unterscheiden sich kaum.
- Auffällig ist, dass vier der Methoden im Mittel einen ESV von 0 haben. Die Tendenz zeigt an, dass diese dabei seltener als 100 mal ein positives Vorzeichen liefern. An den Boxplots ist erkennbar, dass sie nie zu negativen Ergebnissen führen.
- Bei Betrachtung der $loss_g^Y$ -Werte finden sich ähnliche Ergebnisse wie schon bei anderen Testbed-Funktionen: STS und die meisten stückweise konstanten Schätzungen haben einen großen Wert, TS einen kleinen.
- Der Softloss beurteilt die meisten Methoden ähnlich wie der ESV. Dabei zeigt TS aber eine deutliche Abweichung der beiden zugehörigen Ränge von 1 und 7 in Tabelle 7.17. Die Ränge der beiden Normen gleichen sich auch hier.

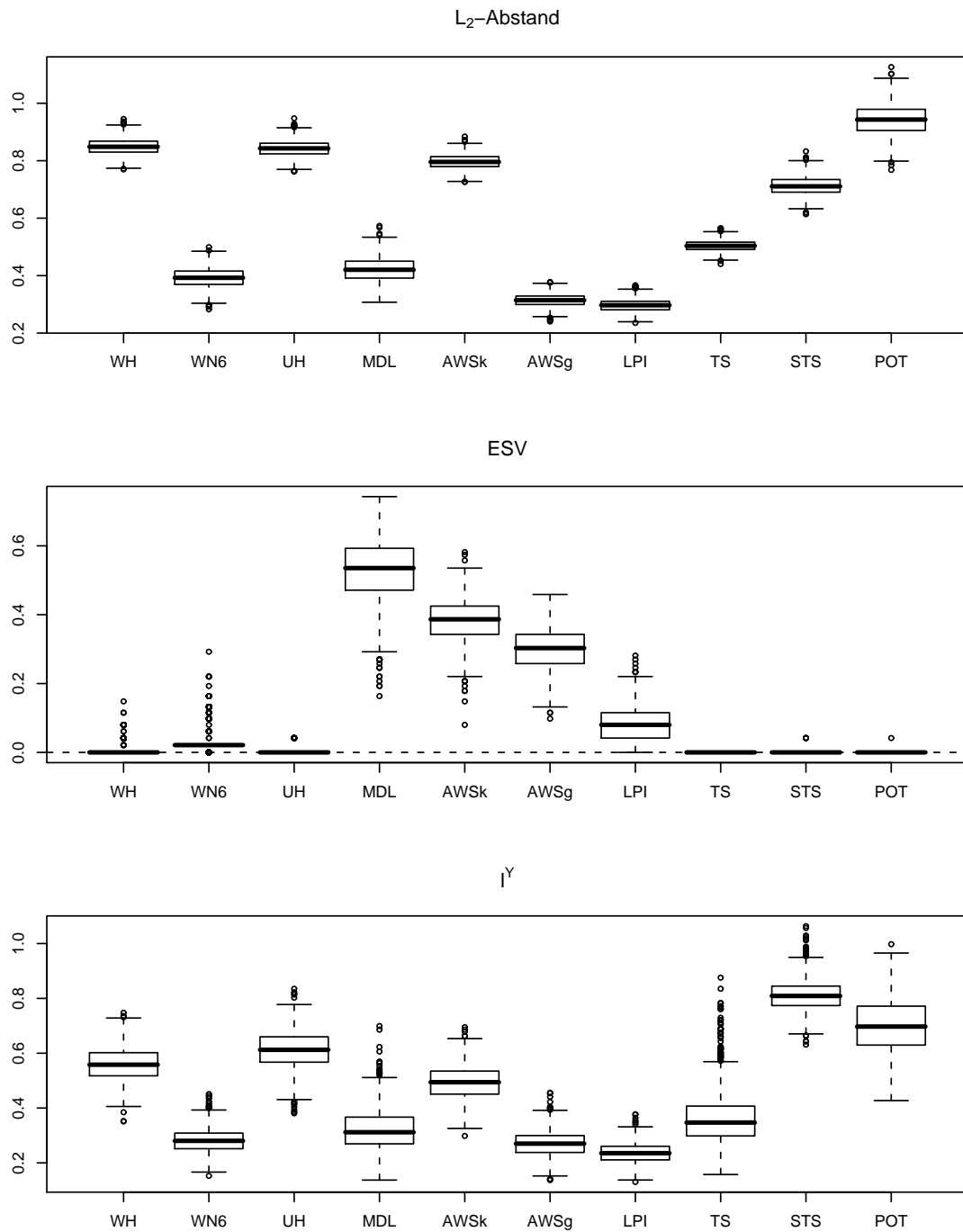
Spikes-Funktion

- In Abbildung 7.19 wird deutlich, dass die Methoden WH, UH und POT zu ähnlichen Ergebnissen führen. Dabei scheint WH die Höhe der Maxima etwas besser zu treffen als die anderen beiden Methoden, liefert aber ebenso überflüssige Extremwerte. Die Werte von ESV und ℓ^Y geben dies wieder. Der Softloss liefert für WH ein besseres Ergebnis als für die anderen beiden Methoden, während beim L_2 -Abstand POT einen größeren Wert aufweist.
- Auch bei dieser Funktion zeigen sich kaum Unterschiede zwischen den Hausdorff-Abstandswerten.

