

Informationsverdichtung durch den gleitenden Mittelwert

MICHAEL HIMMELSBACH, HELGA M. BÖHM, LINZ

Zusammenfassung: Mit Hilfe von Mittelwerten kann man Datensätze rasch interpretieren, vergleichen und zusammenfassen. Wir diskutieren hier den gleitenden arithmetischen Mittelwert: Er hängt stark vom Mittelungsintervall und -zeitpunkt ab. An Hand von Beispielen verdeutlichen wir die Auswirkungen dieser Parameter; wobei eine Langzeituntersuchung von Temperaturdaten und eine Darstellung von Börsenkursen näher betrachtet werden. Anschließend erweitern wir den diskreten Ansatz auf funktionale Zusammenhänge. Durch ein Beispiel aus der Physik werden die Auswirkungen des gleitenden Mittels dargestellt.

1 Einleitung

Mittelwerte haben im Mathematikunterricht einen hohen Stellenwert: Ihre Berechnung und Interpretation ist im österreichischen Kompetenzmodell der zentralen Reifeprüfung explizit in zwei Grundkompetenzen formuliert (BIFIE, 2013). Mittelwerte werden schon früh im Unterricht thematisiert – arithmetisches Mittel, Median und Modus sind zusammen mit den Streumaßen bereits im Lehrplan der 8. Schulstufe enthalten (https://www.bmbf.gv.at/schulen/unterricht/lp/lp_ahs_unterstufe.html).

Schülerinnen und Schüler entwickeln sehr rasch eine Vorstellung vom Durchschnitt einfacher Daten. Viele können bereits zu Beginn der Sekundarstufe I das arithmetische Mittel gemäß dem Berechnungsalgorithmus bestimmen. Bei der Ermittlung des Notenschnitts werden Lernende auch mit gewichteten Mittelwerten konfrontiert – allerdings fällt eine Aufgabenvariation den Schülerinnen und Schülern häufig schwer, da ein inhaltliches Verständnis diverser Mittelwerte nicht vorhanden ist. Bei der Behandlung in der Schule muss der Sinn einer Mittelwertbildung thematisiert und der Interpretationskompetenz im Vergleich zu operativen Kompetenzen einen größeren Stellenwert als bisher im Lernprozess eingeräumt werden (Russell & Mokros, 1995).

In dieser Arbeit diskutieren wir ausschließlich eine Erweiterung des gängigen arithmetischen Mittelwertbegriffes: den „gleitenden Mittelwert“. In Kapitel 2 wird das arithmetische Mittel gleich großer Teile eines Datensatzes berechnet: Die Teilmengen werden also abschnittsweise in ihrem Mittelwert – einem Blockmittel – abgebildet. Anschließend wird dieser Block „gleitend“ über den Datensatz verschoben:

Der Reihe nach wird so jeder Datenpunkt über eine feste Zahl von Nachbarwerten gemittelt. Die Ausgangsgrößen müssen demnach als metrische Größe (entweder als diskrete Werte oder als stetige Funktion) vorliegen. Spätestens an dieser Stelle wird klar, dass sich eine sinnvolle Datenanalyse nicht nur auf die algorithmische Berechnung diverser Kenngrößen bzw. Werte reduziert. Essentiell ist neben der Bedeutung der Datenwerte auch die sinnstiftende Deutung der verdichteten Beschreibung (Konold & Pollatsek, 2002).

Nach Diskussion der mathematischen Grundlagen in Kapitel 2, insbesondere der zentralen Frage nach Dauer und Zeitpunkt der Mittelung, wenden wir das Verfahren in Kapitel 3 auf unterschiedliche alltägliche Fragestellungen an. Dies zeigt, dass diese Datenanalyse sehr gut fächerübergreifend unterrichtet werden kann. Zudem können alle Repräsentationsstufen des Lernens nach Bruner (1976) abgedeckt werden. Die hier diskutierte Thematik ist also gut für projekt- bzw. problemzentrierten Unterricht geeignet.

Um unsere algorithmische Betrachtung inhaltlich zu verdeutlichen und die Begriffsbildung zu erleichtern, greifen wir in den theoretischen Abschnitten folgendes wiederkehrendes Beispiel auf: Ein sportbegeistertes Kind führt am Ende eines Trainings immer den gleichen, validen Test durch (z. B. sieben Würfe auf einen Basketballkorb) und notiert die Trefferanzahl. Jede Woche finden drei Trainings statt. Anhand der Treffer möchte das Kind seinen Trainingserfolg feststellen.

2 Mathematische Betrachtung

Die Berechnung des arithmetischen Mittels n unabhängiger, identisch-verteilter Zufallsgrößen x_1, \dots, x_n

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \quad (1)$$

stellt die meisten Lernenden vor keine hohe Herausforderung (ausgenommen im Fall großer Datenmengen). Auch der Durchschnittswert kontinuierlicher Funktionen ist graphisch intuitiv verständlich, selbst vor Kenntnis der Integralrechnung am Ende der Sekundarstufe II.

Bei vielen Datensätzen ist jedoch die Abbildung in einen einzigen Wert \bar{x} wie in (1) zu wenig aussagekräftig – mehrere Teilmittelungen über jeweils fixe

Untergruppen der Werte sind zielführender. Die mittlere Trefferanzahl aller Trainingseinheiten des Schuljahres erlaubt keinen Aufschluss über die Leistungsentwicklung.

