

Bearbeitung und Speicherung von Daten aus automatischen Datenaufzeichnungssystemen

Processing and Storing of "Data Logger" Data

J. FANK¹⁾

Inhalt

| | Seite |
|--|-------|
| 1. Einleitung und Problemstellung | 222 |
| 2. Ausgleich von Meßwerten..... | 223 |
| 3. Datenkompression | 225 |
| 4. Spline-Interpolation..... | 228 |
| 5. Beispiele | 230 |
| 5.1. Lufttemperaturmessung Station Wagna im Juli 1992 | 230 |
| 5.2. Bodentemperaturmessung in 110 cm Tiefe Station Wagna im Juli 1992 | 233 |
| 5.3. Oberflächenabfluß Pegel Höhenhansl (Pöllau, Oststeiermark) im Juli und August 1992 | 234 |
| 6. Zusammenfassung und Wertung der Ergebnisse..... | 235 |
| Kurzfassung | 235 |
| Literatur..... | 236 |
| Summary | 236 |

¹⁾ Dr. J. FANK, Institut für Geothermie und Hydrogeologie, Joanneum Research, Elisabethstraße 16–18, A-8010 Graz.

1. Einleitung und Problemstellung

In den unterschiedlichsten Bereichen der Naturwissenschaften kommen in den letzten Jahren verstärkt elektronische Meßsysteme und automatische Datenaufzeichnungssysteme zum Einsatz, um den gesteigerten Ansprüchen hinsichtlich Aufzeichnungsdichte und Registriergenauigkeit nachzukommen. Neben den Vorteilen einer hochauflösenden zeitlichen Messung verschiedenster Parameter bringen die großen Datenmengen aber auch Probleme im Bereich der Datenaufbereitung, Datenverwaltung und Datensicherung.

In den unterschiedlichsten Teilbereichen der Hydrogeologie gewinnt in letzter Zeit die Untersuchung von Einzelereignissen zunehmend Bedeutung. Bedingt durch die Forderung nach einer detaillierten Registrierung unterschiedlichster Parameter während kurzdauernder Ereignisse müssen die entsprechenden Meßwerte in einem sehr engen Zeitintervall erfaßt werden. Für statistische Bearbeitungen von Daten ist es meist auch wünschenswert, wenn alle zu untersuchenden Parameter in einem einheitlichen Zeitraster vorliegen.

Aus Gründen einer effizienten Speicherverwaltung zeichnen die Registriereinrichtungen meist in einem konstanten Zeitintervall auf, sodaß neben den Einzelereignissen auch jene Phasen, in denen die zu messenden Parameter kaum Schwankungen aufweisen, im vorgegebenen dichten Zeitraster aufgezeichnet werden. Dies entspricht durchaus den Vorgaben für eine effizientere Bearbeitung der Daten, stellt aber andererseits hohe Anforderungen an Datenbanksysteme oder andere Einrichtungen zur Datenspeicherung (Massenspeicherkapazität), die im Bereich der PC's noch keineswegs gegeben sind. Als Beispiel der zu erwartenden Datenmengen sei hier nur angeführt, daß die Wasserstandsjahresganglinie eines Oberflächenpegels bei fünfminütigem Meßintervall – das zur Bewertung von Einzelereignissen durchaus zu fordern ist – 105 120 Datensätze umfaßt.

Für die Speicherung und Evidenthaltung dieser Daten in Datenbanken müssen deshalb Methoden gefunden werden, die den Datenbestand deutlich reduzieren, ohne allerdings wichtige Details in den Ganglinien zu verlieren, d. h. im Zuge von Ereignissen sollen möglichst alle Daten, in Phasen mit geringem Schwankungsverhalten der Parameter sollen nur jene Datensätze gespeichert werden, die die Ganglinie charakterisieren. Diese Forderung schließt die Speicherung lediglich von Mittelwerten über konstante Zeiträume aus, da die Informationen über die Einzelereignisse verloren gehen.

Im Gegensatz zur Forderung nach möglichst geringen zu speichernden Datenmengen steht der Wunsch, zum Zeitpunkt der Bearbeitung und Auswertung wieder einen vollständigen, von den Meßwerten möglichst wenig unterschiedlichen Datenbestand zur Verfügung zu haben. Das bedeutet, daß bei der Verwendung von Algorithmen zur Datenreduktion auch solche zur Verfügung zu stellen sind, die den komprimierten Datenbestand wieder auf einen vorzuziehenden Zeitraster expandiert. Erfolgt dieser Vorgang durch Anwendung numerischer Verfahren, so ist mit einem Fehlbetrag zwischen Originaldaten und expandierten Daten zu rechnen.

Jede Meßwertaufnahme über elektrische Signale ist mit einem meßtechnischen Fehler behaftet, der die Meßgenauigkeit des jeweiligen Systems definiert. Dieses Hintergrundrauschen fällt bei größerem Schwankungsbereich der Parameter nicht ins Gewicht, kann sich allerdings bei der Dateninterpretation störend auswirken, wenn die tatsächlichen Schwankungen des zu messenden Parameters in ruhigen

Phasen außerordentlich gering sind. In vielen Fällen ist es daher nötig, dieses Rauschen aus den Datenbeständen vor der Auswertung zu eliminieren.

2. Ausgleich von Meßwerten

Wie schon oben erwähnt, ist jede Meßwertaufnahme über elektrische Signale mit einem Hintergrundrauschen behaftet, das dann negativ ins Gewicht fällt, wenn die tatsächliche Schwankung des zu messenden Parameters sehr gering ist. Sollen die Meßwertreihen einer späteren Datenreduktion mit darauffolgender Interpolation unterzogen werden, ist das Hintergrundrauschen jedenfalls vor der Reduktion aus den Datenbeständen zu entfernen und ein Meßwertausgleich durchzuführen.

In Fig. 1 sind die Meßergebnisse einer Drucksonde an einem Grundwasserspiegel in Wagna (Steiermark, Österreich) für den Zeitraum vom 19. bis 31. Juli 1992 dargestellt. Während dieses Zeitraumes herrschte ein kontinuierlicher Rückgang des Grundwasserspiegels, der durch die installierte Drucksonde mit elektrischer Meßwertübertragung relativ stark verrauscht innerhalb der Meßgenauigkeit der Drucksonde wiedergegeben wird. Der in den folgenden Absätzen beschriebene Meßwertausgleich wurde durchgeführt und das Ergebnis in Fig. 1 dargestellt. Es zeigt sich ein sehr gut weiterzubearbeitender kontinuierlicher Verlauf der Ganglinie, der wesentlich besser den natürlichen Verhältnissen angepaßt ist. Die Messung der Drucksonde erfolgt alle fünf Minuten, im Datensammler wird ein Mittelwert über