Abbildung 7.17: Beispiele für Schätzungen der Wave-Funktion, $n = 1024$

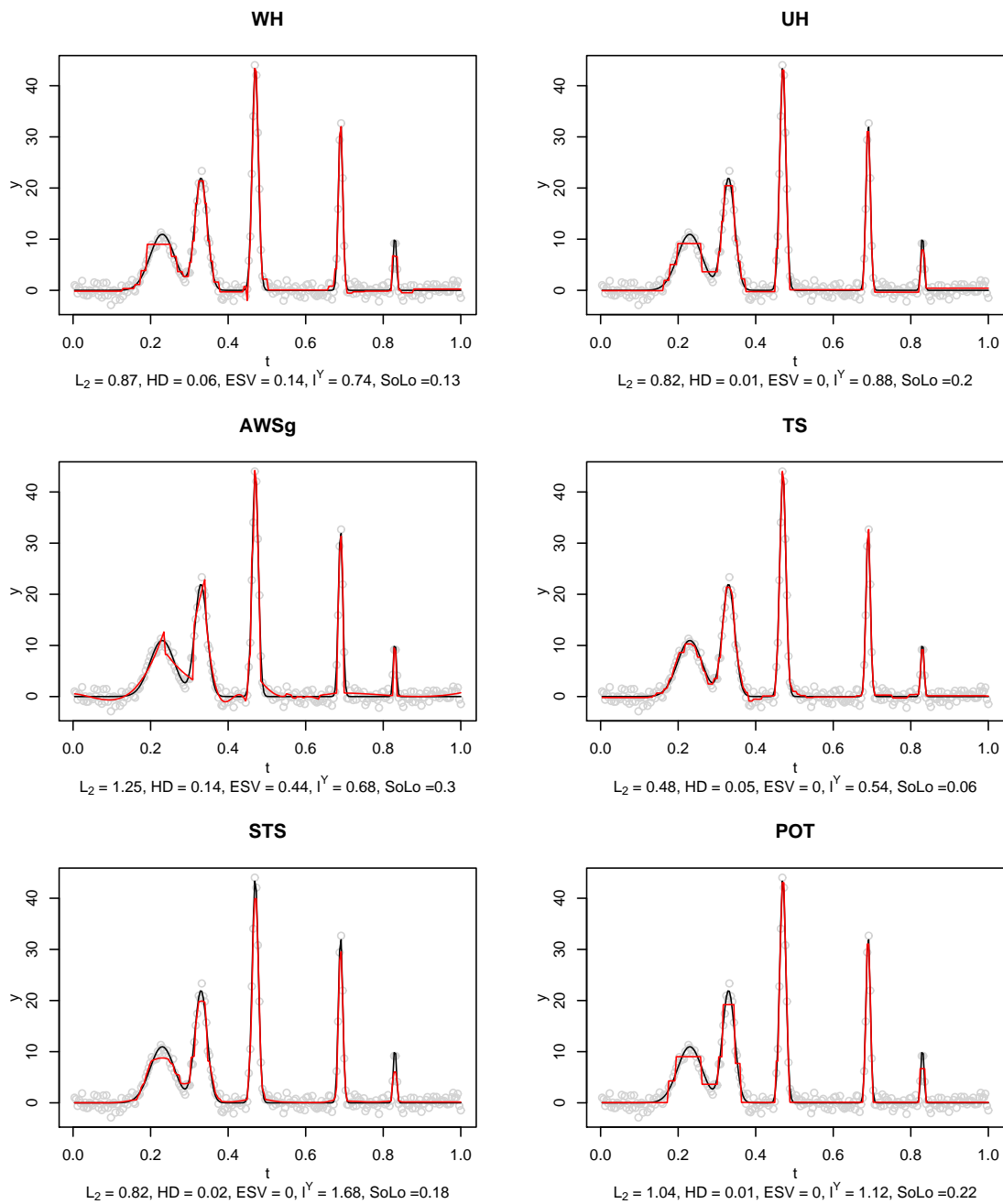
	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,68	0,31	0,68	0,3	0,63	0,25	0,24	0,4	0,58	0,76
L_2 -Abstand	0,85	0,39	0,84	0,42	0,8	0,31	0,3	0,5	0,71	0,94
Hausdorff-Abstand	0,01	0,04	0	0,04	0,03	0,03	0,01	0,01	0	0
ESV (Betrag)	0,01	0,03	0	0,52	0,38	0,3	0,09	0	0	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(o)	(o)	(o)
ℓ^Y	0,56	0,28	0,61	0,32	0,49	0,27	0,24	0,37	0,81	0,7
$loss_g^Y$	0,87	0,39	0,94	0,40	0,90	0,43	0,64	0,37	1,00	0,95
Softloss	0,63	0,51	0,5	1,56	1,4	0,9	0,56	0,77	0,49	0,47