Blockmittelung diskreter Daten

Insbesondere bei einem Datensatz mit n lokal stark fluktuierenden Werten ist eine Mittelung über eine feste Zahl von $m < n$ Datenpunkten sinnvoll. Man erzielt dadurch eine Darstellung, die die Interpretation eines Trends in den Daten wesentlich erleichtert. Es werden also jeweils m Werte zu Blöcken zusammengefasst und gemittelt:

$$\left(\underbrace{x_1, \dots, x_m}_{\bar{x}_1}, \underbrace{x_{m+1}, \dots, x_{2m}}_{\bar{x}_2}, \dots, \underbrace{\dots, x_n}_{\bar{x}_{\lfloor n/m \rfloor}} \right) \quad (2)$$

So erhält man $\lfloor n/m \rfloor$ Mittelwerte (die Klammer $\lfloor \dots \rfloor$ bezeichnet die nächst kleinere natürliche Zahl des Bruchs n/m). Wie erwartet, weisen diese \bar{x}_k eine glattere Struktur als die Originaldaten auf: Die waagrechten Linien in Abb. 1 (a) und (c) filtern praktisch alle Fluktuationen heraus; in den rechten Abbildungen (b) und (d) geben sie eine detailreichere Information, aber ohne die extremen Spitzen. Diese Einteilung der Daten in feste Glättungsintervalle („Bins“) und die Mittelwertbildung in diesem Intervall führen zum sogenannten Regressogramm (Engel, 2009).

Neben der Intervall- bzw. Blockgröße m kann auch der Startpunkt x_k der Mittelung frei gewählt werden, da diese Blockmittelung nicht beim ersten Datenwert x_1 beginnen muss. In Abb. 1 wurden 23 Datenpunkte (durch die graue Linie interpoliert) über $m = 7$ bzw. $m = 3$ Punkte (schwarze Graphen) gemittelt und in einem Regressogramm dargestellt. Startet man dabei bei x_1 („Vorwärtsmittelung“), werden die letzten beiden Werte nicht mehr abgebildet, wählt man hingegen die Blöcke ausgehend von x_{23} („Rückwärtsmittelung“), verliert man analog die ersten beiden Datenpunkte. Der Einfluss des Anfangspunkts x_k auf die Ergebnisse \bar{x}_k ist in Abb. 1 deutlich sichtbar: Das Abfallen der Kurve zwischen x_{15} und x_{16} ist (bei gleicher Blockgröße) in (b) stärker ausgeprägt als in (d).

Zusammenfassend gilt also: i) die Blockmittelwerte glätten starke Schwankungen, ii) je größer die Blöcke umso weniger Eigenschaften der ursprünglichen Daten werden abgebildet und iii) je kleiner die Blöcke umso stärker ist der Einfluss des Startpunkts auf die Ergebnisse.

In unserem eingangs erwähnten Beispiel bedeutet also eine Blockmittelung über sieben Würfe eine Zusammenfassung jedes einzelnen Trainingstests. Die-

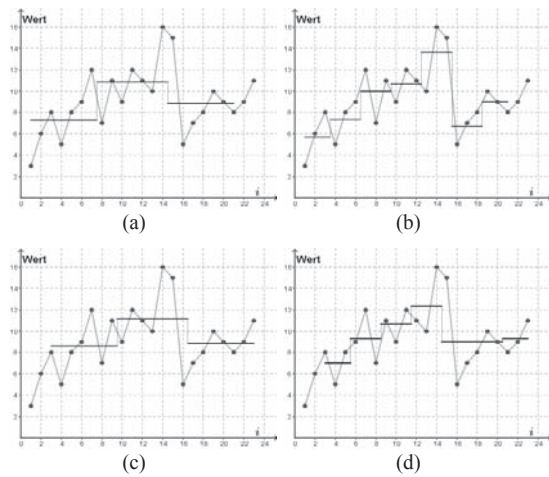


Abb. 1: Mittelung (in schwarz) einer Datenreihe (graue Interpolation) von 23 Punkten: links jeweils 3 Blöcke mit $m = 7$ Werten; rechts jeweils 7 Blöcke mit $m = 3$ Werte (schwarz). Oben: Beginn der Mittelung jeweils beim ersten Datenwert $k = 1$. Unten: Ende der Mittelung beim letzten Datenwert (entspricht dem Startwert $k = 3$).

se Ergebnisse lassen sich wiederum gut miteinander vergleichen, um einen Trainingstrend zu finden.

Gleitender Mittelwert diskreter Daten

Eine Glättung der Rohdaten ohne Auswirkung der willkürlichen Wahl des Startpunkts wird erreicht, indem man bei der blockweisen Mittelung zu einem gleitenden Startwert übergeht. Dies bedeutet, dass nun jeder Wert x_i mit einer fixen Zahl m seiner Nachbarwerte in einen Blockmittelwert abgebildet wird:

$$\bar{x} := \bar{x}_{i,m} := \frac{1}{m} \sum_{j=0}^{m-1} x_{i+j} \quad (3)$$

also konkret in obigen Beispiel

$$\bar{x}_{i,m} = \begin{cases} \bar{x}_{i,3} = \frac{1}{3}(x_i + x_{i+1} + x_{i+2}) & m = 3 \\ \bar{x}_{i,7} = \frac{1}{7}(x_i + \dots + x_{i+6}) & m = 7 \end{cases}$$

Somit werden fast alle n Daten (ausgenommen die Randpunkte) in Mittelwerte abgebildet; dieser „gleitet“ über die Datenwerte und liefert eine glattere Darstellung der Urdaten (siehe schwarze Linien in Abb. 2). Die Wahl des Startpunktes spielt nun keine Rolle mehr. Jedoch stellt sich in Gleichung (3) wiederum die Frage, ob x_i besser mit den folgenden oder den vorangegangenen m Werten, also ob vorwärts oder rückwärts gemittelt wird. Ebenso bietet sich eine symmetrische Wahl an (z. B.: $\bar{x}_i = (x_{i-1} + x_i + x_{i+1})/3$).

