

Analytische und chemische Methoden

Kurze Originalarbeit

Trendanalysen mit Mann-Kendall-Test und Theil-Schätzer Teil 2: Gewichtung der Daten

Dieter Helm

Robert Koch-Institut, FG 22/Umweltmedizin, Seestraße 10, D-13353 Berlin (helmd@rki.de)

DOI: <http://dx.doi.org/10.1065/uwsf2007.04.188>

Bitte zitieren Sie diesen Beitrag wie folgt: Helm D (2007): Trendanalysen mit Mann-Kendall-Test und Theil-Schätzer. Teil 2: Gewichtung der Daten. UWSF – Z Umweltchem Ökotox 19 (4) 240–242

Zusammenfassung

Ziel und Absicht. Es wird eine Erweiterung für Trendanalysen mit dem Mann-Kendall-Test und der linearen Regression nach Theil vorgestellt, die eine Gewichtung der zu analysierenden Daten ermöglicht.

Methoden. An Hand eines erdachten Beispiels wird die Auswirkung der Gewichtung auf die Ergebnisse des Mann-Kendall-Tests und der Regression nach Theil demonstriert. Die Gewichtungsfaktoren leiten sich dabei von den Stichprobenumfängen ab, auf die sich aggregierte Daten oder Mischproben gründen. Ein Wert, der die Information von mehr Individuen repräsentiert, soll stärkere Beachtung finden als ein Wert, der auf weniger Individuen zurückgeht.

Ergebnisse und Schlussfolgerungen. Die Erweiterung wirkt sich auf die Ergebnisse der Theil-Regression aus. Unter Gewichtung neigt sich die Ausgleichsgerade stärker den Datenpunkten mit größeren Stichproben zu.

Ausblick. Die Umsetzung der vorgeschlagenen Erweiterungen zur Trendanalyse mit dem Mann-Kendall-Test und der Regression nach Theil (Berücksichtigung von Messunsicherheiten und Gewichtung der Daten) steht in der Praxis noch aus.

Schlagwörter: Gewichtung; Mann-Kendall-Test; Theil-Schätzer; Trendanalyse; Umweltbeobachtung

Abstract

Weighting of data in trend analysis

Goal and Scope. An extension for trend analysis with Mann Kendall test and Theil estimator is presented which enables the weighting of data by weight factors.

Methods. It is shown in which way the weighting can effect the outcome of a trend analysis. Weighting is done in relation to the different sample sizes of the data points. A value or an observation derived from many specimens will be considered more reliable than a value derived from only a few specimens and will thus get a higher weighting factor.

Results. When weighting is applied, this will effect the Theil estimator rather than the Mann Kendall test. As a result the regression line is tilted towards those points with higher weights.

Outlook. Future must show whether the extensions presented here and in part 1 prove to be practical for the trend analysis in environmental monitoring or not.

Keywords: Environmental monitoring; Mann-Kendall test; Theil estimator; trend analysis; weighting

Einleitung

Im ersten Teil des Beitrags (Helm 2007) wurde eine Erweiterung des Mann-Kendall-Tests auf Trend vorgeschlagen, bei der die Unsicherheit von Messdaten (measurement uncertainty) berücksichtigt werden kann. Es konnte gezeigt werden, dass durch die Berücksichtigung von Messunsicherheiten Trendaussagen sicherer werden. Im jetzt vorliegenden, zweiten Teil des Beitrags wird vorgestellt, wie der Mann-Kendall-Test um die Gewichtung von Daten erweitert werden kann. Hierzu werden wieder die Beispieldaten des ersten Teils verwendet. Bezüglich der Grundlagen von Mann-Kendall-Test und Theil-Schätzer sei ebenfalls auf den ersten Teil verwiesen.

Gewichtung von Daten. In der Praxis der Umweltbeobachtung weicht die realisierbare Probenanzahl häufig von der angestrebten ab. Ein Beispiel: Nach den Verfahrensrichtlinien der Umweltprobenbank sollen für jede Probenahme fläche mindestens 20 Brassen (*Abramis brama*) mit einem Alter von 8–12 Jahren entnommen werden (UBA 1996). Aufgrund von Veränderungen der Populationsstrukturen können diese Vorgaben jedoch nicht immer eingehalten werden. Obwohl die Fanggröße in der Regel zwischen 20–52 Individuen beträgt (Helm 2003), waren es beispielsweise im Jahr 2002 für die Probenahme flächen Cumlosen und Prossen (beide: Elbe) nur 12 bzw. 14 Tiere (UBA 2007).