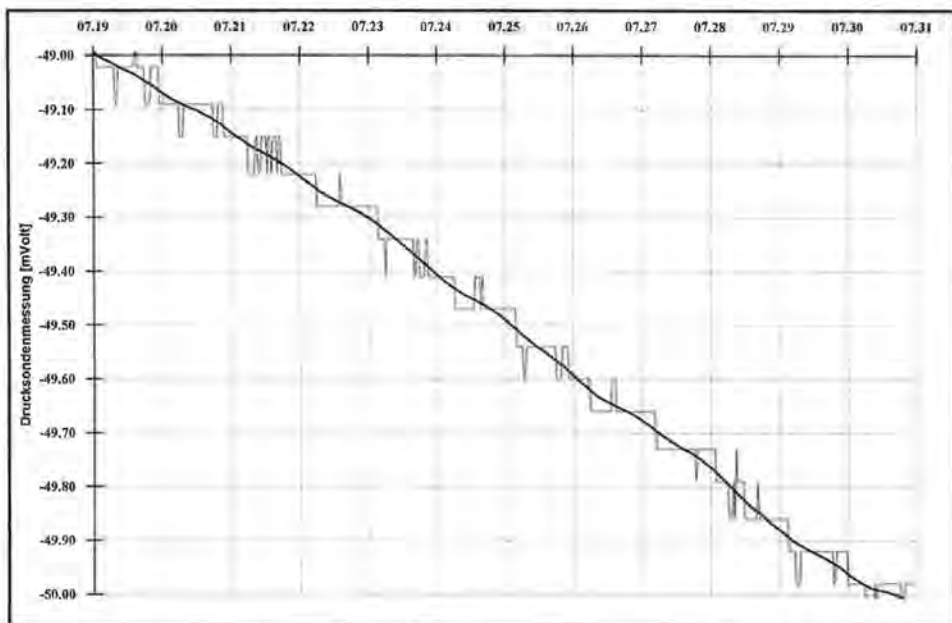


Fig. 1: Hintergrundrauschen bei der Aufnahme der Grundwasserspiegellage an der Station Wagna (Steiermark, Österreich) mittels einer Drucksonde. Für den Zeitraum 19. bis 31. Juli 1992 mit kontinuierlich fallendem Grundwasserspiegel wurde ein Meßwertausgleich durchgeführt und dargestellt.

jeweils sechs Messungen (1/2 Stundenmittel) gespeichert. Nach Durchführung des Meßwertausgleichs stehen gleichfalls Halbstundenwerte zur Verfügung.

Das hier vorgestellte Verfahren des Meßwertausgleichs beruht darauf, daß die benachbarten Stützstellen eines i -ten Punktes entsprechend einer festzulegenden Gewichtung mit in die Berechnung des Ausgleiches einfließen. Dies darf jedoch nur in dem Rahmen geschehen, in dem die Meßwerte zuverlässig sind. Die Gewichte, mit Werten zwischen 0 und 1 sind für einen anzugebenden Bereich von Nachbarstützstellen ($i - m : i + m$) definiert. Im einfachsten Fall ist die Gewichtung linear, es sind jedoch auch andere Funktionen denkbar (D. HILBERG, 1989).

Da die Stützstellenabstände auf der x -Achse (Zeitachse) ungleichmäßig sein können und näher liegende Punkte höher als entferntere bewertet werden sollen, muß ein normierter Abstand

$$x_{\text{norm}} = (x_n - x_1)/(n - 1)$$

eingeführt werden (n = Anzahl der Wertepaare). Der Gewichtungsfaktor für jede Stützstelle j im Intervall ($i - m : i + m$) errechnet sich somit nach der Gleichung:

$$G_j = 1 - \left(\frac{1}{m} \cdot \frac{\text{abs}(x_i - x_j)^p}{x_{\text{norm}}} \right),$$

wobei p der gewünschte Polynomgrad ist ($p = 1$ für linearen Verlauf). Der korrigierte y -Wert an der i -ten Stelle ist dann die Summe aller bewerteten y -Werte, normiert auf die Summe aller Bewertungen (D. HILBERG, 1989):

$$y_i = \frac{\sum_{j=i-m}^{i+m} W_j \cdot y_j}{\sum_{j=i-m}^{i+m} W_j}$$

Die Bewertungsfunktion zur Berechnung der y -Werte der Ausgleichsganglinie kann somit durch die Parameter m und p gesteuert werden, wobei sich für p ein Wert zwischen 0,5 und 2 als sinnvoll erwiesen hat (D. HILBERG, 1989). Für m gibt D. HILBERG einen sinnvollen Bereich zwischen 2 und 5 an, jedoch erwies sich im Zuge der Anwendung der vorgestellten Funktion bei sehr dichtem Aufzeichnungsintervall von mehr als 24 Wertepaaren pro Tag ein Wert von 10 als geeigneter (Die Ausgleichsfunktion in Fig. 1 wurde mit $p = 1$ und $m = 10$ ermittelt, wobei eine maximale Abweichung des errechneten Wertes vom Meßwert von $\pm 0,05$ mVolt zugelassen wurde).

Ein weiterer indirekter Steuerungsfaktor für den Meßwertausgleich ist ein Rahmen, in dem die Meßwerte verändert werden dürfen. Ergibt sich aus der Berechnung eine größere Abweichung als die zulässige, so wird die Abweichung an dieser Stützstelle auf die maximal zulässige gesetzt.

Bei Verfahren, bei denen eine Approximierungsfunktion im mathematischen Algorithmus der Interpolationsfunktion implementiert ist, kann man den Grad des Ausgleiches nicht unabhängig von der Größe der Fehler gestalten. Im Gegensatz dazu ist die hier vorgestellte Ausgleichsfunktion, die mathematisch sehr einfach und damit auch sehr schnell ist, wesentlich flexibler. Diese Flexibilität gestaltet andererseits die Arbeit mit der Funktion als etwas zeitaufwendig, weil praktisch für jeden unterschiedlichen Ganglinientyp und für unterschiedliche Sensoren unterschiedliche Steuerungsparameter notwendig sein können.

Wie das Beispiel in Fig. 1 zeigt, paßt sich die Ausgleichsfunktion sehr schön an die mittlere Meßwertverteilung an und ergibt einen kontinuierlichen Verlauf der

berechneten Ganglinie, die damit bestens für die im folgenden vorgestellten Datenkompressionsalgorithmen vorbereitet ist.

3. Datenkompression

In Fig. 2 ist ein Ablaufplan für ein Programm zur Reduktion von Daten in Zeitreihen dargestellt.

Der hier vorgestellte Algorithmus zur Datenreduktion beruht auf der Tatsache, daß die 2. Ableitung einer Ganglinie Aussagen über die Krümmung der Ganglinie an deren Stützstellen trifft. Die mathematischen Grundlagen für die Berechnung der 2. Ableitung sind in W. MEIER & D. HARTMANN (1988), in C. H. REINISCH (1967) und in H. SPATH (1986) ausführlich ausgeführt.

Berechnet man die 2. Ableitung einer Kurve an den Stützstellen der Ganglinie, so zeigen die Werte der 2. Ableitung umso stärker von 0 verschiedene Werte, je stärker die Krümmung der Kurve an dieser Stützstelle ist. Je nach der Richtung der Krümmung ändert sich das Vorzeichen des Wertes der 2. Ableitung. Dieser Tatsache zufolge können die charakteristischen Punkte einer Kurve dadurch ermittelt werden, daß die Maxima und die Minima der 2. Ableitung an den Stützstellen für die Auswahl der Stützstellen der reduzierten Ganglinie herangezogen wird. Werden nun die bereits ausgewählten Stützstellen für die Weiterbearbeitung nicht mehr herangezogen, kann diese Auswahl von Elementen mit maximaler und minimaler 2. Ableitung solange durchgeführt werden, bis entweder die gewünschte Anzahl an ausgewählten Elementen, oder eine bestimmte Grenze des Wertes der 2. Ableitung erreicht ist.

Voraussetzung für die Bearbeitung einer Zeitreihe nach diesem Verfahren ist, daß die Zeitreihe nach Datum und Uhrzeit aufsteigend sortiert vorliegt. Als Steuerungselemente für das Kompressionsprogramm in Fig. 2 stehen die Parameter „zahl“ und „grenze“ zur Verfügung. Die Auswahl neuer Stützstellen für die reduzierte Ganglinie wird solange fortgeführt bis die angegebene Zahl [zahl] von Elementen, bzw. der angegebene Grenzwert [grenze] der 2. Ableitung erreicht ist (Fig. 2). Zur besseren Handhabbarkeit des Grenzwertes „grenze“ werden die Werte der 2. Ableitung an den Stützstellen im Wertebereich zwischen -1 und $+1$ normiert. Um sicherzustellen, daß der Anfangs- und Endpunkt der Originalganglinie jedenfalls in die reduzierte Ganglinie übernommen wird, wird der Wert der zweiten Ableitung an der ersten und der letzten Stützstelle auf $+1$ (= Maximum) gesetzt.