In Abb. 2 ist der gleiche Datensatz wie in Abb. 1 dargestellt, wobei bei der gleitenden Mittelung gleiche Blockgrößen wie in Abb. 1 (in (a) $m = 7$ Datenwer-

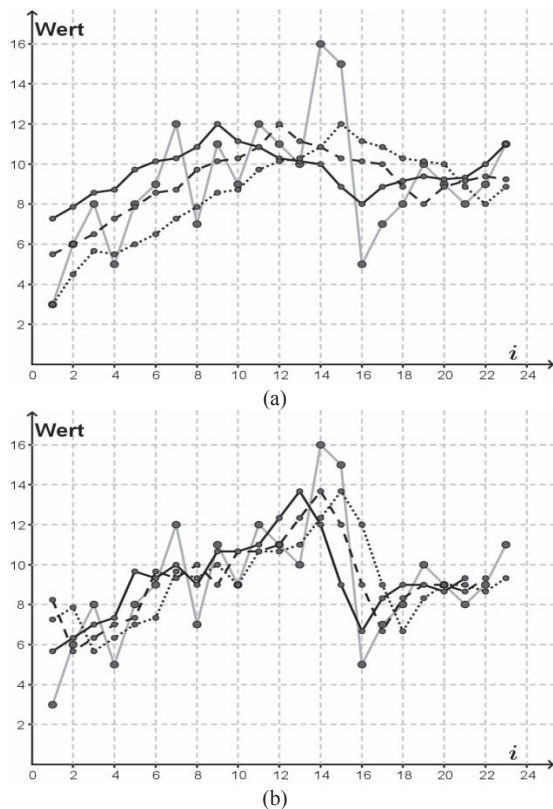


Abb. 2: Interpolierter Datensatz (grau) wie in Abb. 1. In (a) ist der schwarze Graph über $m = 7$ Werte vorwärts (durchgehende Linie), symmetrisch (strichliert) und rückwärts (punktiert) gleitend gemittelt; in (b) analoge Betrachtung mit $m = 3$ Mittelungspunkten.

te, in (b) $m = 3$ Datenwerte) verwendet werden. Die Mittelung erfolgt nun gleitend: Die Punkte, die durch die durchgezogene Linie interpoliert werden, sind mit Gleichung (3) generiert (vorwärts), die strichlierte Linie mit einer symmetrischen Mittelung und bei der gepunkteten Linie wird rückwärts gemittelt.

Man erkennt, dass sich die drei Fälle nur durch eine horizontale Verschiebung unterscheiden. Betrachtet man wieder den „Einbruch“ der Kurve bei x_{15} , so hinkt die Rückwärtsmittelung den Rohdaten nach, die Vorwärtsmittelung nimmt hingegen den Einbruch vorweg. Solche Veränderungen sind umso größer, je breiter die Blöcke sind. In (a) erkennt man bei der Betrachtung der Datenwerte $i = 9, \dots, 13$ bei der Vorwärtsmittelung (durchgehende Linie) den Abfall der Werte, die rückwärts gemittelte Linie (strichpunktiert) steigt in diesem Bereich allerdings!

Für die Trainingsauswertung des sportbegeisterten Kindes ergibt der gleitende Mittelwert eine glattere Kurve, die den Leistungstrend (also die Veränderung der Trefferanzahl mit zunehmenden Training) besser anzeigt. Fluktuationen wie Tagesleistungen und periodische Schwankungen werden durch eine sinnvolle Wahl der Mittelungslänge geglättet.

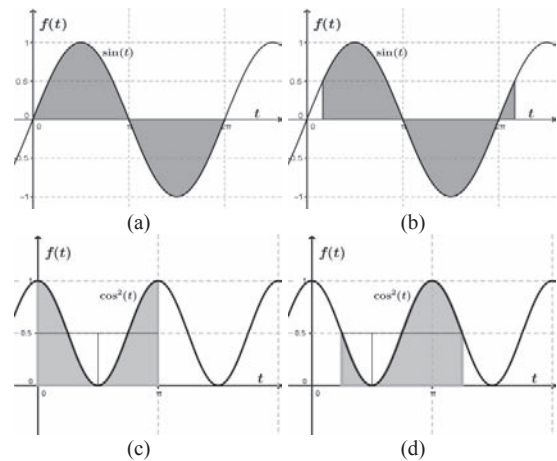


Abb. 3: oben: Mittelung einer Sinusfunktion (a) im Intervall $[0; 2\pi]$ und (b) im Intervall $[\frac{\pi}{6}; 2\pi + \frac{\pi}{6}]$; unten: Mittelung der Cosinus-Quadratfunktion über eine Periode: (c) im Intervall $[0; \pi]$ und (d) in $[\frac{\pi}{4}; \pi + \frac{\pi}{4}]$.

Bevor stetige Funktionen betrachtet werden, führen wir zwei Anmerkungen an: Erstens werden viele Datensätze zeitlich sequentiell erhoben und können daher grundsätzlich nur rückwärts gemittelt werden (außer a posteriori). Zweitens kann die hier präsentierte Betrachtung des arithmetischen Mittelwerts für weitere statistische Kenngrößen, wie den Median und die Quantile (bzw. die Zusammenfassung in Boxplots) erweitert werden (Engel, 2010). In dieser Arbeit beschränken wir uns aber auf arithmetische Mittelwerte.