Tabelle 7.16: Mittelwert der Ergebnisse für die Wave-Funktion, $n = 1024$

Abbildung 7.18: Ergebnisse für die Wave-Funktion, $n = 1024$

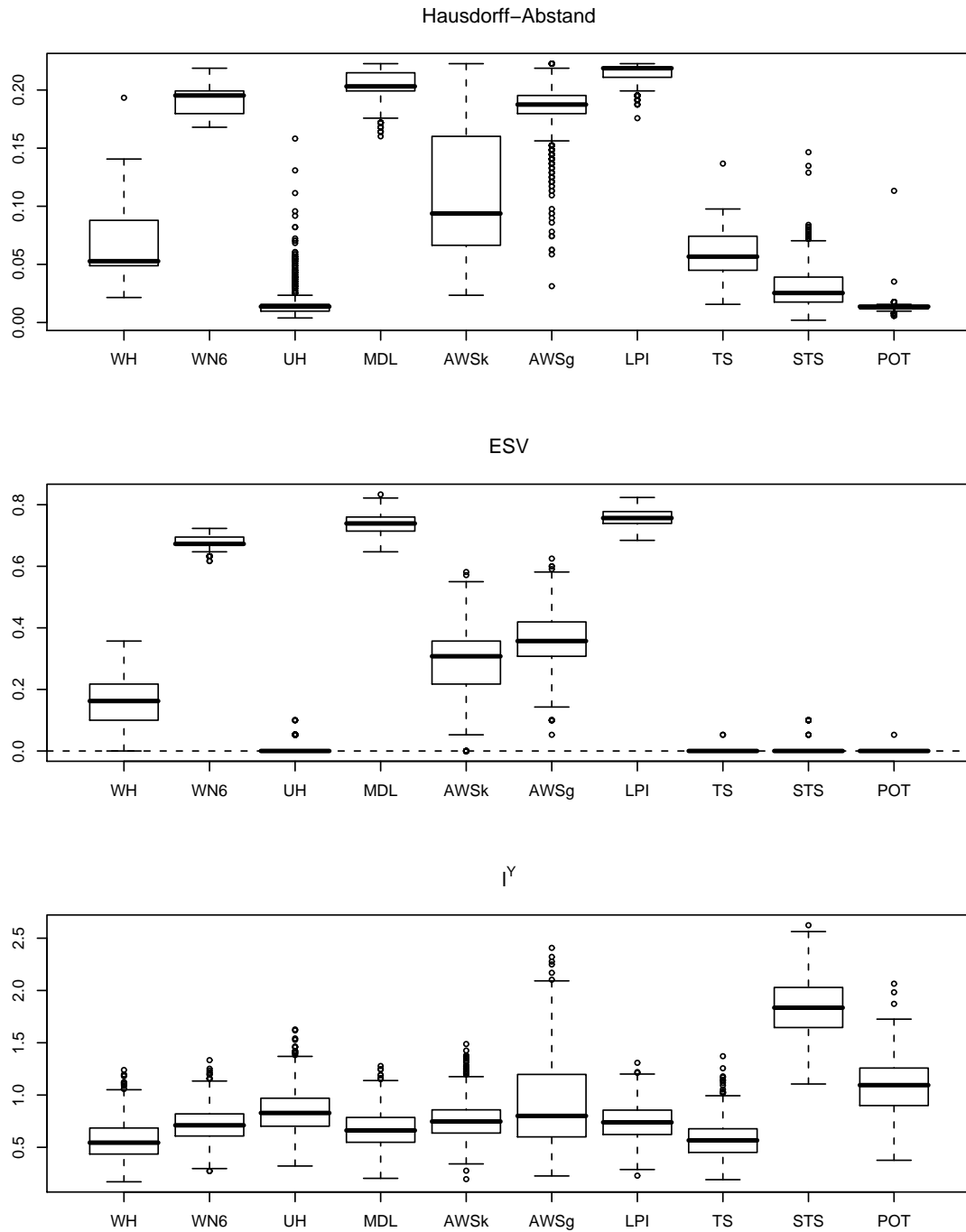
	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	9	4	8	3	7	2	1	5	6	10
L_2 -Abstand	9	3	8	4	7	2	1	5	6	10
ESV (Betrag)	5	6	1	10	9	8	7	1	1	1
Softloss	6	4	3	10	9	8	5	7	2	1

Tabelle 7.17: Ränge der mittleren Ergebnisse für die Wave-Funktion, $n = 1024$

Abbildung 7.19: Beispiele für Schätzungen der Spikes-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	0,52	0,58	0,53	0,53	0,57	0,62	0,49	0,37	0,53	0,6
L_2 -Abstand	0,86	0,81	0,87	0,72	1,02	1,22	0,64	0,54	0,88	1,06
Hausdorff-Abstand	0,06	0,19	0,02	0,2	0,12	0,18	0,22	0,06	0,03	0,01
ESV (Betrag)	0,16	0,68	0	0,74	0,29	0,36	0,76	0	0	0
Tendenz	(+)	(+)	(o)	(+)	(+)	(+)	(+)	(o)	(o)	(o)
l^Y	0,57	0,72	0,84	0,67	0,76	0,92	0,74	0,57	1,83	1,08
$loss_g^Y$	0,54	0,50	0,82	0,45	0,90	0,60	0,53	0,44	0,94	0,88
Softloss	0,2	0,26	0,2	0,31	0,41	0,31	0,28	0,14	0,16	0,18