Es ist verständlich, dass eine Messung, die auf einem größeren Stichprobenumfang fußt, zu einer belastbareren Aussage führt als eine Messung, die sich nur auf eine kleine Stichprobe stützt. Durch die Einführung einer unterschiedlichen Gewichtung der Daten kann diesem Umstand Rechnung getragen werden. Dabei werden Daten, die auf einem größeren Stichprobenumfang beruhen, stärker gewichtet als solche, die auf einem kleinen Stichprobenumfang beruhen. Es gilt: großer Umfang = hohes Gewicht, geringer Umfang = niedriges Gewicht.

Streng zu scheiden von dieser Art der Gewichtung ist jene, die in bevölkerungsrepräsentativen Erhebungen zur Korrektur von Ziehungsabweichungen eingesetzt wird, wie etwa beim Umwelt-Survey (Becker et al. 2002). Hier geht es darum, geringe Abweichungen zwischen Ziehung und der bekannten Bevölkerungsstruktur auszugleichen. Angenommen, Männer jüngeren Alters seien in der Stichprobe schwächer vertreten als es ihrem Anteil in der Gesamtbevölkerung entspricht (was häufig der Fall ist). Man wird dann für die Bildung bevölkerungsrepräsentativer Mittelwerte die Daten dieser Gruppe etwas höher gewichten und gleichzeitig überrepräsentierte Gruppen untergewichten. Hier gilt: zu großer Umfang = niedriges Gewicht, zu geringer Umfang = höheres Gewicht.

In einem weiter gefassten Sinn spielen Wichtungsfaktoren auch bei anderen Methoden zur Zeitreihenanalyse innerhalb der Ökologie/Ökotoxikologie eine Rolle. Basierend auf der Annahme, dass die Varianz der Unsicherheiten proportional zur Höhe der gemessenen Konzentration ist, wird bei Regressionsanalysen mit dem Kehrwert der quadrierten Messwerte gewichtet (Conquest 2000, Pröhl et al. 2004). Glättungsverfahren wie gleitender Durchschnitt (moving average) berechnen für jeden Messwert die gewichtete Summe über eine bestimmte Anzahl von Werten (French und Lindley 2000, Smith und Anderson-Cook 2000), wobei die Gewichte gleich (z.B. bei 3 Werten jeweils 1/3, bei 5 Werten 1/5) oder ungleich sein können. Auch die Spline-Funktionen können als eine Art lokaler, gewichteter gleitender Durchschnitt aufgefasst werden. In Bereichen geringer Messwertdichte wird stärker geglättet als in Bereichen zeitlich hoch aufgelöster Messwertdichte (Pesch und Schroeder 2006).

1 Methoden

Wir verwenden wieder das fiktive Beispiel aus dem ersten Teil, dessen Ergebnisse für die ungewichtete Trendanalyse noch einmal in **Schema 1** zusammengefasst sind. Aus der Sicht des Mathematikers enthält auch dieses Schema bereits Gewichte, die jedoch implizit alle auf den Wert von '1' gesetzt sind und deswegen im Schema nicht auftauchen.

Schema 1

Jahr	2001	2002	2003	2004
Zeit t	1	2	3	4
Messwert y	2,0	5,0	7,0	6,0
$\Delta y / \Delta t = (y_j - y_i) / (t_j - t_i)$ für i = 1 ... (n-1) und j = i+1 ... n		+3/1	+5/2	+4/3
sgn (Δy)		+1	+1	+1
			+1	-1

Um die 'echten' Gewichte (ungleich '1') aus den Daten ableiten zu können, schreiben wir ein zweites Schema mit den Messwerten und den Stichprobenumfängen. Zu jedem Messwert y_i gehört ein Stichprobenumfang z_i , der zugleich der Gewichtungsfaktor für diesen Messwert ist (4. Zeile in **Schema 2**). Die hier eingesetzten Stichprobenumfänge sind wiederum fiktiv; sie wurden passend gewählt, um eine Änderung der Regressionsgeraden zu bewirken, aber ihre Größenordnung lehnt sich an den 'Echtfall' Beprobung von Brassens für die UPB an.