Stetige Funktionen/Daten

Der Übergang von diskreten Datenwerten zur stetigen Verteilungsfunktion ist stringent. Wird eine Funktion $f(t)$ zum Zeitpunkt t_0 über das Intervall $[t_0 - a; t_0 + b]$ gemittelt, so geht die Summe der Datenpunkte in ein Integral über. (Der Blockgröße m entspricht nun die Intervalllänge $T = a + b$.) Analog zum diskreten Fall kann der Mittelwert in t_0 vorwärts ($a = 0$), symmetrisch ($a = b = T/2$) oder rückwärts ($b = 0$) gebildet werden. Da sich diese Fälle nur durch eine Verschiebung unterscheiden (vgl. Abb. 2), beschränken wir uns hier auf die symmetrische Wahl. Der Mittelwert einer Funktion f wird also durch die Größen t_0 und T charakterisiert:

$$\bar{f} := \bar{f}(t_0, T) = \frac{1}{T} \int_{t_0 - \frac{T}{2}}^{t_0 + \frac{T}{2}} f(t) dt. \quad (4)$$

Die Wahl der Variablen t bzw. T ist beabsichtigt, denn häufig werden zeitabhängige Problemstellungen betrachtet. Ähnlich zum diskreten Fall zeigen die zu untersuchenden Funktionen f meist Überlagerungen folgender qualitativer Eigenschaften: „Trends“ (im

Beispiel etwa eine Leistungsänderung mit zunehmendem Training), „Rauschen“ (nie vermeidbare, kaum erklärbare Schwankungen), „Ereignisse“ (z. B. Schularbeiten direkt nach Trainingstagen) und „Oszillationen“ (periodische Effekte wie wöchentlich regelmäßig schlechtere Performance z. B. an Montagen).

Um den Aspekt periodischer Effekte zu untersuchen, betrachten wir dies an Hand von einfachen Funktionen genauer: Als Unterrichtseinstieg eignen sich die Sinus- oder die quadrierte Cosinusfunktion. Für sie lässt sich das Integral (4) über eine Periode analytisch einfach berechnen. Die Aufgabenstellung ist aber auch ohne Integralkenntnisse graphisch durch den Vergleich von Flächeninhalten lösbar und kann somit vor Kenntnis der Integralrechnung durchgeführt werden.

In Abb. 3 (a) ist eine Periode der Sinus-Funktion und ihr Mittelungsintegral dargestellt. Da sich die Flächeninhalte im Intervall $[0; 2\pi]$ aufheben, verschwindet der Mittelwert von $f(t) = \sin(t)$ über eine volle Periode: $\bar{f}(0, 2\pi) = \bar{f} = 0$. In Abb. 3 (b) ist ersichtlich, dass die Anfangsstelle t_0 keine Auswirkung auf das Ergebnis der Mittelung hat. In Abb. 3 (c) ist die Funktion $f(t) = \cos^2(t)$ (Periodenlänge $T = \pi$) gezeigt, die positiv definit ist. Durch einen Vergleich der Flächen erhält man den Mittelwert $\bar{f} = 1/2$, wiederum unabhängig von t_0 (siehe 3 (d)). Wird über die Periodenlänge gemittelt, kann also das Argument t_0 weggelassen werden.

Somit drängt sich die Frage nach Mittelungsintervallen auf, die kein natürliches Vielfaches der Periodendauer sind. Generell ist dabei festzuhalten, dass eine Vergrößerung der Mittelungslänge tendenziell zu immer glatteren Kurven führt. (Engel, 2009).

In Abb. 4 (a) und (b) sind wieder die Sinus- und die quadrierte Cosinusfunktion (jeweils in grau dargestellt); sie werden nun *gleitend* über jeweils unterschiedliche Fensterbreiten gemittelt (strichlierte Kurven), wobei diese von einem kleinen $T = \pi/4$ bis hin zu fast einer gesamten Periode dauernden Wert reichen.

Zwei Auswirkungen der unterschiedlichen T -Werte sind deutlich erkennbar: Erstens hängen die gemittelten schwarz-strichlierten Kurven nun vom Zeitpunkt t_0 der Mittelung ab. Zweitens nähert sich die strichlierte Kurve mit der Fensterbreite $T \rightarrow$ Periode immer mehr dem in Abb. 3 dargestellten Mittelwert. (Mittelungslängen größer als eine Periodendauer können außerdem zu Phasenverschiebungen führen, wie in Abb. 4 bei (a) mit $T = 5\pi/2$ und (b) $T = 5\pi/4$ ersichtlich.)

Eine hier nützliche Metapher sieht Daten als ein von „Rauschen“ überlagertes „Signal“ an (Konold & Pollatsek, 2002). Mit Hilfe des gleitenden Mittelwerts kann das überlagernde Rauschen weggeglättet und das Signal – der Trend – herausgefiltert werden. Dabei sind die Wahl von Mittelungslänge und Mittelungszeitpunkt essentiell: Ein zu kleines Fenster interpoliert lediglich den Datensatz und sein Rauschen (siehe Abb. 4 bei $T = \pi/4$). Hingegen mittelt ein zu großes T die Konturen, die ja extrahiert werden sollen, zur Gänze weg (in Abb. 4 a strichpunktiierte Linie mit $T = 19\pi/10$). Eine derartige Datenreduktion schießt also über das Ziel hinaus.

