

Monte Carlo Wiggle Matching.

Zur statistischen Auswertung der mittelneolithischen ¹⁴C-Daten von Hasselsweiler 2, Inden 3, und Inden 1.

Einleitung

Während das Fundmaterial der altneolithischen Linearbandkeramik aufgrund gezielter Forschungen sich heute auf dem zeitlichen Niveau einzelner Keramikphasen und Hausgenerationen gliedern läßt was einer (relativen) Datierungsgenauigkeit von etwa ± 15 Jahren entspricht (STEHLI 1989), steht die Chronologie des nachfolgenden Mittelneolithikums auf vergleichsweise unsicherer Basis. Auch wenn bislang, jedenfalls in Süddeutschland, allgemein von einer zeitlichen Abfolge (im Sinne eines deutlichen Nacheinander) der Stufen Großgartach (GG), Planig-Friedberg (PF) und Rössen (RÖ) ausgegangen wurde, so bleibt die Stellung von Planig-Friedberg weiterhin unklar. Wird die Abfolge GG -> PF -> RÖ postuliert, und wird also PF als eine verbindende, eigenständige Keramikphase zwischen GG und RÖ angesehen, so steht man vor der Schwierigkeit, daß zwar tatsächlich eine Siedlungskontinuität zwischen GG und PF beobachtet wird, sich aber zwischen PF und RÖ ein Bruch abzeichnet, jedenfalls in Hessen (EISENHAUER 1994). So scheint einerseits die Keramikentwicklung GG -> PF nahtlos zu sein, und man könnte somit PF zeitlich an GG anhängen. Andererseits gibt es aber in Baden-Württemberg die Fundplätze Forchheim und Weisweil, mit allen Keramikmerkmalen, die für GG und RÖ charakteristisch erscheinen, aber ohne die erwarteten Merkmale der Zwischenstufe PF.

Wie nun Biermann in der vorliegenden Arbeit ausführt, läßt sich PF stilistisch zwanglos an GG anhängen, so daß er sich berechtigt fühlt, die Stufe PF als eine Variante GG III der Großgartacher Kultur zu interpretieren. Dafür spricht, neben den Stilkriterien, daß die Befunde mit Keramik des Typs PF sich auch mit der Hilfe einer Keramikseriation nicht vom späten GG abtrennen lassen. Auf dieser Grundlage postuliert Biermann nun eine Abfolge GG I -> GG II -> GG III. So stellt sich schließlich sogar die Frage, ob nicht in gleicher Region - und Biermann führt hierfür die Fundorte Hasselsweiler, Inden 3 und Inden 1 im Rheinland an - die "Keramikulturen" Großgartach und Rössen zumindest teilweise zeitlich parallel verlaufen könnten.

Dieser Kernhypothese von Biermann soll im folgenden auf der Grundlage verschiedener Experimente zur dendrochronologischen Kalibration der verfügbaren ¹⁴C-Daten weiter nachgegangen werden.

Die Kernhypothese von Biermann

Nach Abschluß seiner Arbeit hat mir Biermann die Frage vorgelegt, ob ich, mit Hilfe einer weiteren statistischen Auswertung der ¹⁴C-Daten und dabei möglichst unabhängig von den Ergebnissen seiner Keramikanalysen, ebenfalls zum Ergebnis käme, daß die etwa 15 Kilometer auseinander liegenden Siedlungen Hasselsweiler 2 (mit Funden der Stufe GG II/III) und Inden 1/3 (Rössen) zur gleichen Zeit bestanden haben könnten. Um dieser Frage nachgehen zu können, habe ich zunächst zu jeder Probe alle archäologischen Informationen zusammengestellt, wie sie in den Probendokumenten der im Kölner ¹⁴C-Laboratorium gemessenen Daten der Fundorte Hasselsweiler, Inden 3 und Inden 1 beschrieben sind. Zu den verfügbaren Probeninformationen gehören Kopien der Grabungsdokumente, kommentierte Profilskizzen, kurze Beschreibungen der Fundumstände, sowie der Bodenbildung und der archäologischen Schichten. Insofern stratigraphische Beziehungen zwischen den datierten Befunden und anderer Befunde des gleichen Fundorts vorliegen, wurden diese ebenfalls dokumentiert, aber nicht weiter ausgewertet. Biermann hat die Dauer der betreffenden Siedlungen wie folgt eingeschätzt:

Hasselsweiler 2: 1-2 Hausgenerationen,
ca. 20-40 Jahre, max. 50 Jahre

Inden 1: 1-2 Hausgenerationen,
ca. 20-40 Jahre, max. 50 Jahre

Inden 3: 4-5 Hausgenerationen,
ca. 80-100 Jahre, max. 150 Jahre

Die Schätzungen zur Dauer der Siedlungen werden unten zur Einengung der kalibrierten Wahrscheinlichkeitsverteilungen eingesetzt. Zum

Vergleich: Für den Siedlungsplatz Hambach 260 (Jülich-Welldorf), Lkr. Düren, der ca. GG IIc/III-zeitlich sein dürfte, schätzt Dohrn-Ihmig eine Besiedlungsdauer von 3 Hausgenerationen (bei insgesamt 11 Hausbefunden) oder 120 Jahre (DOHRN-IHMIG 1983 b, 267), was zu lang erscheint.