Schema 2

Jahr	2001	2002	2003	2004
Zeit t	1	2	3	4
Messwert y	2,0	5,0	7,0	6,0
Stichprobenumfang z	20	40	14	30
Paarstichprobenumfang, $z_{ij} = \min(z_i, z_j)$ für i = 1 ... (n-1) und j = i+1 ... n		20	14	20
			14	30
				14

Für die Bestimmung der Gewichtungsfaktoren der Messwertänderungen $\Delta y / \Delta t$ ist ein anderes Vorgehen notwendig, da hier immer zwei Stichprobenumfänge (z_i und z_j der Zeitpunkte t_i und t_j) vorliegen. Da die statistische Belastbarkeit eines Wert für $\Delta y / \Delta t$ durch den kleineren der beiden Stichprobenumfänge begrenzt wird, bestimmen wir den Paarstichprobenumfang für jedes $\Delta y / \Delta t$ als $\min(z_i, z_j)$, wie in den Zeilen 5 bis 7 von **Schema 2** eingetragen.

Prüfung auf Trend. Bei der Prüfung auf Trend ändert sich im Vergleich zum ursprünglichen Beispiel aus **Schema 1** nichts, da es – ungeachtet von Gewichtungsfaktoren – immer nur einen Anstieg der Daten, eine Abnahme oder ein Gleichbleiben geben kann. Demzufolge ist die Prüfgröße C wiederum +4, was nach der Kendall'schen K-Statistik für $n=4$ (Conover 1980) auf 5%-Niveau ($p = 0,05$) einen signifikant ansteigenden Trend der Daten ergibt.

Bestimmung der Regression. Benötigt werden die Gewichtungsfaktoren jedoch für die Bestimmung der Mediane von y und $\Delta y / \Delta t$ nach der Methode der gewichteten Mediane (Aghajanyan 2005). Dazu werden die Werte für y bzw. für $\Delta y / \Delta t$ aufsteigend angeordnet (**Schemata 3,4**), wobei die Gewichtungsfaktoren (= Stichprobenumfänge) hinzugefügt werden. In einer zusätzlichen Zeile werden die kumulierten Stichprobenumfänge eingetragen. Der gewichtete Median ist jener Wert von y bzw. $\Delta y / \Delta t$, auf den die Hälfte des Summenwertes der von links nach rechts kumulierten Stichprobenumfänge zeigt. Bei den Messwerten y (**Schema 2**) ist dies $y_2 = 5,0$, da die Hälfte der Summe aller Stichprobenumfänge ($104/2 = 52$) auf diesen Wert verweist ($20 < 52 < 60$).

Entsprechendes gilt für den Median von $\Delta y / \Delta t$. Hier zeigt die Hälfte der kumulierten Paarstichprobenumfänge ($112/2 = 56$) auf $\Delta y / \Delta t = +4/3$ oder 1,33 (da $44 < 56 < 64$). Die gewichteten Mediane sind also $\text{Median}(y) = 5,0$ und $\text{Median}(\Delta y / \Delta t) = 1,33$ – ungewichtet waren es $\text{Median}(y) = 5,5$ und $\text{Median}(\Delta y / \Delta t) = 1,7$. (Falls Programme zur Verfügung stehen, mit denen gewichtete Mediane berechnet werden können, wie z.B. SPSS, kann die Bestimmung natürlich mit diesen Programmen erfolgen.)

Die resultierende Regressionsgerade nach Theil (Conquest 2000) lautet: $y'_{(\text{gew})} = 1,68 + 1,33t$. Pro Jahr steigt der Messwert um 1,33 Einheiten. Damit verläuft die gewichtete Regressionsgerade deutlich flacher als die ungewichtete, im ersten Teil (Helm 2007) bestimmte ($y' = 1,3 + 1,7t$), wie in **Abb. 1** dargestellt ist.