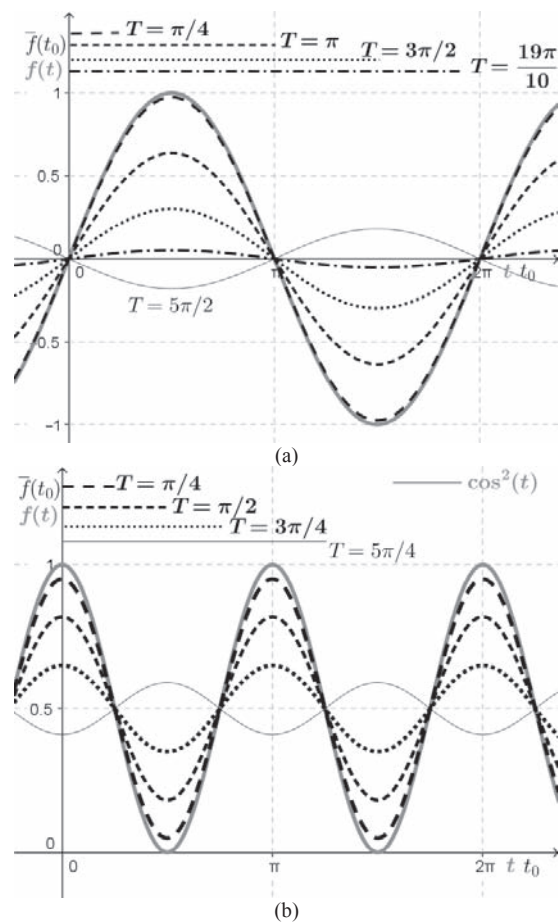


Abb. 4: Gleitender Mittelwert von (a) $f(t) = \sin(t)$ und (b) $f(t) = \cos^2(t)$. Die zu mittelnde Funktion ist dabei grau durchgezogen, die Mittelwerte schwarz strichliert. Die unterschiedlichen Mittelungslängen T sind in der Legende visualisiert: Die Strichlänge zeigt den Vergleich zu einer Periode. (Die dünne Kurve entspricht (a) $T = 5\pi/2$ und (b) $T = 5\pi/4$.)

Das Ziel, einen Trend aus den verrauschten Daten herauszufiltern, bedarf also einer moderaten, kontextbezogenen Wahl der Fensterbreite: Sie muss geeignet zwischen Signal und Rauschen vermitteln, um eine sinnvolle Interpretation zu ermöglichen.

Sitzt das Signal auf einer oszillierenden „Trägerwelle“, ist es sinnvoll, diesen periodischen Teil herauszumitteln. Die Wahl einer Periodenlänge als Fensterbreite glättet dann genau diejenigen Schwingungen, die für die Interpretation irrelevant sind. Steckt hingegen echte Information in periodischen Anteilen, möchte man diese entdecken und nicht wegmitteln.

Auch bei nichtperiodischen Funktionen muss sich die Wahl der Fensterbreite am Kontext orientieren. Bei Ereignissen, die auf eine Zeitspanne τ begrenzt sind, ist es meist sinnvoll, die Fensterbreite nicht wesentlich größer als τ zu wählen. Im einleitenden Beispiel ergibt sich aus dem Kontext eine mögliche Fensterbreite von drei bzw. zwölf Datenpunkten. Die kurze Intervalldauer mittelt über die Testresultate einzelner Trainings pro Woche und wird „Rauschen“ und periodische Effekte abschwächen. Bei einer Mittelungslänge von zwölf Trainingseinheiten, was der Dauer eines Monats entspricht, wird die Varianz in den Schwankungen stark gemindert und der Trend sollte gut erkennbar sein.

Im nächsten Abschnitt betrachten wir Beispieldatensätze, deren Trend durch die Bildung eines gleitenden Mittelwerts verdeutlicht wird.

3 Beispiele Klimaforschung

Der Einsatz gleitender Mittelwerte ist bei zahlreichen Problemstellungen aus alltäglichem Kontext hilfreich. Wir betrachten zwei Beispiele für die höheren Schulstufen. Dass dabei die Auswahl der Mittelungslänge Platz für unterschiedliche Darstellungen und Interpretationen lässt, überrascht nicht. Dieser Aspekt muss bei der Erstellung, Verwendung und Interpretation von Grafiken – im Unterricht immer wieder – hervorgehoben werden (vgl. zu dieser Thematik u. a. Rahmstorf, 2010). Auch ohne quantitative Analyse kann eine qualitative Thematisierung dieser Problematik in der Schule Platz finden.

Ein häufiges Diskussionsthema ist die Erwärmung des Weltklimas. Entsprechende Untersuchungen mitteln Aufzeichnungen von Wetterdaten gleitend über einen längeren Zeitraum T . Wie in Kapitel 2 diskutiert, muss man qualitativ unterschiedliche Effekte beachten: Viele klimatische Kenngrößen wie etwa Temperatur oder Konzentration der Treibhausgase unterliegen Schwankungen, die für eine Aussage über die Entwicklung des Klimas nicht von Bedeutung sind und das Erkennen längerfristiger Trends erschweren. Darunter fallen z. B. punktuell auftretende Wetterextreme („Rauschen“, hier: Dürre, Starkregen,

usw.), oder die jährlich periodischen Schwankungen (Jahreszeiten, Meeresströmungen, Monsun-Winde, usw.). All diese werden bei Wahl einer geeigneten Mittelungslänge eliminiert, um das eigentliche „Signal“ – den Anstieg der Temperatur – zu extrahieren.

Inwiefern singuläre, das Wetter potentiell weltweit über mehrere Jahre beeinflussende Ereignisse wie Vulkaneruptionen durch die Wahl entsprechender Fensterbreiten noch deutlich sichtbar sein sollen oder eben nicht, ist möglichst wertfrei zu diskutieren (Ziel *hier* ist die Festigung des mathematischen Zusammenhangs zwischen dargestellter Kurve und Länge des Mittelungsfensters, um entsprechende Interpretationskompetenz zu erwerben).