Methoden der Auswertung von archäologischen ^{14}C -Daten

Einleitend werden verschiedene Probleme der dendrochronologischen Kalibration von archäologischen ^{14}C -Daten geschildert. Es werden die Gründe dargelegt, warum die bisher existierenden Verfahren, so leistungsfähig sie im einzelnen schon sein mögen, sicherlich noch nicht das letzte Wort darstellen. Allgemein ist zu fordern, daß eine Auswertungsmethode für ^{14}C -Daten der großen Vielfalt der archäologischen Funde und Befunde gerecht wird, andererseits aber auch theoretischen Modellen angepaßt werden kann (zum Beispiel der Siedlungs- oder Kulturentwicklung, die zur Beschreibung und Erklärung der Befunde dienen). Am nützlichsten scheinen dabei Verfahren zu sein, die nicht nur statistische Maßzahlen zur Altersbestimmung liefern, wie zum Beispiel die Balken-Methode, sondern zugleich auch eine Plattform für explorative Hypothesen darstellen, sich also zum Testen von archäologischen Modellen eignen. Die herkömmlichen Auswertungsverfahren für ^{14}C -Daten genügen diesen Forderungen zur Zeit nur unvollkommen, was sicherlich unter anderem daran liegt, daß die Methodenentwicklung eng an der Formulierung geeigneter Rechenprogramme gebunden ist, und die dendrochronologische Kalibration von archäologischen ^{14}C -Daten recht aufwendige Algorithmen voraussetzt. Vermutlich sind es aber auch die Grenzen der archäologischen Erkenntnismöglichkeiten, welche zugleich die Grenzen zum Ausdenken geeigneter Verfahren zur Auswertung von ^{14}C -Daten bilden. Wir wollen diese Grenzen nun so weit stecken, wie es irgend geht.

Bayessche Kalibration

Wie zu erkennen ist, verläuft die Forschung bei der Entwicklung von Auswertungsmethoden für ^{14}C -Daten zur Zeit in verschiedene Richtungen. Im englischsprachigen Raum wird schon seit einigen Jahren viel Mühe in die Ausarbeitung von

Anwendungen der Bayesschen Wahrscheinlichkeitstheorie investiert (BUCK, LITTON, & SMITH 1992; BUCK, LITTON & SCOTT 1994). Kennzeichen der Bayesschen Kalibration ist die Berechnung von bedingten Wahrscheinlichkeiten. Im einfachsten Fall werden bei der Bayesschen Kalibration Modelle entwickelt, um verfügbare stratigraphische Informationen der Form "älter", "jünger" oder "zeitgleich" zur Präzisierung der ^{14}C -Chronologie nutzbar zu machen. Dabei werden die in aller Regel zunächst recht breiten Sicherheitsintervalle der kalibrierten Daten mit Hilfe von zusätzlichen, archäologischen Erkenntnissen eingeengt. Im einfachsten Fall vergleicht man zu diesem Zweck die Überlappungsbereiche der kalibrierten Wahrscheinlichkeitsverteilungen einzelner ^{14}C -Daten mit dem als bekannt vorausgesetzten zeitlichen Nacheinander der datierten Proben. Damit lassen sich die Wahrscheinlichkeiten gewisser Altersbereiche, die in einem ersten Rechenschritt zunächst nur grob anhand der Einzeldaten ermittelt wurden, in einem zweiten Schritt mit Hilfe der vorgegebenen archäologischen Informationen weiter eingengen. Die archäologischen Informationen dienen dabei zur Entzerrung bzw. zur Änderung der Form der ursprünglichen Wahrscheinlichkeitsverteilung. Wegen des nichtlinearen Verlaufs der Kalibrationskurve, und damit zusammenhängend der Existenz multipler Ablesemöglichkeiten der meisten ^{14}C -Daten, gelingt diese Entzerrung allerdings zumeist nur unvollständig.

Archäologisches Wiggle Matching

Neben der Verknüpfung der ^{14}C -Daten mit Rangskalierten Probeneigenschaften ("älter" oder "jünger") lassen sich die Daten auch mit archäologischen Informationen höherer Skalenwertigkeit verbinden. Besonders interessant ist die Einbeziehung von Intervall-skalierten stratigraphischen Informationen, mit deren Hilfe (addierbare) Zeitdifferenzen erfaßt werden. Um ein Beispiel zu nennen, könnte eine Probe mit einer gewissen (archäologischen) Sicherheit nicht nur "älter als", sondern vielmehr "20 ± 10 Jahre älter als" eine andere Probe sein. Die rechnerische Berücksichtigung derartiger Bedingungen führt hin zur Methode des "Archäologischen Wiggle Matching" (AWM). Desse Kennzeichen ist die Einpassung von Intervall-skalierten Datenserien in die Kalibrationskurve. Aus mathematischer Sicht läßt sich das AWM als eine Variante der Bayesschen Kalibration charakterisieren. Beim AWM wird die Betonung aber

weniger auf die statistischen Eigenschaften der einzelnen Daten, als vielmehr auf die archäologischen Eigenschaften der datierten Proben gelegt. Insofern nun beide Methoden je nach Informationsvorgabe (bzw. Eingabe-Qualität) unterschiedliche Voraussetzungen haben, lassen sie sich nicht nur terminologisch sinnvoll unterscheiden, sondern ergänzen sich auch in ihren Anwendungsbereichen.