Nachdem die Auswahl der Stützstellen der neuen Ganglinie durch Erreichen einer der beiden Grenzvorgaben abgeschlossen ist, werden die ermittelten Elemente nach Datum und Zeit aufsteigend sortiert und als neue, reduzierte Ganglinie auf einen Datenfile geschrieben. Diese neue Ganglinie enthält als Datenpunkte jene Stellen der Originalganglinie, an denen die Krümmung am deutlichsten erkennbar ist, d. h. im Zuge von kurzfristigen Ereignissen ist eine große Datenanzahl in der reduzierten Ganglinie enthalten, in Phasen relativer Ruhe werden nur wenige Datenpunkte ausgewählt.

Das Schwankungsverhalten der Werte der normierten 2. Ableitungen selbst hängt extrem von der Form der Originalganglinie ab. Ist die Ausgangskurve von relativer Gleichförmigkeit, die nur durch wenige, aber extreme Ereignisse unterbrochen wird (z. B. Oberflächenabfluß), so liegen die Mittelwerte der normierten positiven und negativen 2. Ableitungen sehr nahe bei ± 0 , sind die Einzelereignisse weniger extrem

Programm KOMPRESS

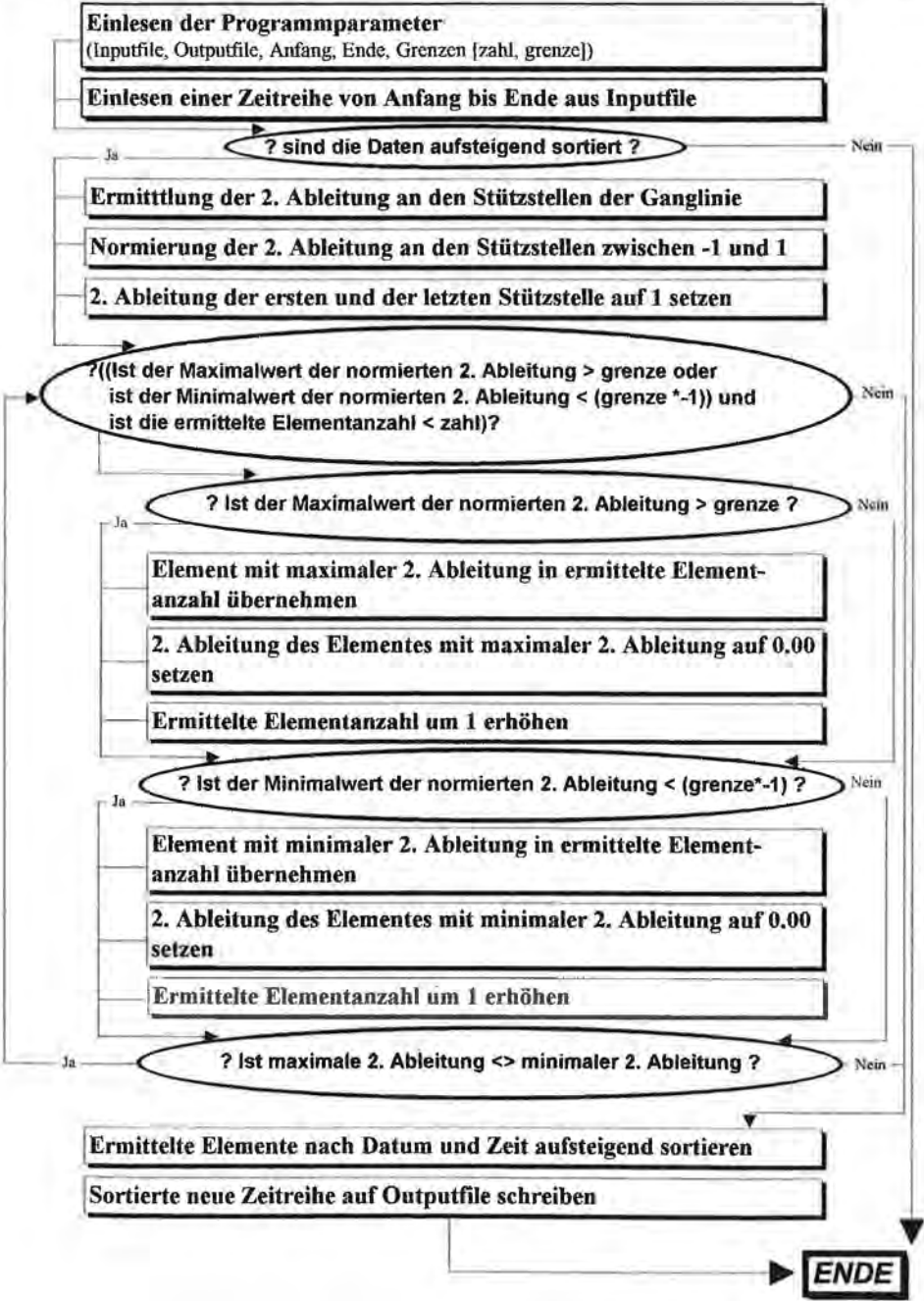


Fig. 2: Ablaufplan eines Programmes zur Reduktion von Datensätzen in Zeitreihen.

ausgebildet bzw. wird die Gleichförmigkeit der Ganglinie stärker (z. B. Grundwassertemperatur), so geht der Mittelwert der normierten positiven 2. Ableitungen gegen +1, der negativen gegen -1. Das bedeutet, daß die Grenzwerte [grenze] zur Berechnung der neuen Ganglinie für jeden Ganglinientyp neu zu suchen und anzugeben sind.

Die Zahl der Elemente, die notwendig sind, den Verlauf einer Ganglinie ausreichend zu charakterisieren, hängt ebenfalls von der Form der Ganglinie ab. Die Zahl der auszuwählenden Elemente hängt aber auch sehr stark vom Verwendungszweck der reduzierten Ganglinie ab bzw. von den Anforderungen an die genaue Repräsentanz der Originalganglinie durch die reduzierte Ganglinie. Auch dieser Steuerungsparameter ist daher für jeden Ganglinientyp zu eruiieren und anzugeben.

In Fig. 3 ist die Wirkungsweise der Datenreduktion dargestellt. Die Oberflächenabflußganglinie am Pegel Höhenhansl in Pöllau (Steiermark, Österreich) ist als Originalganglinie und als reduzierte Ganglinie mit unterschiedlichen Reduktionsfaktoren für den 1. August 1992 dargestellt. Zwecks der besseren Darstellung wurden die reduzierten Ganglinien gegenüber der darüberliegend dargestellten um jeweils 2 l/s vermindert.

Die Daten, die in Fig. 3 dargestellt sind, entstammen einer Wasserspiegelmessung mittels einer Drucksonde am Pegel Höhenhansl in Pöllau. Die Wasserspiegellagen wurden über eine Schlüsselkurve in Abflüsse umgerechnet. Das Einzugsgebiet des Pegels ist sehr klein (0,43 km²), sodaß anthropogene Einflüsse im Quellbereich des Prätsibaches auf das Abflußgeschehen deutliche Auswirkungen zeigen.