Weiters können regelmäßige Oszillationen extrem langer Periodendauer thematisiert werden (wie etwa die mehrere 10.000 Jahre dauernde Milanković Zyklen des Sonneneinfalls auf die Erde). Entsprechend all dieser Aspekte werden unterschiedliche Empfehlungen für die Mittelungslänge T gegeben; typische Werte liegen zwischen 5 und 50 Jahren (vgl. u. a. WMO, 2015).

Als erstes Beispiel ist in Abb. 5 die Temperaturänderung zwischen 1880 und 2014 dargestellt: Die schwarzen Datenpunkte (interpoliert durch die graue Linie) zeigen die jährlichen Abweichungen vom Mittelwert des Zeitraums 1951–1980, im Vergleich mit dem gleitenden Mittelwert über 5 (durchgehender schwarzer Graph) und 10 Jahre (strichliert). Offensichtlich werden die Temperaturspitzen und -einbrüche abgeflacht und glattere Graphen entstehen. Aus diesen gemittelten Werten lassen sich Aussagen über den Langzeittrend der Temperaturveränderungen in den letzten 130 Jahren treffen. Darüber hinaus wird anhand dieser Kurve deutlich, dass der 10-Jahres-Trend im Zeitraum 1945–1960 negativ war; zur vor-

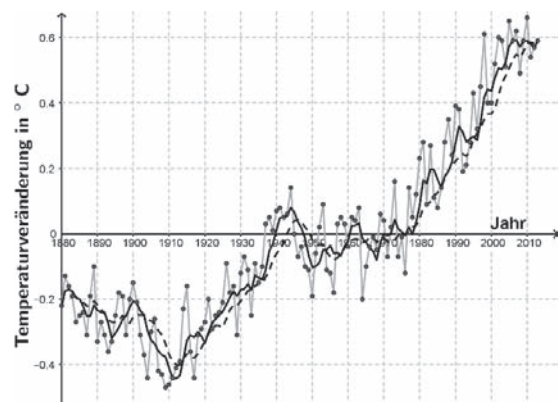


Abb. 5: Die Temperaturveränderung zwischen 1880 und 2014 (Rohdaten grau interpoliert) als Abweichung vom Mittel 1951–1980 (Nasa, 2015). Die durchgehende Linie (5 Jahre) und strichlierte (10 Jahre) stellen gleitende Mittelwerte dar.

letzten Jahrhundertwende sogar über einen Zeitraum von 30 Jahren, wohingegen ein nahezu konstanter Anstieg zwischen 1980–2014 zu beobachten ist.

Für die Unterrichtspraxis ist die Quelle der Daten aus Abb. 5 hilfreich (Nasa, 2015): Das *Goddard Institute for Space Studies* der NASA stellt auf seiner Webseite nicht nur Rohdaten zahlreicher Langzeittemperaturaufzeichnungen zur Verfügung, sondern zeigt auch in monatlich aktualisierten Grafiken vorgefertigte Auswertungen globaler Temperaturänderungen. Weiters finden sich dort Anleitungen, wie aus den Rohdaten diverse graphische Darstellungen gewonnen werden können. Obwohl die Schülerinnen und Schüler nicht enaktiv selbst über Jahrhunderte Daten erheben können, haben sie doch somit die Möglichkeit, mit echten Datensätzen und entsprechender Anleitung ihre eigenen Grafiken zu erzeugen, die Auswirkungen der Mittelungslängen zu testen. Somit bietet sich die Möglichkeit, Thesen durch Visualisierung in Graphiken zu untermauern.

Börsenkurse

Auch in der Finanzmathematik spielt der gleitende Mittelwert eine wichtige Rolle. Als Beispiel dient der Börsenkurs einer Aktie. Der Wert ändert sich wieder durch Rauschen (An-/Verkäufe schwanken mehrmals täglich), diskrete Ereignisse (Wahlen, Terroranschläge, Panikverkäufe, ...), periodische Effekte (Winterpause in der Baubranche, Weihnachtsgeschäft im Spielzeughandel, bekannte Konjunkturzyklen, ...) und tatsächliche Trends (Wert des Unternehmens). Um die Performance der Aktie (das „Signal“) zu verdeutlichen, werden die kurzfristigen Schwankungen wiederum durch eine gleitende Mittelung abgeschwächt.

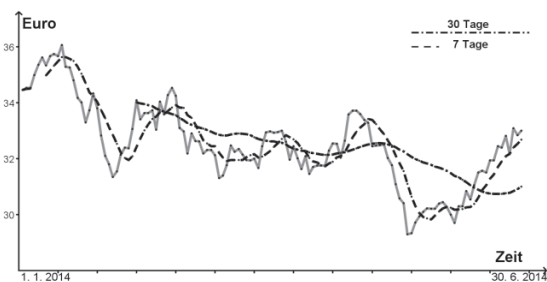


Abb. 6: Ein Börsenkurs (grau – jeweiliger Tageschlusskurs; Sparda, 2014) und seine täglichen Schwankungen. Glättung mittels gleitender Rückwärtsmittelung. strichliert: Mittelung über $T = 7$ Tage; strichpunktirt: $T = 30$ Tage.

Abbildung 6 zeigt einen konkreten Aktienkurs im Lauf eines halben Jahres (graue Linie). Die raschen Kursänderungen und starken Schwankungen innerhalb weniger Tage („Rauschen“) spiegeln nur bedingt

den Wert des Unternehmens wider. Aus der Graphik ist sowohl klar ersichtlich, dass der Mittelwert die Kurve glättet, und auch wie stark die resultierende Information von der Fensterbreite abhängt: Die strichlierte Linie stellt ein 7-Tage Mittel dar; sie folgt den Rohdaten etwas zeitverzögert, die Fluktuationen sind abgeschwächt, aber noch klar vorhanden. Die strichpunktirte Kurve (30-Tage Mittel) verdeutlicht den Abwärtstrend der Aktie im betrachteten Zeitraum und unterdrückt die Kursfluktuationen.