Tabelle 7.18: Mittelwert der Ergebnisse für die Spikes-Funktion, $n = 256$

Abbildung 7.20: Ergebnisse für die Spikes-Funktion, $n = 256$

	WH	WN6	UH	MDL	AWSk	AWSg	LPI	TS	STS	POT
L_1 -Abstand	3	8	6	4	7	10	2	1	5	9
L_2 -Abstand	5	4	6	3	8	10	2	1	7	9
Hausdorff-Abstand	4	8	2	9	6	7	10	5	3	1
ESV (Betrag)	5	8	1	9	6	7	10	1	1	1
Softloss	5	6	4	8	10	9	7	1	2	3

Tabelle 7.19: Ränge der mittleren Ergebnisse für die Spikes-Funktion, $n = 256$

- Die AWSg-Methode liefert überflüssige Extremstellen, vor allem in den Intervallen, in denen die Spikes-Funktion 0 ist. Außerdem zeigt die Methode große Abweichungen im Intervall $[0,2; 0,4]$. Alle Verlustfunktionen zeigen dies an, die Methode liefert die schlechtesten Ergebnisse.
- Die TS-Methode liefert den besten Graphen, was die meisten Verlustfunktionen wiedergeben. Nur der Hausdorff-Abstand zeigt einen leicht größeren Wert als bei anderen Methoden.
- Die Boxplots des Hausdorff-Abstands in Abbildung 7.20 zeigen, dass der Eindruck aus Abbildung 7.19 von wenig Unterschieden zwischen den Methoden bei der Spikes-Funktion nicht richtig ist. Die Methoden UH und POT liefern sehr geringe Werte, WN6, MDL, AWSg und LPI sehr große. Wie bereits bei anderen Funktionen zeigt auch hier die AWSk-Methode eine große Bandbreite an Ergebnissen, fast alle Werte im Bereich $[0; 0,25]$ kommen unter den 1000 Stichproben vor.
- Die Methoden UH, TS, STS und POT zeigen bezüglich der Extremwernerkenung immer so gute Ergebnisse wie in Abbildung 7.19. Die Methoden WN6, MDL und LPI liefern auch bei dieser Funktion zahlreiche überflüssige Extremstellen.
- Der ℓ^Y -Wert zeigt wenig Unterschiede zwischen den Methoden. Offensichtlich sind die Methoden bei der Schätzung der richtigen Höhe gleich gut. Nur STS und POT liefern größere Abstände.
- Bei Betrachtung der Mittelwerttabelle 7.18 fällt auf, dass die meisten Methoden einen L_1 -Abstand zwischen 0,5 und 0,6 liefern, nur TS ist mit einem Wert von 0,37 deutlich besser. Beim L_2 -Abstand zeigen sich mehr Abweichungen.
- Die Mittelwerte des Hausdorff-Abstands zeigen bei dieser Funktion mehr Unterschiede in der Güte der Methoden als bei anderen Funktionen.
- Die Ränge in Tabelle 7.19 zeigen wiederum, dass die verschiedenen Gütekriterien verschiedene Methoden als gut bewerten. Ähnliche Bewertungen werden nur durch Hausdorff-Abstand und ESV geliefert.

Zusammenfassung und Fazit

Insgesamt lassen die durchgeführten Simulationen folgende Schlussfolgerungen zu.

Der ESV beurteilt die Güte einer Schätzmethode in Bezug auf die Extremstellen besser als andere Maße, insbesondere besser als der Hausdorff-Abstand, der in vielen Fällen keine Güteunterschiede zwischen den Methoden feststellen kann. Methoden, für die bereits bekannt ist, dass sie häufig zu viele Extremstellen schätzen, werden durch den ESV als solche erkannt. Werden zu wenige Extremstellen geschätzt, zeigt der ESV dies durch das negative Vorzeichen an.

Die gewünschten Eigenschaften aus Bemerkung 3.1 in Kapitel 3 werden durch den ESV erfüllt. Insbesondere wird in der Praxis der Definitionsbereich ausgeschöpft und schlechte Schätzungen werden in sinnvollem Maße bestraft. Zudem wird das Verhältnis verschieden guter Schätzungen zueinander durch die zugehörigen ESV-Werte angemessen wiedergegeben. Mittels Betragbildung lassen sich die ESV-Ergebnisse über alle 1000 Simulationen sinnvoll zusammenfassen. Die angegebenen Tendenzen des Vorzeichens der ESV-Werte ermöglichen sogar zusätzliche Informationen über zu viele bzw. zu wenige geschätzte Extremwerte gemittelt über alle Simulationen. Dadurch bleibt das Kriterium Informationsverlust nicht nur für einzelne Stichproben sondern auch in der Zusammenfassung zumindest ansatzweise erfüllt.