Die mathematischen Grundlagen des Wiggle Matching werden in einer Reihe von Arbeiten schon ausführlich behandelt (PEARSON 1986), und wir können uns an dieser Stelle kurz fassen. Beim Wiggle Matching wird mit geeigneten statistischen Testverfahren (z.B. Methode der kleinsten Fehlerquadrate, Chi-Quadrat-Test, t-Test) die beste, gleichzeitige Einpassung aller Daten eines zeitlich strukturierten Datensatzes in die Kalibrationskurve ermittelt. Dabei wird die zeitliche Struktur des Datensatzes nach archäologischen Kriterien vorgegeben. Der Datensatz wird dann sukzessive, in festen Schritten, parallel zur Kalenderzeitskala verschoben. In jedem gerade erreichten Kalenderjahr wird die Summe der Abstände der einzelnen Daten zur Kalibrationskurve ermittelt, gewichtet für variierende Standardabweichungen. Als laufendes Ergebnis erhält man nach diesem Verfahren für jedes Kalenderalter einen mittleren Abstand der Daten zur Kalibrationskurve, in Richtung der ^{14}C -Skala gemessen. Schließlich wird als "beste" Einpassung dasjenige Kalenderjahr ermittelt, bei welchem der gemeinsame Abstand der archäologischen Daten zur Kurve ein Minimum erreicht. Zu diesem Endjahr existiert nun ein Sicherheitsintervall, dessen genaue Berechnung aufgrund der zahlreichen statistischen und archäologischen Parameter, die es zu berücksichtigen gilt, allerdings noch gewisse Schwierigkeiten bereitet. Ein weiteres Problem, jedenfalls was die Interpretation der Ergebnisse angeht, besteht darin, daß die herkömmlichen statistischen Tests immer nur ein Ablehnungsintervall, niemals aber ein Akzeptanzintervall zu definieren erlauben. Das Ablehnungsintervall wird um so schmaler, je schlechter die Einpassung ist (weil es nämlich um so einfacher wird, eine Hypothese zu falsifizieren, je schlechter die Hypothese zu den empirischen Daten paßt). Hinsichtlich der genauen Berechnung des kalenderzeitlichen Datierungsfehlers bestehen insofern beim AWM grundlegende, wenn auch sicher lösbare methodische Schwierigkeiten (die vor allem auf der archäologischen Ebene zu suchen sind). Aus pragmatischer Sicht genügt es aber, jedenfalls zur Orientierung, als einfaches Maß zur Beschreibung der Güte des

AWM-Ergebnisses vorerst die mittlere Standardabweichung der archäologischen Daten, geteilt durch die Wurzel aus der Datenzahl N , zu verwenden. Dieses Maß hat den Vorteil, langsam (mit $N^{1/2}$) um so kleiner zu werden, je mehr und je genauere Daten vorliegen. Mit zunehmender Datenzahl N erhält man damit eine Fehlerverkleinerung, welche sich auch als "Gewinn an chronologischer Präzision" oder auch als "Informationsgewinn" veranschaulichen läßt.

Anhand von weiteren Überlegungen zum Informationsgewinn, den man bei der Kalibration von ^{14}C -Datenserien im Kontext von archäologischen Informationen erzielen kann, lassen sich die verschiedenen Anwendungsbereiche des AWM wie folgt nach Gütekriterien unterteilen.

1. AWM und Dendrochronologie

Den größten Informationsgewinn bietet das Wiggle Matching (wie auch die Bayessche Kalibration) für ^{14}C -Datenserien an Holzproben mit möglichst vielen erhaltenen Jahrringen. Nachdem die Kurve der Jahrringbreiten gemessen (und für spätere Kontrollen aufgehoben) ist, wird durch mechanische Aufteilung des Holzstücks eine Serie von Jahrringblöcken hergestellt. Die Probenserie ist nun in idealer Weise nach Zeitintervallen gegliedert, denn die zeitlichen Abstände zwischen den Proben sind im Prinzip fehlerfrei bestimmbar. Ferner ist auch die zeitliche Richtung der Serie von "alt" zu "jung" bekannt. Sind die ^{14}C -Messungen abgeschlossen, so wird die Datenserie nach dem oben beschriebenen Verfahren kalibriert. Sofern das Splintholz erhalten ist, wird das berechnete Endjahr mit dem Kalenderjahr der Fällung des Baumes identisch sein. Wenn nicht, sind Schätzungen der Anzahl der fehlenden Jahrringe erforderlich. Weil nun bei historischen Bauten der Zeitpunkt der Fällung der Hölzer in aller Regel mit dem Baujahr übereinstimmt, ist damit auch das interessierende archäologische Ereignis, nämlich das Baujahr des betreffenden Gebäudes einigermaßen wohldefiniert.