Solch ein „Verschleiern“ vorangegangener Kurseinbrüche könnte von unseriösen Anbietern bewusst gewählt werden. Offensichtlich ist eine solide Interpretation von Abbildungen nur möglich, wenn die Mittelungslänge explizit angegeben wird und dem Kontext der Fragestellung (z. B. „Muss das Wertepapier eventuell innerhalb eines Monats verkauft werden?“) angepasst ist.

Die betrachteten Problemstellungen sind aus Bereichen gewählt, die fächerübergreifenden Unterricht erlauben. Ihr Kontext liegt in Geographie (Klima) und Wirtschaftskunde (Finanzmarkt); im ersteren Fall ergibt sich ein weiterer Überlapp mit Chemie (Treibhausgase) und Physik (Temperatur), im letzteren mit Psychologie (z. B. Panikverkäufe). Die Anwendung der gleitenden mathematischen Mittelwertbildung bietet also eine Vielzahl von Möglichkeiten problemzentriert zu unterrichten.

Zudem bedient die hier vorgestellte Herangehensweise alle Repräsentationsmöglichkeiten des Lernens, insbesondere wenn Daten selbst aufgezeichnet und gesammelt werden können.

4 Anwendung in der Physik

Zeitreihen

Zahllose weitere Anwendungsgebiete ergeben sich innerhalb der Naturwissenschaften. Ähnlich zu den bereits in Kapitel 3 präsentierten Fällen sind Messwerte oft von Störungen überlagert, die aus der Apparatur oder der Experimentierumgebung stammen. In der Elektrotechnik erkennt man im gleitenden Mittelwert beispielsweise einen Tiefpassfilter, der hochfrequente Störungen beseitigt.

Materialeigenschaften

In einem Festkörper variiert beispielsweise die *räumliche* Ladungsverteilung auf atomarer Skala. Die Elektronenwolken stellen negative Ladungsverteilungen dar, die Atomrümpfe lokale positive Verteilungen (in idealen Kristallen sind beide zudem periodisch). Für die makroskopische Betrachtung des

Festkörpers sind solche Fluktuationen auf atomarer Ebene meist irrelevant. Um die makroskopischen Materialeigenschaften zu bestimmen, werden die mikroskopischen Ladungsfluktuationen der einzelnen Atome bzw. Moleküle räumlich gleitend gemittelt. Die für die Praxis unwichtigen Fluktuationen auf atomarer Ebene werden geglättet und die Ladungsverteilung in einen gemittelten, nur mehr schwach variierenden Wert zusammengefasst.

Im Gegensatz zu Zeitreihen, wo die Zukunft unbekannt ist, liegt hier kein Grund für eine Rückwärtsmittlung vor. Bei periodischen Strukturen ist das Resultat für die gemittelte Ladungsdichte unabhängig vom Startwert der Mittelung (siehe Diskussion und Abb. 3). Entsprechend erhält man „Materialkonstanten“ (im betrachteten Beispiel führt die Ladungsverteilung zur sogenannten Dielektrizitätskonstante).

Trägheit von Detektoren

In der Physik variieren Messgrößen oft schneller mit der Zeit als ein Detektor dies auflösen kann. Dies gilt insbesondere bei der Untersuchung von Licht im sichtbaren Bereich.

Jeder Detektor integriert die auftreffenden Impulse über seine Reaktionszeit und zeichnet somit ein gleitend gemittelt Signal auf. Als Beispiel kann hier durchaus die Messung jedes rasch oszillierenden Signals mit Hilfe eines „trägen“ Messapparats thematisiert werden – diese Problemstellung beinhaltet *konzeptionell* die Frage, wie sich das gleitende Mittel auf „schmale“, nicht-periodische Funktionen auswirkt.

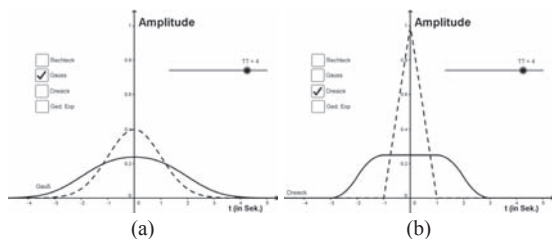


Abb. 7: (a) Ein gaußförmiger Lichtimpuls (gestrichelt) wird über eine Mittelungszeit von $T = 4$ s symmetrisch gemittelt. Der gemittelte Graph (schwarz) wird breiter und flacher. (b) Ein Dreiecksimpuls, der die gleiche Fläche wie die Gaußkurve umschließt, wird mit gleichem T gemittelt.

In Abb. 7 (a) ist die normierte Gaußfunktion (die Fläche unter der Kurve ist auf die Einheitsfläche normiert) dargestellt, mit

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

welche oft für die Beschreibung von Lichtimpulsen verwendet wird.

Diese Funktion wird nun zu jedem Zeitpunkt t über eine Zeitspanne $T = 4$ s gemittelt. Die schwarze Kurve zeigt das gemittelte Signal, also das, was ein Detektor aufzeichnet, der ja mit der Integrationszeit T reagiert.