Neben dem ESV bieten die Ergebnisse des Abstands ℓ^Y und des Glattheitsverlustes $loss_g^Y$ hilfreiche Ergänzungen. ℓ^Y ist ein geeignetes Kriterium, um die Güte der Schätzung in Y -Richtung zu beurteilen. $loss_g^Y$ zeigt an, wie stark eine Methode die Extremwerte der Signalfunktion überglättet.

Der Softloss differenziert weniger zwischen den Methoden als der ESV. Insgesamt zeigen die Ergebnisse dieses Verlustes kein eindeutiges Verhalten über alle Funktionen. Für einige Testbeds scheint der Softloss eine Art Kompromiss zwischen dem ESV und dem L_2 -Abstand oder ℓ^Y zu liefern. Bei anderen Funktionen beurteilt der Softloss die Methoden gänzlich anders als die anderen Verluste. Die Ergebnisse des Softloss sind also stark von der Testbed-Funktion abhängig, da durch die Reduzierung auf die Extremwerte die Unterschiedlichkeit der Funktionen auf Anzahl und Größe der Extremwerte reduziert wird. Zusätzlich führt die Faltung mit $\sigma = 1$ dazu, dass in den zu vergleichenden Funktionen Extremwerte mit wenig Masse entstehen. Diese fallen beim Vergleich

der Kurven mit der L_1 -Norm nicht so stark ins Gewicht.

Die Studie zeigt, dass die Methoden, die stückweise konstante Schätzungen liefern, häufig andere Ergebnisse erzielen, als die Methoden mit glatten Schätzungen. So sind erstere gemessen mit dem ESV den glatten Schätzungen für nahezu alle Testbed-Funktionen überlegen. Ausnahmen bilden bezüglich der Funktionen die Sine-Funktion und bezüglich der Methoden AWSk. Als einzige glatte Schätzung liefert STS ähnlich gute Ergebnisse wie WH, UH, TS und POT. Das liegt daran, dass STS das Monotonieverhalten einer stückweise konstanten Schätzung übernimmt. Weiterhin ist auffällig, dass die stückweise konstanten Schätzungen die Signalfunktionen häufig überglätten. Eine Ausnahme bildet die TS-Methode. Von den Methoden, die glatte Schätzungen liefern, zeigt die STS-Methode häufig überglättete Ergebnisse.

ZUSAMMENFASSUNG UND AUSBLICK

Die Beurteilung der Güte von nichtparametrischen Regressionsmethoden erfolgt über verschiedene Kriterien. Häufig finden L_p -Normen Verwendung, obwohl diese nur den vertikalen Abstand zwischen einer Testbed- und einer Schätzfunktion zur Beurteilung heranziehen. Je nach Ziel der Regression kann dies zu Fehleinschätzungen bezüglich der Qualifikation von Regressionsmethoden führen. Beispielsweise sind diese Normen nicht geeignet, wenn die Extremwerte der zugrunde liegenden Funktion von Bedeutung sind. In dieser Arbeit wird ein neues Gütekriterium hergeleitet, das in der Lage ist, eine Schätzfunktion in Bezug auf ihre Extremstellenerkennung zu beurteilen.

Kapitel 2 dient der Erläuterung des zugrunde liegenden Modells der nichtparametrischen Regression. Jeweils zehn verschiedene verwendete Testbed-Funktionen und Regressionsmethoden werden vorgestellt sowie übliche Gütekriterien genannt.

In Kapitel 3 erfolgt die Herleitung des Extremstellenverlustes. Zunächst werden Anforderungen an eine Verlustfunktion, die das Schätzverhalten bezüglich der Extremwerte beurteilt, aufgestellt. Darunter fallen Forderungen wie Normiertheit, keine Definitionslücken und wenig Informationsverlust dahingehend, dass erkennbar sein soll, ob der Verlust durch zu viele oder zu wenige Extremstellen zustande kommt. Als weitere gewünschte Eigenschaft gilt die Ausschöpfung des Definitionsbereichs. Dazu gehört auch eine sinnvolle Bestrafung und ein geeignetes Verhältnis von verschiedenen Möglichkeiten des Fehlverhaltens einer Schätzfunktion. Eine letzte Anforderung ist die Möglichkeit einer sinnvollen Zusammenfassung der Ergebnisse über mehrere Simulationsläufe.

Für die Herleitung der Verlustfunktion wird zunächst eine Extremstelle der Testbed-Funktion als identifiziert definiert, wenn in einem Toleranzintervall um sie herum eine

geschätzte Extremstelle liegt, wobei zwischen Minima und Maxima unterschieden wird. Daraufhin können Fehler bezüglich *zu vieler* und *zu weniger* Extremstellen in der Schätzfunktion beschrieben werden. Die Zusammenfassung dieser Fehler in vier verschiedene Verlustfunktionen wird ausführlich diskutiert. Schließlich wird einer dieser Verluste, der die genannten Fehler mit der Anzahl wahrer und geschätzter Extremstellen normiert, ausgewählt und unter Berücksichtigung von Sonderfällen als Extremstellenverlust definiert. Dieser neue Verlust erfüllt alle Anforderungen.