2. AWM und historische Chronologien

Die Methode des Wiggle Matching kann mit guten Ergebnissen auch bei ^{14}C -Datenserien angewandt werden, bei denen die erforderlichen zeitlichen Abstände der Proben anhand von historischen oder

astronomischen Überlieferungen hergeleitet werden. Anwendungen liegen für die ägyptische historische Chronologie vor (WENINGER 1986, 1990, 1992), wie auch andere Chronologien des Vorderen Orients und der Mittelmeerländer ab etwa 3000 v. Chr. in Frage kommen.

3. AWM und Tellstratigraphien

Auch bei Tellstratigraphien lassen sich die erforderlichen zeitlichen Intervalle anhand von archäologischen Kriterien aufstellen. Beispiele sind Schätzungen der zeitlichen Tiefe verschiedener Schichten eines Tells und die Dauer aufeinanderfolgender Bauphasen. Problematisch bleibt dabei allerdings immer die Möglichkeit, daß Siedlungsunterbrechungen oder Hiaten nicht erkannt werden, oder aber einzelne Proben sich nicht in primärer Lage befinden. Bei Tellstratigraphien mag die Bestimmung geeigneter Abstandshypothesen zuweilen schwierig sein, doch bietet das AWM gegenüber der Einzelkalibration der Daten immerhin den Vorteil, unterschiedliche Hypothesen vergleichen und damit auch testen zu können. Man erhält in jedem Fall statistische Aussagen darüber, wie gut die Hypothesen sich in den Daten widerspiegeln. Damit lassen sich die Hypothesen weiter einengen. Die Mindestinformation, die man beim AWM von stratifizierten Proben aus Tellsiedlungen erhält, ist eine Aussage über das Vorliegen von Ausreißerdaten oder -proben. Alleine schon diese Minimalinformation kann durchaus nützlich sein, denn aufgrund der zahlreichen Schwingungen der Kalibrationskurve, mit Amplituden zwischen 20 und 300 BP (¹⁴C-Jahre) und zeitlichen Abständen zwischen 20 und 400 Kalenderjahren, ist es mit herkömmlichen statistischen Testverfahren erfahrungsgemäß kaum möglich, zwischen stratigraphischen Inversionen der Proben und Inversionen der Daten auf der ¹⁴C-Skala zu unterscheiden. Fallstudien liegen für das bulgarische Neolithikum (WENINGER 1992) und für die Frühbronzezeit in Nordwestanatolien vor (WENINGER 1995).

4. AWM und Keramiksequenzen

Als weiteres Anwendungsgebiet des Wiggle Matching in der Archäologie ist schließlich noch die Bestimmung von Zeitintervallen mit Hilfe von Keramikseriationen zu nennen. Besonders bewährt

hat sich die Methode der Korrespondenzanalyse, mit deren Hilfe die Abfolge der datierten Befunde als Schwerpunktkoordinaten auf einer linearen Skala in einem höher dimensional Faktorenräum, also Intervall-skaliert dargestellt werden können. Auch bei dieser Anwendung ist eine unabhängige Skaleneichung nach Kalenderjahren erforderlich, die man experimentell durch lineare Dehnung und Streckung der Keramiksequenz und damit verknüpft der ¹⁴C-datierten Proben erhält (STEHLI 1989; EASTON AND WENINGER 1993; WENINGER 1995). In zukünftigen Studien wird weiter darauf zu achten sein, wieweit die Beziehung zwischen den Faktorenskalen und der Kalenderzeitskala tatsächlich linear ist, wie für das AWM gefordert werden muß (JABLONKA 1995).

Zur Schätzung von Zeitintervallen in der Archäologie

Manning (1995) hat in einer detailreichen Monographie zur Chronologie der Ägäischen Frühbronzezeit zahlreiche Anwendungen des AWM zusammengestellt und dessen Probleme und Möglichkeiten ausführlich dargestellt. Als Ergebnis der inzwischen durchaus umfangreichen Menge an AWM-Studien kann man zusammenfassend folgende Erkenntnisse herleiten, die Probleme der Schätzung von Zeitintervallen in der Archäologie betreffen:

Auch in dem extrem "schlechten" Fall, wenn überhaupt keine brauchbaren (quantifizierbaren) stratigraphischen Informationen zur zeitlichen Reihenfolge der Proben vorliegen, kann das Verfahren des AWM trotzdem noch gewisse chronologische Informationen bereitstellen. Das gilt jedenfalls solange, wie überhaupt Schätzungen zur ungefähren Dauer einer archäologischen Siedlung (oder Schicht) möglich sind. In diesem Fall bleibt die zeitliche Reihenfolge der datierten Proben gewiß unsicher, doch kann aller Erfahrung nach zumindest ausgesagt werden, daß die Proben einem Zeitintervall gewisser Länge, zum Beispiel mindestens 20 Jahre und höchstens 80 Jahre, entstammen. Ein Informationsgewinn ist aus physikalischer Sicht übrigens nur möglich, wenn man einem System von außen Energie zuführt. Aus dieser Sichtweise muß man beim Archäologischen Wiggle Matching "etwas dafür tun", um präzisere Datierungsergebnisse zu erzielen. Insofern sind beim AWM auch zusätzliche Fehlerquellen zu berücksichtigen. Bevor das Verfahren des Monte Carlo Wiggle Matching näher

beschrieben wird (siehe unten), soll an dieser Stelle zunächst eine Begründung dafür vorgelegt werden, warum wir "vor lauter Verzweiflung" den Weg gehen, die Probenreihenfolge mit Hilfe eines Zufallsgenerators festzulegen.

2-Dimensionale Dispersionskalibration

Zu Beginn einer jeden Auswertung von ^{14}C -Daten scheint es ebenso selbstverständlich wie auch fast schon trivial zu sein, daß umfangreiche Datenserien zu genaueren chronologischen Ergebnissen führen müßten, als Einzeldaten. "Ein Datum ist kein Datum", so pflegt man zu formulieren, und das "Gesetz der großen Zahl" biete die Gewähr dafür, daß umfangreiche Datensätze eine Präzisierung der chronologischen Ergebnisse erlauben. Ein bewährtes Verfahren zur Kalibration von umfangreicheren ^{14}C -Datengruppen ist nun die Methode der 2-D-Dispersionskalibration, mit deren Hilfe die Wahrscheinlichkeitsverteilung der Daten von der ^{14}C -Skala auf die Zeitskala des julianischen Kalenders bzw. der Dendrochronologie übertragen wird. Die Voraussetzungen der Anwendung dieser Methode sind vergleichsweise wenig streng. Erforderlich ist lediglich eine Zusammenstellung der ^{14}C -Daten nach archäologischen Begriffen, zum Beispiel Tellschichten, Kulturstufen oder Siedlungsphasen (JAGUTTIS-EMDEN 1977; PAPE 1979; BREUNIG 1987). Zur Durchführung der Dispersionskalibration werden die Daten entsprechend dem interessierenden Begriff als Tabelle zusammengestellt und als Liste mit Überschrift in den Rechner eingegeben. Die Reihenfolge der eingegebenen Daten ist beliebig, und wird auch nicht weiter ausgewertet. Alles weitere regelt der Kalibrations-Algorithmus. Das Ergebnis ist eine graphische Darstellung der Verteilung der kalibrierten Daten, geplottet im Zusammenhang mit statistischen Aussagen zur Lage der Verteilungsmitte und -breite.

Antinomie der Deutung von ^{14}C -Wahrscheinlichkeitsverteilungen

Bei der Kalibration von archäologischen ^{14}C -Daten stößt man aller Erfahrung nach relativ schnell, etwas Übung im Umgang mit den Daten vorausgesetzt, gerade bei großen Datenserien auf die Frage, wo denn nun die in der großen Datenserie enthaltene, gegenüber der Einzelmessung doch vermeintlich verfeinerte chronologische Information geblieben ist. Im Datierungsergebnis,

jedenfalls, ob graphisch oder numerisch ausgedrückt, drückt sich der Informationsgewinn nicht aus. Das Problem hat J.H. Ottaway (1986) sehr anschaulich formuliert, und ich zitiere:

"The theories of probability that underlie the statistical treatment of experimental observations were developed largely from games theory, and they accept that every so often a highly improbable event will occur....In calibration methods that convert each central value (i.e. each individual date) into a range, there is no means of reestablishing a central tendency. Consequently, as an increase in the number of replicates is, if probability theory holds, sooner or later going to throw up an extreme value, it is inevitable that an *increase* in the number of observations leads to a *decrease* in the precision with which the extremes of the individual ranges, the only remaining numerical values, can be quoted, even for a single site.

As one who regards archaeology as a branch of biology with historical overtones, I find this situation *absurd*. It is unique in any province of biology, or for that matter in any physical science, and I would not tolerate it in respect of any data for which I was responsible."

Somit wird eine Vergrößerung der Datenbasis (Anzahl N Daten) schließlich zu einer (scheinbaren) Verschlechterung der erreichten Datierungsgenauigkeit führen. Die Verteilung der Daten (ob kalibriert oder nicht) wird in der Theorie ja nur immer breiter, je mehr Daten vorliegen. Das theoretische Argument dafür ist, folgt man den Ausführungen von J.H. Ottaway, daß mit zunehmendem N eben irgendwann ein Extremwert auftreten wird, und zwar mit absoluter Sicherheit.