Die Wasserspiegellagen werden von einer Drucksonde aufgezeichnet, die einen Meßbereich von 3,5 m hat. Durch diesen relativ hohen Meßbereich der Sonde ist

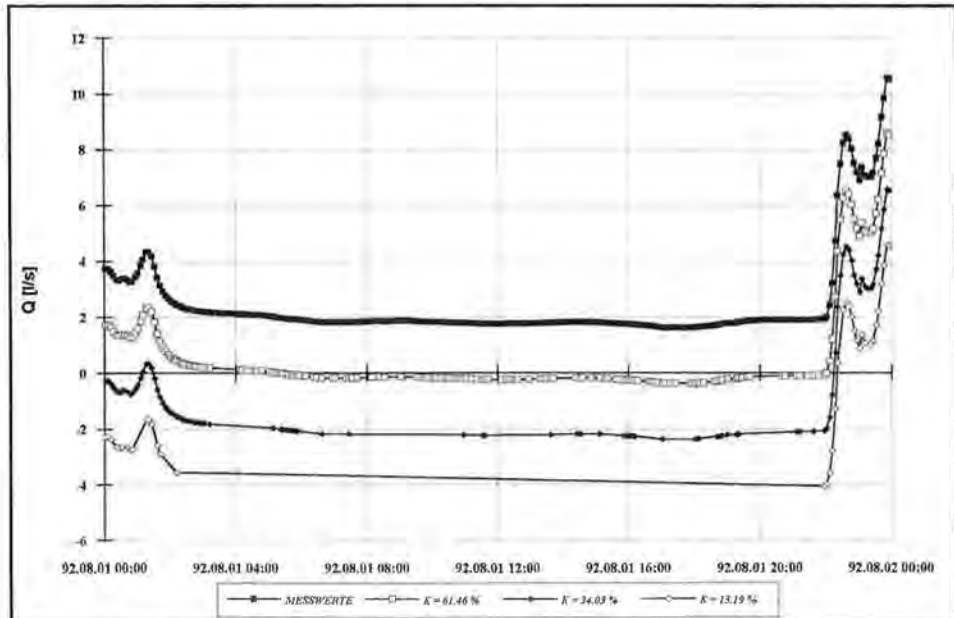


Fig. 3: Wirkungsweise der Datenreduktion mit unterschiedlichen Reduktionsfaktoren. Die Werte der einzelnen Datenreihen wurden gegenüber den zuvor dargestellten um jeweils 2 l/s vermindert, um den Verlauf der Reduktion besser erkennbar gestalten zu können.

die Meßgenauigkeit etwa $\pm 0,2$ mm. Vor der Umrechnung in Abflüsse wurden die Wasserstandswerte mit dem Meßwertausgleichsverfahren von Kap. 2. bearbeitet. Der Wasserspiegel wird von einem Datensammler in einem konstanten Zeitintervall von fünf Minuten aufgezeichnet.

Der 1. August 1992 an diesem Pegel war charakterisiert durch einen leichten Abflußpeak um etwa 2^h morgens und einem darauf folgenden Abflußrückgang mit leichten, anthropogen bedingten Schwankungen während des ganzen Tages. Um etwa 22^h führte beginnender Regen zu einem Anstieg des Wasserstandes, der bis zum Tagesende mit einer leichten Unterbrechung anhielt.

Während die Datenreduktion auf ca. 61 bzw. etwa 34% der Originaldatensätze noch sehr gut auch die kleinen Details der Ganglinie widerspiegeln, werden bei einer Reduktion auf ca. 13% nur noch die Abflußereignisse befriedigend wiedergegeben. Die leichten Abflußschwankungen während des Tages gehen dabei verloren. Deutlich erkennbar ist die Auslichtung des Datenbestandes an den Ganglinienteilen mit nur geringfügigen Veränderungen.

4. Spline-Interpolation

Für eine große Zahl von Auswertungen, v. a. von zeitabhängigen Daten müssen die Daten in einem einheitlichen Raster vorliegen. Für diese Auswertungen müssen die reduzierten Datensätze (vgl. Kap. 3.) wieder erweitert werden, d. h. es müssen in die Zeitreihen in einem vorzugebenden Zeitraster Daten interpoliert werden. Mittels der „Spline-Interpolation“ kann ein geschmeidiger Kurvenzug konstruiert werden, der exakt durch die vorgegebenen Punkte verläuft. Da die gesplinte Kurve in ihrem ganzen Verlauf als mathematisch exakt definierte Form vorliegt, kann für jeden Punkt der Zeitachse aus dieser Funktion ein Wert berechnet werden, der genau auf der Kurve liegt.

Eine bekannte Möglichkeit zur Interpolation von Punkten ist das Lagrangesche Polynom (R. BULIRSCH, 1968), das zwar durch alle Stützpunkte verläuft, aber mit wachsender Datenanzahl sehr stark oszilliert und daher für den vorgegebenen Zweck völlig unbrauchbar ist. Bessere Ergebnisse liefern kubische Polynome, die für jedes Intervall zwischen zwei Stützstellen definiert werden. Dadurch ist die Forderung nach Glattheit innerhalb der Intervalle automatisch erfüllt. Wenn zusätzlich auch noch die Stetigkeit der 2. Ableitung an den Stoßstellen erfüllt ist, erhält man als Ergebnis die kubische Spline-Interpolation. Setzt man die 2. Ableitung an den Rändern der ersten und letzten Intervallfunktion auf 0, entstehen natürliche kubische Splines.

In Fig. 4 sind verschiedene Interpolationsverfahren bei der Interpolation einer Treppenfunktion, sicherlich eines der schwierigsten Kapitel der Spline-Interpolation, dargestellt. Es ist klar zu erkennen, daß die Interpolation mittels kubischer Splines in diesem Fall viel zu sehr überschwingt und die interpolierten Werte sehr weit vom Verlauf der Verbindung der Datenpunkte entfernt ist.

Eine wesentliche Verbesserung läßt sich durch die Einführung sogenannter „rationaler“ Splines erreichen (Fig. 4). Hier wird die intervallweise Interpolation mittels nichtpolynomischer Funktionen durchgeführt, die ebenfalls zweimal stetig differenzierbar aneinander gesetzt werden. Durch die Vorgabe eines Glättungsparameters läßt sich das Überschwingen der Kurve mehr oder weniger stark dämpfen.

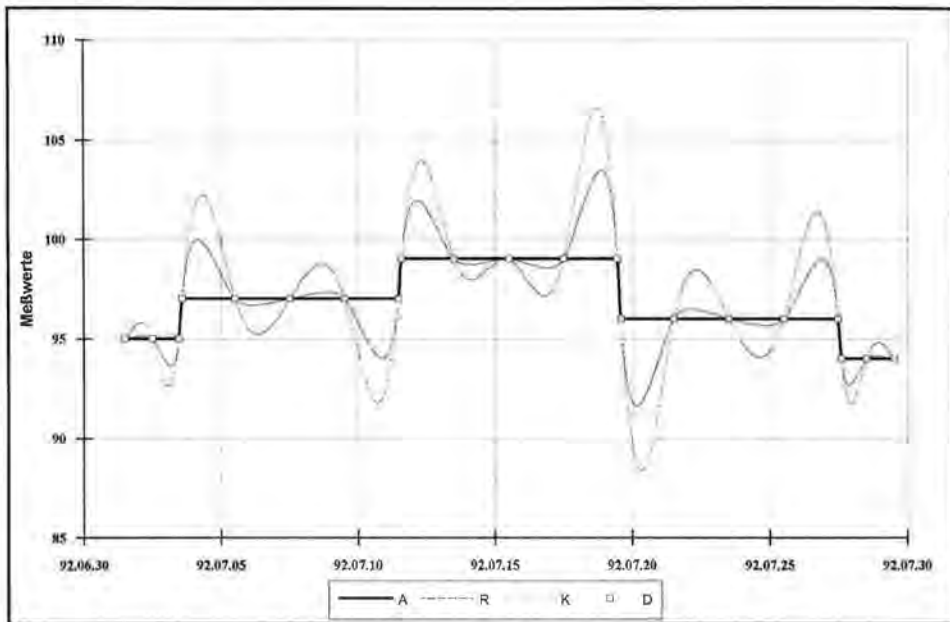


Fig. 4: Interpolation einer Treppenfunktion. Unterschiedliche Spline-Interpolationsverfahren zeigen deutlich unterschiedliche Kurvenverläufe. A = Akima-Splines, R = Rationale Splines, K = Kubische Splines, D = Datenpunkte.