Zum Vergleich wird in Abb. 7 (b) ein Dreieckspuls gemittelt, dessen Fläche ebenfalls normiert ist. Dieser Puls und das von einem Detektor aufgezeichnete Signal mit der gleichen Integrationszeit von wiederum $T = 4$ s sind dargestellt. Das gemittelte Signal ist durch die schwarze Kurve gegeben, wobei hier der Graph ebenfalls breiter und flacher wird.

In der qualitativen Betrachtung im Unterricht können daraus zwei Schlüsse gezogen werden: Erstens ist das gemessene Signal immer breiter und niedriger als das ursprüngliche. (Daraus lässt sich vermuten, dass die Flächen unter den beiden Kurven identisch sind. Dies lässt sich auch beweisen; bei Licht entspricht es dem fundamentalen Prinzip der Energieerhaltung.) Zweitens ist nun für diese nicht-periodischen Funktionen die gemittelte Größe nicht mehr unabhängig vom Mittelungszeitpunkt t_0 (schließlich ist das Ergebnis ja nicht einfach ein konstanter Wert \bar{f} , sondern die schwarze Kurve $\bar{f}(t_0, 4)$).

Die explizite Berechnung der gleitenden Mittelwerte kontinuierlicher Funktionen ist im regulären Unterricht zu herausfordernd. In unserem Applet (GeogebraTube, 2014) kann die Integrationszeit T mit Hilfe eines Schiebereglers verändert werden, was die Auswirkung einer langsameren Reaktion eines Detektors widerspiegelt. Weiters können neben einem gaußförmigen Impuls auch Rechteck-, verschiedene Dreieckssignale sowie gedämpfte Exponentialfunktionen untersucht werden. Diese Problemstellung tritt in der Physik nicht nur bei optischen Problemen, sondern insbesondere auch bei elektrotechnischen Anwendungen auf. Dies ermöglicht den Lernenden die Auswirkung eines trägen Detektors auf eine Messgröße zu untersuchen und gegebenenfalls mit einem Experiment zu vergleichen.

5 Zusammenfassung

In dieser Arbeit wurde der Einsatz von gleitenden Mittelwerten gezeigt. Dieses Konzept ist für vielfältige Bereiche anwendbar und bietet die Möglichkeit für Unterricht auf allen Repräsentationsstufen des Lernens nach Bruner.

Der gleitende Mittelwert glättet Schwankungen in diskreten Datensätzen, schwächt ihre Extrema („Ausreißer“) ab und verdeutlicht zugrunde liegende Trends („Signale“). Die gleichen Effekte treten bei kontinuierlichen Daten (Funktionen) auf, die mit

Hilfe eines Integrals gleitend gemittelt werden. Die Kontur der Funktion wird durch den Mittelungsprozess bei gleicher Fläche glatter, flacher und breiter.

Diese Auswirkungen können sowohl im diskreten als auch im kontinuierlichen Fall den Schülerinnen und Schülern der Sekundarstufe nicht nur quantitativ verdeutlicht werden. In einer qualitativen Behandlung kann man schon bald in der Sekundarstufe II thematisieren, wie die Intervallgröße, der Startwert und die Art der Mittelung (vorwärts, symmetrisch, rückwärts) die Resultate beeinflussen. Dadurch wird insbesondere die Interpretationskompetenz von Datendarstellungen im Alltag geschult.

Der hier präsentierte Ansatz bietet eine Erweiterung der mathematischen Kompetenzen von Schülerinnen und Schülern der Sekundarstufe mit vielen Einsatzmöglichkeiten im fächerübergreifenden, problemorientierten Unterricht.

Die Autoren danken dem Institut für Theoretische Physik an der Johannes Kepler Universität Linz, dem Europagymnasium Baumgartenberg sowie dem Gutachter für das hilfreiche Feedback.

Literatur

- BIFIE (2013): Praxishandbuch Mathematik AHS Oberstufe. Auf dem Weg zur standardisierten kompetenzorientierten Reifeprüfung. Teil 1. Wien.
- Bruner, J. (1976): *Der Prozess der Erziehung*, zit. nach: Barzel, B., Büchter A., Leuders. T. (2007). *Mathematik-Methodik*. Berlin: Cornelsen.
- Engel, J. (2009): *Anwendungsorientierte Mathematik: Von Daten zur Funktion. Eine Einführung in die mathematische Modellbildung für Lehramtsstudierende*. Heidelberg: Springer.
- GeogebraTube (2014): Zeitliche Mittelung von Signalen. <http://www.geogebraTube.org/material/show/id/312821> (Zugriff: 15.12.2014)
- Konold, J.; Mokros, J. (1995): Children's Concepts of Average and Representativeness. In: *Journal for Research in Mathematics Education*, 26 (1), 20–39.
- Nasa (2015): Goddard Institute for Space Studies, <http://data.giss.nasa.gov/gistemp/> (Zugriff: 03.01.2015)
- Rahmstorf, S. (2010): SciLogs <http://www.scilog.de/klimalounge/rwe-manager-vahrenholt-zum-klima/> (Zugriff: 15.12.2014)
- Russell, S. J.; Mokros J. (1996): What do Children Understand about Average? In: *Teaching Children Mathematics*, 2(6), S. 360–364.
- Sparda (2015): Tagesschlusskurs der ÖMV Aktie im Zeitraum 01.01.2014–30.06.2014.
- WMO (2015): Climate Data and Data Related Products, https://www.wmo.int/pages/themes/climate/climate_data_and_products.php (Zugriff: 12.08.2015).

Anschrift der Verfasser

Michael Himmelsbach
Anastasius-Grün-Straße 18
A – 4020 Linz
michi.himmelsbach@gmail.com

Helga M. Böhm
Johannes Kepler Universität Linz
Altenbergerstraße 69
A – 4040 Linz
helga.boehm@jku.at