In Kapitel 4 wird der Einfluss der verwendeten Toleranzen auf den Extremstellenverlust untersucht. Dabei stellt sich heraus, dass verschiedene Toleranzen nur einen geringen Einfluss auf das Ergebnis des Extremstellenverlustes aufweisen. Gleichzeitig macht die Möglichkeit, verschiedene Toleranzen zu verwenden, den Extremstellenverlust flexibel. Je nach Anwendung können die Toleranzen größer oder kleiner gewählt werden. Entsprechendes gilt für die Rechengenauigkeit, mit der die Extremstellen von Signal- und Schätzfunktion bestimmt werden. Damit lässt sich beispielsweise einstellen, wenn Extremwerte erst ab einer bestimmten Größe auch als solche in die Berechnung des Verlustes eingehen sollen.

Da auch die richtige Höhe von Extremwerten in der nichtparametrischen Regression von Interesse ist, wird in Kapitel 5 ein Extremwertverlust in Y -Richtung diskutiert. Eine Definition wie die des Extremstellenverlustes stellte sich als nicht praktikabel heraus, da dazu die Höhen der Extremwerte mittels einer Toleranz in *akzeptabel oder nicht* einzuteilen sind. Die dabei verwendete Toleranz hat im Gegensatz zu der für Extremstellen einen derart großen Einfluss auf das Ergebnis, dass es nicht sinnvoll erscheint, einen solchen Verlust zu verwenden. Stattdessen wird ein Abstandsmaß ℓ^Y eingeführt. Dieses gibt Aufschluss über die Genauigkeit, mit der eine Schätzfunktion die Höhe der wahren Extremwerte bestimmt. Zusätzlich wird ein Glattheitsverlust vorgeschlagen, der anzeigt, ob eine Methode die Extremwerte einer Testbed-Funktion zu stark glättet, also Minima zu groß und Maxima zu klein schätzt.

In Kapitel 6 wird eine gänzlich andere Herangehensweise an die Problematik der Beurteilung der Güte bezüglich Extremstellen und -werten aufgezeigt. Dafür sollen Testbed- und Schätzfunktion mittels einer L_p -Norm verglichen werden, aber dennoch die Extremwerte als Grundlage dienen. Das wird erreicht, indem zunächst beide Funktionen auf ihre jeweiligen Extremwerte reduziert und die entstandenen Kurven mit einer

Gausskurve gefaltet werden. Die gefalteten Kurven können mit einer L_1 -Norm verglichen werden. Der auf diese Weise definierte Verlust trägt die Bezeichnung Softloss. Er hängt von der Standardabweichung der verwendeten Gausskurve ab. Da für Signal- und Schätzfunktion dieselbe Standardabweichung verwendet wird, ist der Einfluss jedoch verhältnismäßig gering.

In Kapitel 7 werden schließlich die vorgestellten Verlustfunktionen in einer großen Simulationsstudie angewendet. Zudem findet ein Vergleich mit bestehenden Gütekriterien wie der L_1 - und L_2 -Norm und dem Hausdorff-Abstand statt. Dabei zeigt sich, dass der Extremstellenverlust als einziger Verlust in der Lage ist, die Güte der Schätzmethode bezüglich der Extremstellen zu beurteilen. Der Abstand ℓ^Y dient als hilfreiche Ergänzung zum Extremstellenverlust.

Obwohl der Softloss so konstruiert ist, dass er Extremstellen und -werte gleichzeitig erfasst, ist er nicht bei allen Testbed-Funktionen in der Lage, die Ergebnisse des Extremstellenverlustes und des Abstands ℓ^Y gemeinsam wiederzugeben. Da er ein klassisches Gütekriterium mit der Extremwernerkenntnis verknüpft, liefert der Softloss in einigen Simulationssituationen einen Kompromiss zwischen einer Gütebeurteilung durch die L_1 -Norm und den Extremstellenverlust. Weitere Forschung an einem ähnlich aufgebauten Verlust könnte hier Abhilfe schaffen.

Der Extremstellenverlust lässt sich leicht auf andere interessierende Merkmale von Regressionsfunktionen übertragen. Statt der Extremstellen können beispielsweise Sprungstellen oder Wendepunkte in die Formel eingesetzt werden. In der Dichteschätzung sind häufig die Modalwerte einer gesuchten Dichte von Interesse. Hier kann ein ähnlicher Verlust Verwendung finden, der sich auf Modalwerte bezieht.

Zudem ist der Extremstellenverlust auch geeignet, wenn ein abweichendes nichtparametrisches Regressionsmodell, etwa mit nichtäquidistanten Designpunkten oder mit verschiedenen Rauschverteilungen zugrunde liegt.