Eine Interpolation mit Glättungsparameter = 0 liefert wieder kubische Splines. Bei extrem großen Glättungsparameter nähert sich der Spline mehr und mehr einem kantigen Linienzug und interpoliert sogar Treppenfunktionen. Der Rationale Spline in Fig. 4 wurde mit dem Glättungsparameter von 3 interpoliert.

Der Akima-Spline Ansatz orientiert sich an der Art, in der man beim manuellen Einzeichnen einer Interpolationskurve vorgeht. H. AKIMA entwickelte 1970 eine neue Methode, die ebenfalls Polynome dritten Grades zwischen den Stützstellen definiert, jedoch auf die Stetigkeit der 2. Ableitung in den Stützstellen verzichtet. Gerade diese Bedingung erzeugte die lästigen „Überschwinger“ bei starkem Gradientenwechsel in anderen Interpolationsverfahren (H. AKIMA, 1970). Die mathematischen Grundlagen für die Akima-Interpolation sind in D. HILBERG (1989) detailliert ausgeführt. Neben dem Vorteil, daß die Akima-Interpolation auch in schwierigsten Fällen (Treppenfunktion in Fig. 4) die „natürlichsten“ Kurven liefert, hat diese Methode noch weitere Vorteile:

- Der Speicherbedarf ist durchwegs geringer als bei einer vergleichbaren Spline-Interpolation,
- die einfacheren Algorithmen führen zu wesentlich kürzeren Rechenzeiten.

Von ihrer Definition her ist die Akima-Interpolation in jenen Fällen überfordert, in denen für nachfolgende Berechnungen auf die Stetigkeit der 2. Ableitung nicht verzichtet werden kann.

Für die aktuelle Fragestellung der Datendekompression von Zeitreihen wurde ein Programm entwickelt, das Datenreihen eines einheitlichen Zeitraumes mit

unterschiedlichem Rhythmus in den einzelnen Reihen auf ein einheitliches Zeitraster interpoliert. Der Programmablauf folgt in etwa folgendem Weg:

- Einlesen der Programmparameter.
- Solange Zeitreihenfiles vorhanden sind:
 - Einlesen einer komprimierten Zeitreihe,
 - Auswahl des Spline-Interpolationsverfahrens,
 - Berechnung der Polynomkoeffizientenfelder A, B, C, D,
 - Interpolation von Werten nach einem vorgegebenen konstanten Zeitintervall,
 - Schreiben der interpolierten Ganglinie auf einen temporären File.
- Erstellen eines Files mit den interpolierten Daten aller Ganglinien in einem Rasterfile.

Das Ergebnis der Dekompression ist ein Datenfile, der in der Art einer Matrix aufgebaut ist, in dem spaltenweise die einzelnen Ganglinien gleichabständigen Zeitdaten in den Zeilen zugeordnet sind. Als Interpolationsverfahren stehen rationale Splines mit der möglichen Angabe des Glättungsparameters und Akima-Splines zur Verfügung. Da bei der Datenkompression nach dem Schema in Kap. 3. nur jene Datensätze reduziert werden, an denen die Ganglinie keine oder nur äußerst untergeordnete Krümmungen aufweist, liefert in den meisten Fällen die Akima-Interpolation die besten Ergebnisse bei der Dekompression, da mit diesem Verfahren „wiggles“ (Überschwinger) am besten verhindert werden und die beste Anpassung der rückinterpolierten Datensätze an die Meßdaten erreicht werden kann.

5. Beispiele

Im folgenden soll anhand von drei Beispielen aus unterschiedlichen Bereichen die Anwendbarkeit der vorgestellten Algorithmen diskutiert und der zu erwartende Fehlerbereich abgeschätzt werden. Die Lufttemperatur und die Bodentemperatur werden an einer Meßstation in Wagna in der Südsteiermark (Österreich) gemessen. Die Messung der Sonden erfolgt in einem konstanten Intervall von fünf Minuten. Der Mittelwert von jeweils sechs Meßwerten wird in einem Datensammler abgelegt, sodaß als Originaldatensätze Halbstundenmittelwerte zur Verfügung stehen. Die Meßanordnung zur Registrierung des Oberflächenabflusses am Pegel Höhenhansl in Pöllau (Steiermark, Österreich) ist bereits in Kap. 2. näher erläutert.

5.1. Lufttemperaturmessung Station Wagna im Juli 1992

Die Ganglinie der Lufttemperatur an der Station Wagna im Juli 1992 ist durch einen relativ starken Tagesgang charakterisiert, dessen Regelmäßigkeit nur unter dem Einfluß des Wechsels der Wetterlagen gebrochen wird.

Der Originaldatensatz mit Halbstundenmittelwerten wurde mehreren Kompressionsphasen, wie in Kap. 2. beschrieben, unterzogen. Die Grenzwerte der normierten 2. Ableitung in den einzelnen Durchläufen lagen bei 0,05, 0,10, 0,15 und 0,20. Die Meßdaten und die Ganglinien der einzelnen Reduktionsphasen sind in Fig. 5 dargestellt. Der Verlauf der Datenreduktion unter Anwendung unterschiedlicher Gewichtungsfaktoren sowie deren Kennzahlen sind in Tab. 1 dargestellt.

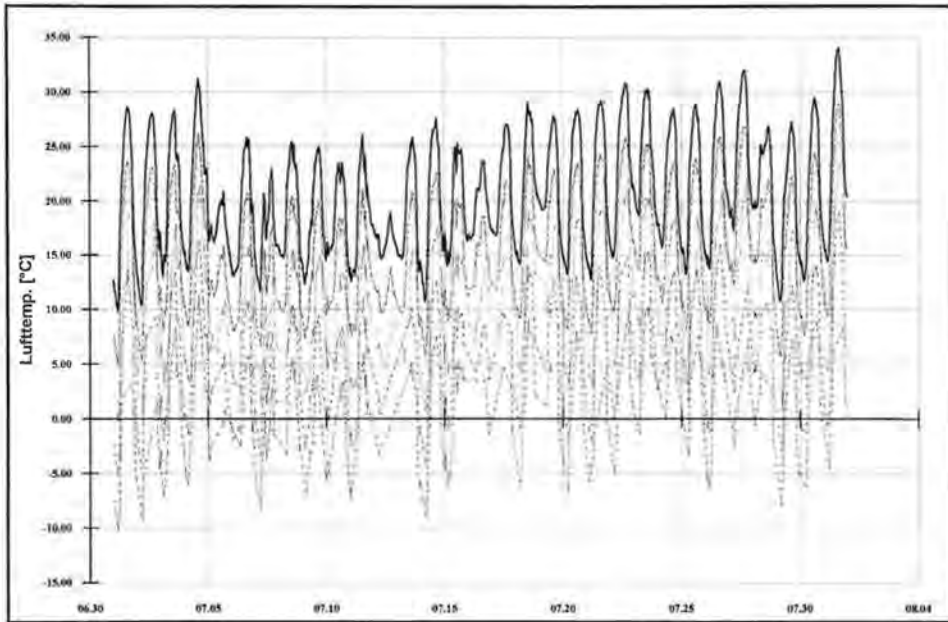


Fig. 5: Reduktion der Datensätze der Ganglinie der Lufttemperatur an der Station Wagna im Juli 1992. Die Ergebnisse der einzelnen Reduktionsphasen wurden um jeweils 5°C vermindert, um sie anschaulicher darstellen zu können.