Nun vermute ich, daß J.H. Ottaway als Beispiel für seine Überlegungen unter anderem die Stapelung der statistischen Sicherheitsintervalle von vielen, einzelnen Daten vor Augen hatte. Das von den Daten überdeckte chronologische Intervall wird mit zunehmender Datenzahl tatsächlich nur immer breiter, und niemals enger. Das ist in der Tat eine

"absurde" Situation, aber das Problem scheint sogar noch schlimmer zu sein, als es Ottaway schon sieht. Man benötigt nämlich nicht einmal den ab und zu sicherlich zu erwartenden Extremwert, damit die Datenstreuung mit zunehmendem N breiter wird. Die Verteilung kann auch ohne Extremwerte immer nur mit N breiter werden. Zur Illustration dieser Vermutung benötigen wir nicht einmal sehr viele Daten; dazu genügen schon zwei. Liegt nämlich das erste Datum schon vor, und wird damit ein Sicherheitsintervall gewisser Länge auf der ^{14}C -Skala (oder auf der kalibrierten Zeitskala) definiert, so wird das zweite Datum entweder links oder rechts daneben liegen. Das Sicherheitsintervall ist schon im Übergang von $N=1$ zu $N=2$ breiter geworden. Davon ist natürlich auch die Breite der ^{14}C - wie auch der kalibrierten Wahrscheinlichkeitsverteilung betroffen, so daß das Problem nicht nur bei der Stapelung einzelner Intervalle, wie auch bei der Dispersionskalibration auftritt, sondern bei jeder anderen, vergleichbaren Methode der Kalibration von nicht verknüpften ^{14}C -Daten.

Das Problem besteht darin, um es nun präziser zu formulieren, daß man weder durch Stapelung der Sicherheitsintervalle, noch durch die Addition der Gauß'schen Glockenkurven der einzelnen Daten zu einem Gesamthistogramm, einen Zugang zu der mit zunehmendem N sicher vorhandenen Verbesserung der Datierungsgenauigkeit hat. Damit steht man weiterhin vor dem Rätsel, warum eine Vergrößerung der Datenbasis zu einer Verschlechterung der Ergebnisse führt, was J.H. Ottaway doch sehr zu recht bemängelt

Gewichtete Mittelwerte von archäologischen ^{14}C -Daten

Zur Auflösung des Rätsels könnte man sich zunächst überlegen, daß die gewünschte Verkleinerung des statistischen Intervalls sich dadurch erzielen läßt, indem einfach nur das gewichtete Mittel aller Daten berechnet wird. Die dazu benötigten Formeln haben Ward und Wilson (1978) hergeleitet. Gegeben seien n Daten mit Mittelwerten μ_i , $i = 1, \dots, n$ und Standardabweichungen σ_i . Der gewichtete Mittelwert $\langle \mu \rangle$ wird nach folgender Formel berechnet:

$$\langle \mu \rangle = \left(\sum_1^n \mu_i / \sigma_i^2 \right) / \left(\sum_1^n 1 / \sigma_i^2 \right)$$

Die gewichtete Standardabweichung $\langle \sigma \rangle$ hat den Wert:

$$\langle \sigma \rangle = \left(\sum_1^n 1 / \sigma_i^2 \right)^{-1}$$

Allerdings ist diese Mittelwertbildung der ^{14}C -Daten strikt nur unter der Voraussetzung erlaubt, daß es sich dabei um Proben mit gleichem ^{14}C -Gehalt handelt. Um dieser Voraussetzung zu genügen, reicht ein statistischer Test der Datenidentität wohlgerne nicht aus. Die Daten müssen zwar im Hinblick auf die Identität ihrer Mittelwerte getestet worden sein, um abzusichern, daß deren Streuung nur aufgrund der Zufälligkeiten des radioaktiven Zerfalls entstanden ist, oder aber andere, ebenfalls in den Standardabweichungen ausgedrückte Zufallsfehler (apparative Schwankungen) vorlagen. Die dann aus statistischer Sicht zulässige Mittelwertbildung setzt ebenso streng weiter voraus, daß schon vor der Messung (*a priori*) bekannt war, daß alle Proben den gleichen ^{14}C -Gehalt besitzen. Die Mittelwertbildung mit Fehlerverkleinerung nach den obigen Formeln kann damit, in aller Regel, nur für Interlaborvergleichsmessungen an homogenisierten Proben in Anspruch genommen werden, nicht aber für typische archäologische Proben. Das heißt konkret, auch wenn mehrere Proben ein und der selben archäologischen Schicht entstammen, und damit archäologisch aller Voraussicht nach "gleichalt" sein sollten, daß man trotzdem die Mittelwertbildung mit Fehlerverkleinerung nicht in Anspruch nehmen kann. Es bleibt eben (*a posteriori*) ungewiß, ob die Proben tatsächlich jahrgenau gleichalt sind.

Nun könnte man an dieser Stelle natürlich liberal verfahren, und die Anforderungen an die Definition der Gleichaltrigkeit zweier Proben von 1 auf 20 Kalenderjahre reduzieren, entsprechend der Anzahl der Jahrringe typischer Holzproben der dendrochronologischen Kalibrationskurve. Trotzdem gäbe es bei realen archäologischen Proben, auch wenn diese archäologisch "gleichalt" sein sollten, niemals eine Gewähr dafür, daß sie hinreichend gleichalt sind, um den Voraussetzungen der Mittelwertbildung von Ward und Wilson (1978) zu genügen.