Schema 3

Messwert y	2,0	5,0	6,0	7,0
Stichprobenumfang n_i	20	40	30	14
Σn_i	20	60	90	104

Schema 4

$\Delta y / \Delta t$	-1/1	+1/2	+4/3	+2/1	+5/2	+3/1
Paarstichprobenumfang n_{ij}	14	30	20	14	14	20
Σn_{ij}	14	44	64	78	92	112

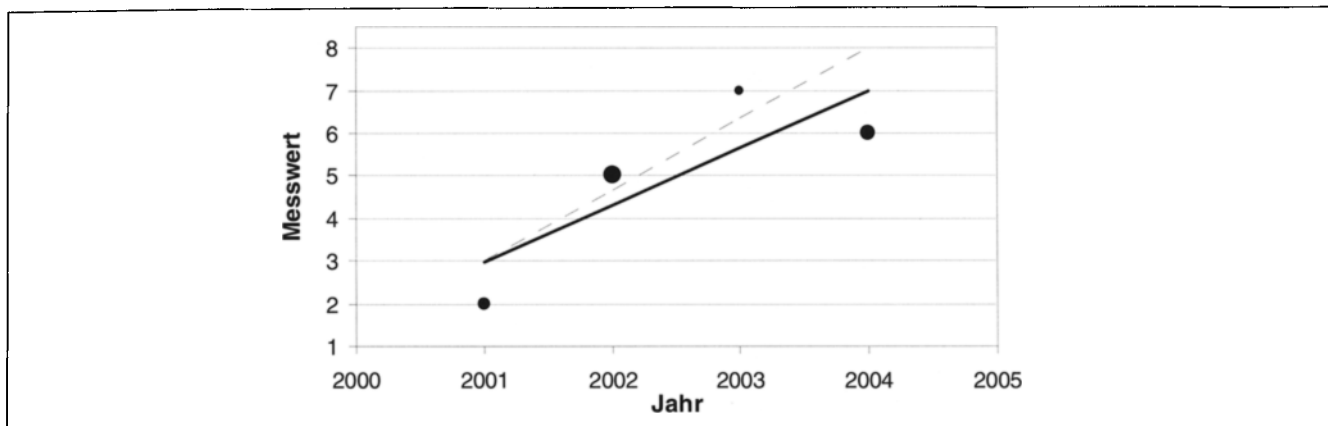


Abb. 1: Regressionsgeraden nach Theil (gestrichelte Linie: ungewichtet; durchgezogene Linie: gewichtet; Kreise: Messwerte; die Stichprobenumfänge [Gewichte] sind durch die Größe der Kreise angedeutet)

2 Ergebnisse und Diskussion

Messwerten, die auf einer größeren Stichprobe beruhen, kann eine höhere Reliabilität zugeschrieben werden als Messwerten, die auf einem geringeren Stichprobenumfang fußen. Indem die unterschiedlichen Stichprobenumfänge als Gewichtungsfaktoren in die Berechnung eingehen, wird diesem Umstand Rechnung getragen. Bezüglich der Trendtests mit Mann-Kendall und Theil-Schätzer bedeutet dies, dass sich die Regressionsgerade den Messwerten mit größerem Stichprobenumfang stärker 'zuneigt' als dies bei einer ungewichteten Analyse der Fall ist. Im gewählten Beispiel geht der Messwert mit 7,0 Einheiten, der auf einer Stichprobe von nur 14 Individuen beruht, schwächer in die Rechnung ein als der Messwert mit 6,0 Einheiten, dem 30 Individuen zugrunde liegen. Demzufolge 'kippt' die Gerade in Richtung des zweithöchsten Wertes. Der Nutzen einer Gewichtung liegt darin, dass eine durch unterschiedlich große Stichprobenumfänge verzerrte Probennahme im nachhinein wieder 'entzerrt' wird.