Ein interessanter Gegenstand weiterer Forschung wäre die Übertragung des Extremstellenverlustes ins Mehrdimensionale. Dafür muss zunächst eine geeignete Definition von mehrdimensionalen Extremwerten gefunden werden. Einerseits kann bereits ein Funktionswert als Extremwert definiert werden, der nur in eine Richtung größer oder kleiner als die umliegenden Werte ist. Andererseits kann auch gefordert werden, dass dies für einen Extremwert in alle Richtungen zutrifft. Zudem muss die Toleranz in einen geeig-

neten Begriff der Umgebung eines mehrdimensionalen Extremwerts überführt werden. Die Untersuchungen in dieser Arbeit haben gezeigt, dass das neue Kriterium *Extremstellenverlust* den Anforderungen der Extremwerterkennung gerecht wird. In allgemeinen Simulationsstudien zur nichtparametrischen Regression, in denen kein explizites Ziel angegeben wird, sollte neben klassischen Gütekriterien der Extremstellenverlust nicht fehlen.

ABKÜRZUNGSVERZEICHNIS

NICHTPARAMETRISCHE REGRESSIONSMETHODEN

AWSg	Adaptive Weights Smoothing, glatt
AWSk	Adaptive Weights Smoothing, stückweise konstant
LPI	Lokaler Plug-in Kernschätzer
MDL	Minimum Description Length
POT	Potts-Schätzer
STS	Glattes Taut String
TS	Taut String
UH	Unbalanced Haar
WH	Wavelet-Methode mit Haar-Wavelet
WN6	Wavelet-Methode mit Daubechies Extremal Phase Family $N = 6$

GÜTEKRITERIEN

ESV	Extremstellenverlust
HD	Hausdorff-Abstand
ℓ^Y	Verlust für den Abstand in Y -Richtung
$loss_g^Y$	Glattheitsverlust
L_p	L_p -Norm, für $p = 1, 2$
SoLo	Softloss

LITERATURVERZEICHNIS

Bernholt, T. und Hofmeister, T., 2006, An Algorithm for a Generalized Maximum Subsequence Problem. In LATIN 2006: Theoretical Informatics (Hrsg.: Correa, J., Hevia, A. und Kiwi, M.). *Lecture Notes in Comput. Sci.*, 3887, S. 178 – 189, Springer, Berlin.

Bernholt, T., Eisenbrand, F. und Hofmeister, T., 2007, A geometric framework for solving subsequence problems in computational biology efficiently, *Proceedings of the 23rd Annual ACM Symposium on Computational Geometry (SoCG 2007)*, S. 310 – 318.

Boysen, L., Kempe, A., Liebscher, V., Munk, A. und Wittich, O., 2009, Consistencies and Rates of Convergence of Jump-Penalized Least Squares Estimators, *The Annals of Statistics*, 37, S. 157 – 183.

Brockmann, M., Gasser, T. und Herrmann, E., 1993, Locally Adaptive Bandwidth Choice for Kernel Regression Estimators, *Journal of the American Statistical Association*, 88, S. 1302 – 1309.

Barner, M. und Flohr, F., 1991, Analysis I, 4. Auflage, de Gruyter, Berlin.

Cai, T., 1999, Adaptive Wavelet Estimation: a Block Thresholding and Oracle Inequality Approach, *The Annals of Statistics*, 27, S. 898 – 924.

Chaudhuri, P. und Marron, J. S., 2000, Scale Space View of Curve Estimation, *The Annals of Statistics*, 28, S. 408 – 428.

Daubechies, T., 1992, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

- Davies, P. L., Gather, U., Nordman, D. J. und Weinert, H., 2009, A Comparison of Automatic Histogram Constructions, *ESAIM: Probability and Statistics*, 13, S. 181 – 196.
- Davies, P. L., Gather, U. und Weinert, H., 2008, Nonparametric Regression as an Example of Model Choice, *Communications in Statistics - Simulation and Computation*, 37, S. 274 – 289.
- Davies, P. L. und Kovac, A., 2001, Local Extremes, Runs, Strings and Multiresolution, *The Annals of Statistics*, 29, S. 1 – 65.
- Davies, P. L. und Kovac, A. <A.Kovac@bristol.ac.uk>, 2008, ftnonpar: Features and Strings for Nonparametric Regression. R package version 0.1-83, <http://www.maths.bris.ac.uk/~maxak/ftnonpar.html>, <http://cran.r-project.org/web/packages/ftnonpar/index.html>.
- Donoho, D. L., 1988, One-sided Inference about Functionals of a Density, *The Annals of Statistics*, 16, S. 1390 – 1420.
- Donoho, D. L. und Johnstone, I. M., 1994, Ideal Spatial Adaptation by Wavelet Shrinkage, *Biometrika*, 81, S. 425 – 455.
- Fryzlewicz, P., 2006, unbalhaar: Function estimation via Unbalanced Haar wavelets. R package version 1.0., <http://cran.r-project.org/web/packages/unbalhaar/index.html>.
- Fryzlewicz, P., 2007, Unbalanced Haar Technique for Nonparametric Function Estimation, *Journal of the American Statistical Association*, 102, S. 1318 – 1327.
- Gasser, T., Kneip, A. und Köhler, W., 1991, A Flexible and Fast Method for Automatic Smoothing, *Journal of the American Statistical Association*, 86, S. 643 – 652.
- Grünwald, P., 2000, Model Selection based on Minimum Description Length, *Journal of Mathematical Psychology*, 44, S. 133 – 152.
- Härdle, W., 1990, Applied Nonparametric Regression, Cambridge University Press, Cambridge.