Als Beispiel dafür, was in der Praxis passieren kann, wenn eine unzulässige Mittelwertbildung

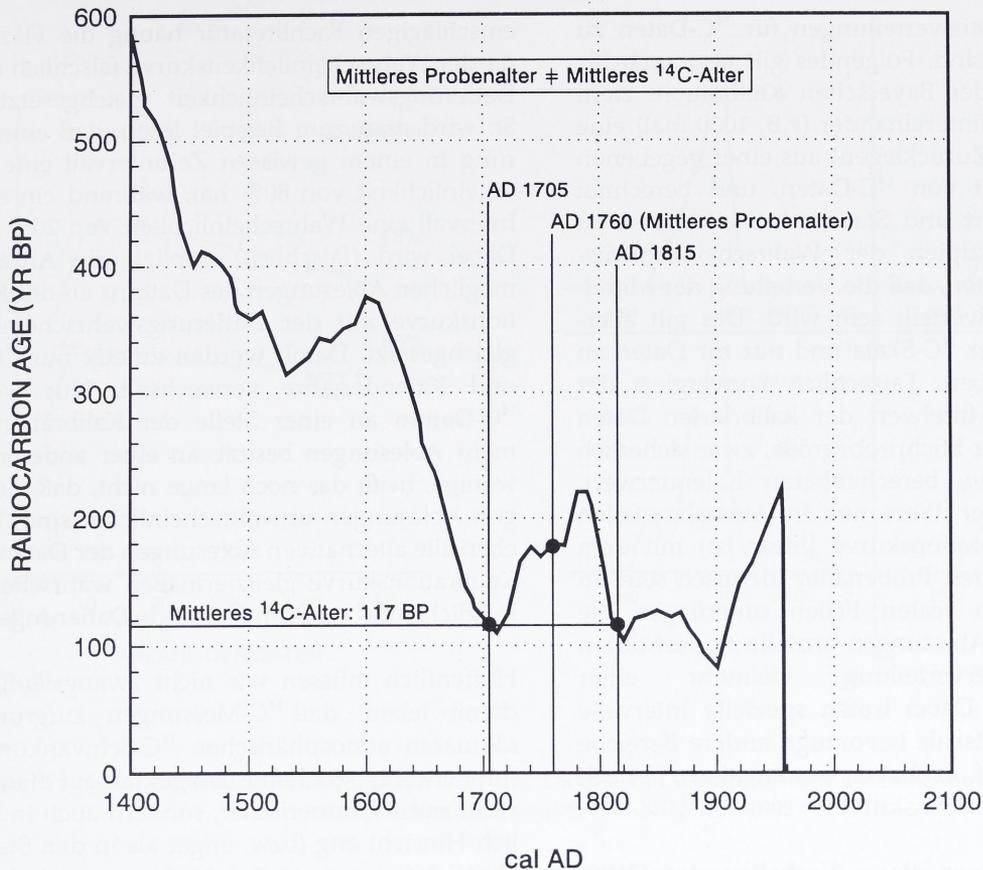


Abb.1 Ausschnitt aus der dekadischen Kalibrationskurve von Stuiver und Becker (1993)

zweier statistisch identischer ¹⁴C-Daten durchgeführt wird, betrachten wir zwei Proben des Kalenderalters AD 1705 und AD 1815 (Abb. 1).

Das mittlere Kalenderalter beider Proben ist $(AD\ 1705 + AD\ 1815) / 2 = AD\ 1760$. Die entsprechenden ¹⁴C-Daten lauten $116 \pm 4\ BP$ und $118 \pm 6\ BP$ (STUIVER UND BECKER 1993, Fig. 2A, und Table 1, S. 57). Die Daten besitzen, nach den obigen Formeln berechnet, einen gewichteten Mittelwert $116 \pm 3\ BP$. Tatsächlich haben Proben mit Alter AD 1760 einen völlig anderen ¹⁴C-Wert, nämlich $170 \pm 4\ BP$ (Abb. 2).

Bei realen archäologischen ¹⁴C-Daten (und das Prädikat real bezieht sich hierbei auf die Probeneigenschaften) wird die Mittelwertbildung mit Fehlerverkleinerung demnach in aller Regel zu falschen Ergebnissen führen. Wie das Beispiel in Abb. 1 weiter zeigt, spielt es aufgrund der vergleichsweise starken Schwingungen der Kalibrationskurve eine gewichtige Rolle, ob Mittelwerte von Proben (auf

der Dendroskala) oder aber Daten (auf der ¹⁴C-Skala) bestimmt werden. Die Ergebnisse unterscheiden sich durchaus beträchtlich.