Alternativ zu den Stichprobenumfängen können auch, sofern vorhanden, die Streuungen der Messwerte als Gewichtungsfaktoren eingesetzt werden. Dabei geht man von dem Gedanken aus, dass – bei einer gleich hoch erwarteten Varianz – eine Messung mit geringerer Streuung (Standardabweichung oder Standardfehler) eine höhere Präzision und damit eine höhere Reliabilität zeigt. Die entsprechenden Gewichtungsfaktoren berechnen sich aus dem Kehrwert der Streumaße. Dabei werden allerdings häufig gebrochene Gewichtungsfaktoren auftreten. Um den gewichteten Median (nach der hier vorgestellten Methode) bestimmen zu können, wird man die gebrochenen Gewichtungsfaktoren so erweitern müssen, dass sie ganzzahlig werden. Alternativ sei wieder auf Statistikprogramme verwiesen, die es ermöglichen, einen gewichteten Median zu berechnen.

3 Schlussfolgerungen und Ausblick

Hauptanliegen der beiden Teile des vorliegenden Beitrages war es, Erweiterungen für die Trendanalyse mittels Mann-Kendall-Test und Theil-Schätzer vorzuschlagen. Es ging um die Berücksichtigung von Messunsicherheiten und Stichprobenumfängen. Eine dritte Möglichkeit, den Test auf nicht-

lineare Daten zu erweitern, wurde im ersten Teil kurz angesprochen. Durch geeignete Transformation werden nicht-linearisierte Daten linearisiert und können danach dem Test zugeführt werden. Das Lehrbuch von Sachs (1984) führt eine Reihe von geeigneten Transformationsfunktionen auf.

Ein erhoffter Nebeneffekt des Beitrages ist, das Augenmerk der Kollegen in der Umweltbeobachtung vermehrt auf die Verwendung von statistischen Methoden für die Auswertung ihrer Messdaten zu lenken, speziell für eine Testung auf Trend.

Literatur

- Aghajanyan GG (2005): Core inflation in a small transition country: Choice of optimal measures. *The European Journal of Comparative Economics* (2) 83–110
- Becker K, Kaus S, Krause C, Lepom P, Schulz C, Seiwert M, Seifert B (2002): Umwelt-Survey 1998 Band III: Human-Biomonitoring – Stoffgehalte in Blut und Urin der Bevölkerung in Deutschland. Berlin, Firma Werbung und Vertrieb
- Conover WJ (1980): *Practical nonparametric statistics*. 2. ed, John Wiley & Sons
- Conquest L (2000): Environmental monitoring: Investigating associations and trends. In: Sparks T (ed), *Statistics in Ecotoxicology*. Chichester, New York
- French D, Lindley D (2000): Exploring the data. In: *Statistics in Ecotoxicology* (ed Sparks T), Chichester, New York
- Helm D (2003): Bewertung von Monitoringdaten der Umweltprobenbank des Bundes mit der Hasse-Diagramm-Technik. *UWSF – Z Umweltchem Ökotox* 15 (2) 85–94
- Helm D (2007): Trendanalysen mit Mann-Kendall-Test und Theil-Schätzer. 1. Teil: Berücksichtigung von Messunsicherheiten. *UWSF – Z Umweltchem Ökotox* 19 (2) 91–95
- Pesch R, Schroeder W (2006): Mosses as bioindicators for metal accumulation: Statistical aggregation of measurement data to exposure indices. *Ecological Indicators* 6 (1) 137–152
- Pröhl G, Fiedler I, Klemm E, Zibold G, Ehlken S (2004): Erfassung ökologischer Halbwertszeiten von ^{90}Sr und ^{137}Cs in terrestrischen und aquatischen Ökosystemen. BMU-2004-640. Schriftenreihe Reaktorsicherheit und Strahlenschutz, BMU
- Sachs L (1984): *Angewandte Statistik*. Springer-Verlag
- Smith E, Anderson-Cook C (2000): Analysis of field studies: Regression analysis. In: Sparks T (ed), *Statistics in Ecotoxicology*. Chichester, New York
- UBA (1996): Umweltprobenbank des Bundes. Verfahrensrichtlinien für Probenahme, Transport, Lagerung und chemische Charakterisierung von Umwelt- und Human-Organproben. Erich Schmidt Verlag, Berlin (<http://193.174.169.38/wwwupb/servelet/upb>)
- UBA (2007): www.umweltprobenbank.de (Stand 07.02.2007)

Erhalten: 11. Februar 2007
Akzeptiert: 22. April 2007
OnlineFirst: 23. April 2007