- Härdle, W., Müller, M., Werwatz, A. und Sperlich, S., 2004, Nonparametric and Semiparametric Models, Springer Series in Statistics, Springer-Verlag Berlin, Heidelberg, New York.
- Herrmann, E., 1997a, Local Bandwidth Choice in Kernel Regression Estimation, *Journal of Computational and Graphical Statistics*, 6, S. 35 – 54.
- Herrmann, E., 1997b, Kernel regression fitting with local and global bandwidth optimization, <http://www.biostat.uzh.ch/research/software/kernel.html>.
- Huttenlocher, D. P., Klanderman, G. A. und Rucklidge, W. J., 1993, Comparing Images Using the Hausdorff Distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, 850 – 863.
- Kovac, A., 2007, Smooth Functions and Local Extreme Values, *Computational Statistics & Data Analysis*, 51, S. 5155 – 5171.
- Kuhls, S., 2008, Planung, Durchführung und Analyse einer klinischen Studie zur Bewertung und zum Vergleich von Alarmsystemen in der Intensivmedizin, Dissertation, Technische Universität Dortmund.
- Liebscher, V. und Winkler, G., 1999, A Potts Model for Segmentation and Jump-Detection, *Proceedings of the SSSSG Conference*, Prag.
- Liebscher, V., 2003, tssegmentation: Potts smoothers and related stuff, R package version: 0.1–2, <http://www.math-inf.uni-greifswald.de/biomathematik/liebscher/misc/>.
- Marron, J. S., Adak, S., Johnstone, I. M., Neumann, M. H. und Patil, P., 1998, Exact Risk Analysis of Wavelet Regression, *Journal of Computational and Graphical Statistics*, 7, S. 278 – 309.
- Marron, J. S. und Tsybakov, A. B., 1995, Visual Error Criteria for Qualitative Smoothing, *Journal of the American Statistical Association*, 90, S. 499 – 507.
- Mildenberger, T., 2008, A Geometric Interpretation of the Multiresolution Criterion in Nonparametric Regression, *Journal of Nonparametric Statistics*, 20, S. 599 – 609.

- Nason, G. P. und Silverman, B. W., 1994, The Discrete Wavelet Transform in S, *Journal of Computational and Graphical Statistics*, 3, S. 163 – 191.
- Nason, G. P., Kovac, A. und Maechler, M., 2008, wavethresh: Software to perform wavelet statistics and transforms, R package version 2.2–10, <http://cran.r-project.org/web/packages/wavethresh/index.html>.
- Polzehl, J. und Spokoiny, V. G., 2000, Adaptive Weights Smoothing with Applications to Image Restoration, *Journal of the Royal Statistical Society, Series B*, 62, S. 335 – 354.
- Polzehl, J. und Spokoiny, V. G., 2003, Varying Coefficient Regression Modeling, Weierstrass-Institut für Angewandte Analysis und Stochastik, Preprint, ISSN 0946–8633.
- Polzehl, J. und Spokoiny, V. G., 2006, Local Likelihood Modeling by Adaptive Weights Smoothing, *Probab. Theory Related Fields*, 135, S. 335 – 362.
- Polzehl, J., 2007, aws: Adaptive Weights Smoothing. R package version 1.3–3.1., <http://wias-berlin.de/project-areas/stat/projects/adaptive-image-processing.html>, <http://cran.r-project.org/web/packages/aws/index.html>.
- Rissanen, J., 1987, Stochastic Complexity, *Journal of the Royal Statistical Society, Series B*, 49, S. 223 – 239.
- Rissanen, J., 1996, Fisher Information and Stochastic Complexity, *IEEE Trans. Information Theory*, 42, S. 40 – 47.
- Rissanen, J., 2000, MDL-Denoising, *IEEE Trans. Information Theory*, 46, S. 2537 – 2543.
- Wassermann, L., 2006, All of Nonparametric Statistics, Springer, New York.
- Winkler, G. und Liebscher, V., 2002, Smoothers for discontinuous signals, *Nonparametric Statistics*, 14, 203 – 222.
- Winkler, G., Kempe, A., Liebscher, V. und Wittich, O., 2003, Parsimonious Segmentation of Time Series' by Potts Models. Sonderforschungsbereich 386, Discussion Paper 348.

Winkler, G., Wittich, O., Liebscher, V. und Kempe, A., 2004, Don't Shed Tears over Breaks, Preprint, Schriftenreihe des IBB 04-08.

Wittich, O., Kempe, A., Winkler, G. und Liebscher, V., 2008, Complexity Penalized Least Squares Estimators: Analytical Results, *Mathematische Nachrichten* 281, S. 582 – 595.

Zhou, X.-H., Obuchowski, N. A. und McClish, D. K., 2002, Statistical Methods in Diagnostic Medicine, Wiley-Interscience, John Wiley & Sons, Inc., New York.