Aus der Darstellung in Fig. 5 ist klar erkennbar, daß mit der Verminderung der Datensätze, die Originalganglinie immer undeutlicher nachgezeichnet wird. Während die Tagesgänge der Lufttemperatur auch noch mit sehr hohem Kompressionsfaktor erkennbar sind, sind die Details der Ganglinie wohl nur noch in der zweiten Kurve von oben ersichtlich, weshalb für die weiteren Betrachtungen ein Kompressionsfaktor von 0,05 zugrundegelegt wird. Wie aus Tab. 1 ersichtlich, werden dadurch die Datensätze von 1488 auf 944 oder auf 66,8% reduziert.

Die Datenreduktion selbst verläuft bei der Lufttemperatur der Station Wagna im Juli 1992 nach einer exponentiellen Funktion in Abhängigkeit vom Gewichtungsfaktor (Tab. 1). Bei einem Faktor von 0,08 werden nur noch 50% der Originaldatensätze zur Beschreibung der reduzierten Ganglinie herangezogen.

Um einen Vergleich der Originaldaten mit der reduzierten Datenreihe zu ermöglichen und um das Datenmaterial statistisch zu vereinheitlichen, wurde nach dem Verfahren der Akima-Spline-Interpolation (Kap. 4.) die reduzierte Datenreihe auf das Meßintervall und den Zeitraster der Originaldatenreihe hochgerechnet. Daraufhin wurden die Originaldatenreihe und die berechnete Ganglinie miteinander verglichen. Die Kennwerte dieses Vergleiches sind in Tab. 1, die Originaldatenreihe, die interpolierte Ganglinie und die Abweichungen zwischen diesen beiden sind in Fig. 6 dargestellt.

Der optische Vergleich der beiden Ganglinien (gemessen und interpoliert) in Fig. 6 läßt praktisch keine Unterschiede erkennen. Minimum, Maximum und Mittelwert der Ganglinie werden nach der Reduktion der Daten auf etwa zwei Drittel der Originaldaten und darauffolgender Interpolation mit einem Fehler von max. $0,08^{\circ}\text{C}$ wiedergegeben (Tab. 1). Der mittlere Fehler zwischen Originaldatenreihe und

Tab. 1: Kennzahlen der Datenreduktion bei der Ganglinie der Lufttemperatur der Station Wagna im Juli 1992 mit den auftretenden Fehlern bei der Rückinterpolation.

| LUFTTEMPERATUR: Station WAGNA (Juli 1992) | | | | | |
|---|---------|--------|--------|--------|--------|
| Gewichtungsfaktor | 0,00 | 0,005 | 0,10 | 0,15 | 0,20 |
| Datensatzanzahl absolut | 1488,00 | 944,00 | 635,00 | 390,00 | 256,00 |
| Datensatzanzahl relativ | 100,00 | 66,80 | 42,67 | 26,10 | 17,20 |
| Minimum (°C) | 9,83 | 9,83 | 9,83 | 9,40 | 9,47 |
| Mittelwert (°C) | 20,58 | 20,57 | 20,50 | 20,67 | 20,84 |
| Maximum (°C) | 34,00 | 34,08 | 34,13 | 34,15 | 33,57 |
| Interpolationsfehler (Meßwert – interpolierter Wert): | | | | | |
| größter negativer Fehler absolut | 0,00 | -1,54 | -1,46 | -7,27 | -7,96 |
| größter negativer Fehler relativ | 0,00 | 107,51 | 107,98 | 149,62 | 154,34 |
| mittlerer Fehler absolut | 0,00 | 0,01 | 0,08 | -0,09 | -0,26 |
| mittlerer Fehler relativ | 0,00 | 99,94 | 99,54 | 100,61 | 101,76 |
| größter positiver Fehler absolut | 0,00 | 1,82 | 2,91 | 4,93 | 4,99 |
| größter positiver Fehler relativ | 0,00 | 91,80 | 85,47 | 71,91 | 68,72 |

interpolierter Datenreihe bei einem Meßintervall von einer halben Stunde liegt bei $0,01^{\circ}\text{C}$, die größten Abweichungen liegen allerdings schon bei mehr als $\pm 1,5^{\circ}\text{C}$ (Tab. 1), was auf den Algorithmus der Spline-Interpolation zurückgeführt werden

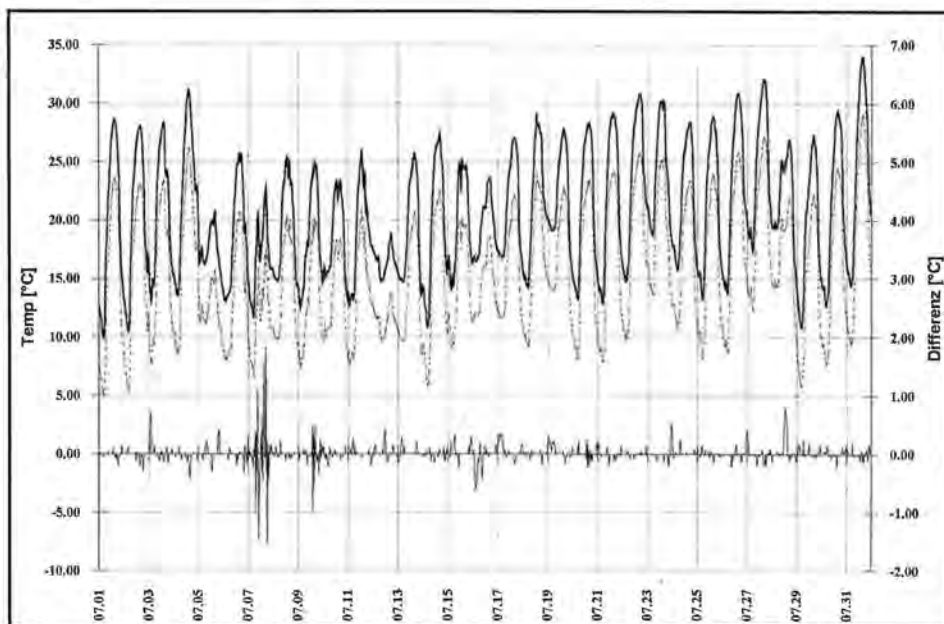


Fig. 6: Gemessene Ganglinie (dicke durchgezogene Linie), reduzierte und rückinterpolierte Ganglinie (strichpunktierte dünne Linie) der Lufttemperatur an der Station Wagna im Juli 1992 im Vergleich. Zusätzlich ist die Differenz zwischen den beiden Ganglinien (durchgezogene dünne Linie) dargestellt. Die Werte der reduzierten und rückinterpolierten Ganglinie sind zwecks der besseren Anschaulichkeit um jeweils 5°C vermindert.

muß. Wie in Fig. 6 weiters erkennbar ist, treten diese maximalen Abweichungen zu jenen Zeiten auf, in denen der „normale“ Gang der Lufttemperatur durch Änderung der Witterung unterbrochen ist. Als Kennzeichen dafür, daß diese maximalen Differenzen praktisch nur die Extremwerte beeinflussen, sei noch angeführt, daß die Tagesmittel der Lufttemperatur errechnet aus den interpolierten Werten als arithmetisches Mittel der Halbstundenmittel sich im Maximum nur um $0,01^{\circ}\text{C}$ (am 9. Juli 1992) von den Tagesmitteln, errechnet aus den Originaldaten unterscheidet.

In diesem Beispiel ist noch zu berücksichtigen, daß vor der Datenreduktion kein Meßwertausgleich durchgeführt wurde, da ein „elektrisches Rauschen“ aus der Meßwertaufnahme aufgrund der großen Wertschwankungen in der Ganglinie nicht erkennbar, aber sicherlich vorhanden war. Dies führt dazu, daß gerade in Zeiten stark wechselnder Gradienten in der Zeitreihe bei der Interpolation leichte „Überschwinger“ nicht zu vermeiden sind, welche in kurzen Bereichen die Extremwerte gegenüber den Originaldaten relativ stark beeinflussen können.

5.2. Bodentemperaturmessung in 110 cm Tiefe Station Wagna im Juli 1992

Die Ganglinie der Bodentemperatur in 110 cm Tiefe, die an der Station Wagna im Halbstundenrhythmus als Ergebnis von fünfminütigen Messungen erfaßt wird, ist aufgrund der Meßumgebung und der Einflußfaktoren (Niederschlag, Witterungsänderungen, Bodenbeschaffenheit, Vegetation etc.) sehr gedämpft. Dieses

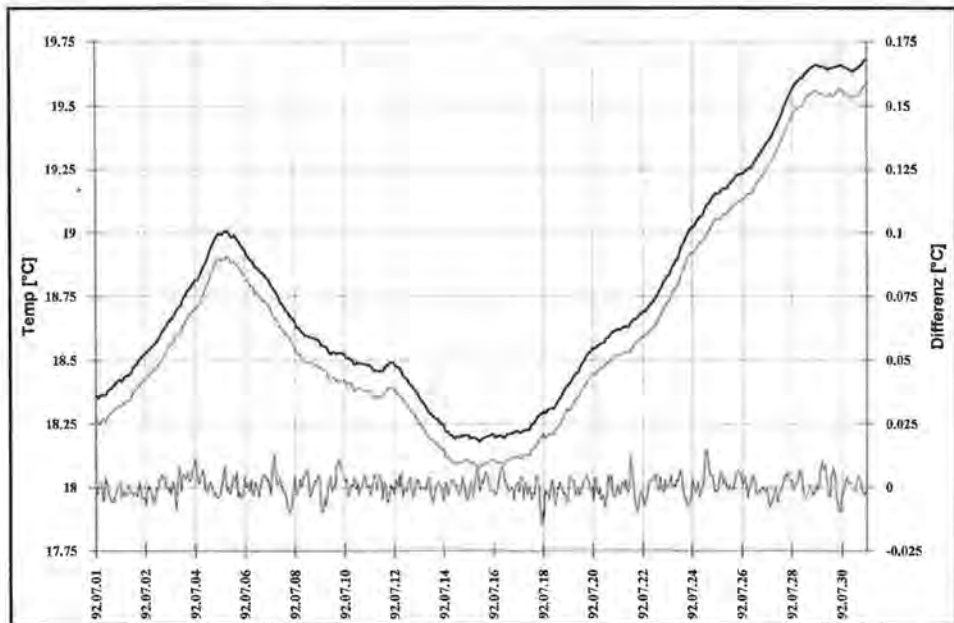


Fig. 7: Gemessene Ganglinie (dicke durchgezogene Linie), reduzierte und rückinterpolierte Ganglinie (strichpunktierte dünne Linie) der Bodentemperatur an der Station Wagna im Juli 1992 im Vergleich. Zusätzlich ist die Differenz zwischen den beiden Ganglinien (durchgezogene dünne Linie) dargestellt. Die Werte der reduzierten und rückinterpolierten Ganglinie sind zwecks der besseren Anschaulichkeit um $0,1^{\circ}\text{C}$ vermindert.

gedämpfte Schwankungsverhalten macht eine Datenreduktion nach den Methoden in Kap. 3. sehr leicht möglich und führt bei einer Reduktion der Originaldaten auf etwa 18% zu einer nahezu identen Ganglinie. Ein Vergleich zwischen gemessener Ganglinie und reduzierter und rückinterpolierter Ganglinie (Fig. 7) führt zu einem maximalen Fehler von nur $0,016^{\circ}\text{C}$, weit innerhalb der Meßgenauigkeit des Systems.

5.3. Oberflächenabfluß Pegel Höhenhansl (Pöllau, Oststeiermark) im Juli und August 1992

Im Bereich der Hydrologie konnten die Datensätze einer Abflußganglinie eines Pegels in einem kleinen kristallinen Einzugsgebiet auf ca. ein Drittel reduziert werden. Trotzdem sind nach der Rückinterpolation des ursprünglichen Meßintervalls von fünf Minuten nicht nur die eingetretenen Abflußereignisse, sondern auch die anthropogen bedingten Tagesschwankungen des Abflusses bei Niederwasser klar ersichtlich (Fig. 8). Die reduzierte und rückinterpolierte Ganglinie wurde gegenüber der Originalganglinie zwecks der besseren Anschaulichkeit um jeweils 2 l/s vermindert. Die durch die Akima-Spline-Interpolation entstandenen Differenzen zwischen der gemessenen und der berechneten Ganglinie betrug maximal $\pm 0,08\text{ l/s}$, entsprechend einer Wasserstandsänderung von weniger als einem halben Millimeter. Die Fehler liegen demgemäß auch hier innerhalb der Ungenauigkeit des eingesetzten Meßsystems.

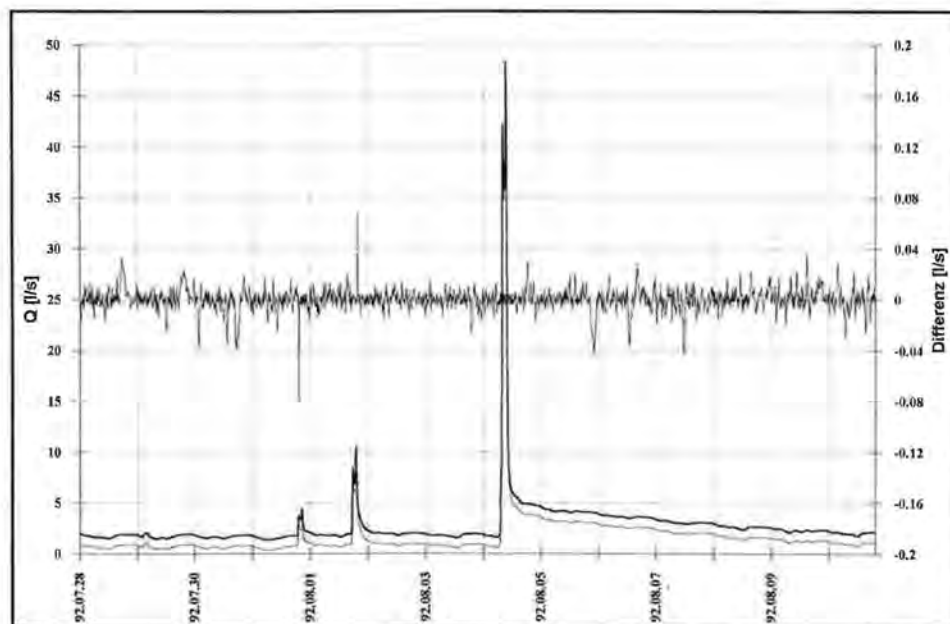


Fig. 8: Gemessene Ganglinie (dicke durchgezogene Linie), reduzierte und rückinterpolierte Ganglinie (strichpunktierte dünne Linie) des Oberflächenabflusses am Pegel Höhenhansl im Juli/August 1992 im Vergleich. Zusätzlich ist die Differenz zwischen den beiden Ganglinien (durchgezogene dünne Linie) dargestellt. Die Werte der reduzierten und rückinterpolierten Ganglinie sind zwecks der besseren Anschaulichkeit um 1 l/s vermindert.

6. Zusammenfassung und Wertung der Ergebnisse

Mit den vorgestellten Algorithmen und Programmen stehen Werkzeuge zur Verfügung, die es ermöglichen, „Datenfriedhöfe“, die sehr leicht bei der automationsunterstützten Erfassung von Meßdaten in Bereichen entstehen, in denen intensiv mit Zeitreihen gearbeitet wird (Hydrologie, Meteorologie, Bodenkunde etc.), zu vermeiden.

Durch die ereignisbezogene Reduktion der Daten können diese unter Steuerung des Benutzers soweit reduziert werden, daß einerseits die charakteristischen Kenndaten der Ganglinie erhalten bleiben, andererseits aber Datenmengen gefunden werden, die auch auf kleineren Datenbanksystemen im PC-Bereich noch verwaltbar sind, ohne entscheidende Einbrüche im Bereich der Datenqualität zu erleiden.

Durch die Möglichkeit der Spline-Interpolation mittels verschiedener Algorithmen können für die statistische Bearbeitung der Daten aus den reduzierten Datenreihen wieder gleichabständige Datenreihen und Zeitraster berechnet werden. Naturgemäß entstehen bei dieser Rückrechnung Fehler, die allerdings in sehr vielen Fällen vernachlässigbar gering sind, wenn vom Anwender größte Sorgfalt bei der Auswahl der Methoden getroffen wurde.

Von besonderer Bedeutung ist in diesem Zusammenhang der Meßwertausgleich, der elektrische Schwankungen aus der Meßwertaufnahme, die die Meßwerte innerhalb der Meßgenauigkeit der Sonden verändern können, eliminiert. Diese geringfügigen Schwankungen stellen für die Interpolation ein relativ großes Problem dar, da sich dadurch bei extrem geringem Schwanken des tatsächlichen Wertes der Gradient der Ganglinie sehr stark ändern kann (Fig. 1).

Die vorgestellten Beispiele zeigen allerdings auch die große Bedeutung der Auswahl der anzuwendenden Methoden und Programmparameter durch den Bearbeiter. Durch die Abhängigkeit der Ergebnisse der Datenreduktion von sehr vielen Faktoren, kommt der Gewichtung der Steuerungsparameter erhebliche Bedeutung zu. Während bei der Reduktion der Lufttemperaturdaten eine Kompression der Originaldaten auf etwa zwei Drittel gerade noch akzeptabel erscheint, liefert die Reduktion der Bodentemperaturdaten auf 18% der Ausgangsdaten nach der Rückinterpolation praktisch eine idente Datenreihe innerhalb der Meßgenauigkeit des eingesetzten Temperaturfühlers. Im Bereich der Registrierung des Abflusses von Oberflächengewässern konnten trotz einer Reduzierung der Meßdaten auf etwa ein Drittel der Fehler zwischen Originaldatenreihe und interpolierter Datenreihe sowohl im Zuge von Abflußereignissen als auch im Rahmen der Aufzeichnung von anthropogen bedingten Tagesschwankungen innerhalb der Meßgenauigkeit der eingesetzten Drucksonde gehalten werden.

Kurzfassung

Die verstärkte Verwendung von elektronischen Meßsystemen und automatischen Datenaufzeichnungssystemen im Bereich der Hydrogeologie bringen neben den Vorteilen einer hochauflösenden zeitlichen Messung verschiedenster Parameter aber auch Probleme im Bereich der Datenaufbereitung und -verwaltung.

Bedingt durch die Forderung nach einer detaillierten Registrierung der Parameter während kurzdauernder Ereignisse müssen die Daten in einem sehr engen Zeit-

intervall registriert werden. Aus Gründen einer effizienten Speicherverwaltung zeichnen die Registriereinrichtungen meist in einem konstanten Zeitraster auf, sodaß neben den Ereignissen auch jene Phasen, in denen die Vorgänge kaum Schwankungen unterworfen sind, in einem sehr dichten Raster registriert werden. Diese hohe Aufzeichnungsdichte bringt entscheidende Schwierigkeiten bei der Speicherung der Meßdaten in Datenbanksystemen, da diese mit einem hohen Overhead von nicht notwendigen Meßdaten belastet werden. Da die Messung der meisten Parameter über elektrische Spannung oder Widerstand erfolgt, sind die Meßergebnisse immer mit einem meßtechnischen Rauschen versehen, das bei der Datenauswertung und -interpretation störend wirkt, wenn die tatsächlichen Schwankungen der Parameter sehr gering sind.

In diesem Artikel soll eine mathematisch einfache Methode des Meßwertausgleichs vorgestellt werden. Weiters sollen Algorithmen präsentiert werden, die – unter Berücksichtigung von Ereignissen und Zeiten mit geringer Schwankungsbreite der Parameter – eine Kompression der Daten in Zeitreihen ermöglichen. Da jedoch für Auswertungen die Daten oft in einem konstanten Zeitraster benötigt werden, werden auch Verfahren vorgestellt, die durch Spline-Interpolation das ursprüngliche Meßintervall wiederherstellen können.

Die Effizienz der vorgestellten Algorithmen wird an drei Beispielen aus den Bereichen Meteorologie, Hydrologie und Bodenkunde gezeigt, wobei durch Kompression und darauffolgende Interpolation die Abweichungen der errechneten Werte von den gemessenen Werten dargestellt werden.

Literatur

- AKIMA, H. (1970): A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures. – *Journal of Assc. for Comp. Mach.*, Vol. 17, No. 4, 589–602.
- HILBERG, D. (1989): Akima-Interpolation. Noch besser als das Spline-Verfahren. – *c't* 6/89, 206–212.
- MEIER, W. & D. HARTMANN (1988): Grafische Darstellung von Meßwerten durch verallgemeinerte Splines. – *c't* 12/88, 138–145.
- BULIRSCH, R. (1968): *Mathematische Hilfsmittel des Ingenieurs*. – Berlin (Springer).
- REINISCH, C. H. (1967): *Smoothing by Spline Functions*. – *Numerische Mathematik*, Bd. 10, 177–183, Berlin (Springer).
- SPÄTH, H. (1986): *Spline-Algorithmen*. – München (Oldenburg).

Summary

The intensive use of electronic measuring systems and data loggers in hydrogeology has advantages in getting data from different parameters in very short intervals, but also brings along problems in processing and managing of data.

During short-time events values of different parameters have to be recorded in very short time intervals. Due to the efficiency of memory management most data loggers record in a fixed time interval. In periods of less oscillation the recording in very short time intervals is not necessary. The recording in fixed time intervals leads to problems with data management systems because of the big overhead of unnecessary data points. The recording of values by measuring electric voltage or resistance leads to an electric oscillation with high frequency disturbing the evaluation and interpretation of data.

In this paper simple methods and algorithms are presented, allowing the reduction of data points in time series without losing any precision. For calculating data points in the measured equivalent time interval, which is necessary for most statistical evaluations, methods of spline interpolation are used.

The efficiency of the algorithms is demonstrated by three examples from meteorology, hydrology and soil sciences. After reduction and reinterpolation of data points in time series the divergences between measured and calculated values are shown.