"Quantisierung" bzw. "Lock-In Effekte" bei kalibrierten ¹⁴C-Daten

Mathematisch ausgedrückt führt die Nichtlinearität der Kalibrationskurve dazu, daß die ¹⁴C-Daten in ihrem statistischem Verhalten nicht der gewohnten "Logik" (WENINGER 1986), sondern einer nicht-kommutativen Algebra folgen (WENINGER 1992). Für zwei Proben mit ¹⁴C-Gehalt A BP und B BP gilt $cal(A+B)$ ungleich $cal(A) + cal(B)$. Das Problem ist symmetrisch auf beiden Skalen: das aus den Axiomen der klassischen Wahrscheinlichkeitstheorie hergeleitete Gesetz der großen Zahl gilt weder für konventionelle ¹⁴C-Daten noch für kalibrierte Daten, jedenfalls nicht, wenn man näher hinschaut. Das hat gewichtige Konsequenzen, die bei jeder Methode der Berechnung von

Wahrscheinlichkeitsverteilungen für ^{14}C -Daten zu berücksichtigen sind. Folgendes gilt also auch für die Ergebnisse der Bayesschen Kalibration: zieht man mehrfach hintereinander (z.B. 1000 mal) eine Stichprobe (mit Zurücklegen) aus einer gegebenen Grundgesamtheit von ^{14}C -Daten, und berechnet dessen Mittelwert und Standardverteilung, so ist nach den Prinzipien der Wahrscheinlichkeitstheorie zu erwarten, daß die Verteilung der Mittelwertsdaten gaußverteilt sein wird. Das gilt allerdings nur auf der ^{14}C -Skala und nur für Daten an identischen Proben. Tatsächlich konvergiert der entsprechende Mittelwert der kalibrierten Daten bei zunehmender Stichprobengröße, zwar sicherlich gegen *irgendeinen*, berechenbaren Kalenderwert. Doch wird dieser Wert nur in Ausnahmefällen (wenn die Kalibrationskurve linear ist) mit dem korrekten, mittleren Probenalter identisch sein. In den allermeisten realen Fällen unterliegen die Daten und ihre Ablesungen anstelle der erhofften Gauß'schen-Fehlerverteilung vielmehr einer "Quantisierung". Dabei treten spezielle Intervalle der Kalenderzeitskala bevorzugt, andere Bereiche benachteiligt auf, und zwar systematisch, je nach Verlauf der Kalibrationskurve.

Wie Bernd Kromer dieses Verhalten der Daten anschaulich zu charakterisieren pflegt (pers. Mitt. 1996), unterliegen die Daten und ihre Ablesungen einem "Lock-In" Effekt: sie haken sich bevorzugt in die "wiggles" bzw. flachen Bereiche der Kalibrationskurve ein. Damit ist das Verhalten der Daten insofern gut charakterisiert, als spezielle Bereiche der ^{14}C -Skala, wie auch der Kalenderzeitskala, aufgrund der multiplen Ablesemöglichkeiten der Daten nicht gleichmäßig, sondern unterschiedlich oft angetroffen werden. Davon ist auch die Amplitude beider Wahrscheinlichkeitsverteilungen betroffen, welche ebenfalls im Rhythmus der "wiggles" zu- und abnimmt. Die Fläche unter den Kurven verhält sich insofern auch nicht linear zur Wahrscheinlichkeit. Man sollte deswegen, anstelle des Begriffs "kalibrierte Datierungswahrscheinlichkeit" vielleicht besser von einer "Pseudo-Wahrscheinlichkeit" der kalibrierten Daten sprechen. Wenn wir das nicht tun, so nur aus sprachlichen Gründen. Jedenfalls ist die Fläche unter der Kurve der kalibrierten Altersablesungen, entgegen den Voraussetzungen der klassischen Wahrscheinlichkeitstheorie, kein gutes Maß der Datierungswahrscheinlichkeit.

An dieser Stelle ist vielleicht auch die weitere Anmerkung nützlich, daß sogar in der

einschlägigen Fachliteratur häufig die Fläche unter der Wahrscheinlichkeitskurve fälschlich mit der Datierungswahrscheinlichkeit gleichgesetzt wird. So wird man zum Beispiel lesen, daß eine Datierung in einem gewissen Zeitintervall eine Wahrscheinlichkeit von 80% hat, während ein anderes Intervall eine Wahrscheinlichkeit von 20% besitzt. Dabei wird (fälschlich) implizit die Anzahl der möglichen Ablesungen des Datums an der Kalibrationskurve mit der Datierungswahrscheinlichkeit gleichgesetzt. Damit werden zuletzt nur ^{14}C -Jahre und Kalenderjahre verwechselt. Nur weil ein ^{14}C -Datum an einer Stelle der Kalibrationskurve mehr Ablesungen besitzt, an einer anderen Stelle wenige, heißt das noch lange nicht, daß die wenigen Ablesungen unwahrscheinlicher sind. Es sind eben alle alternativen Ablesungen der Daten an der Kalibrationskurve gleichermaßen wahrscheinliche, mögliche und ernstzunehmende Datierungen.

Hoffentlich müssen wir nicht zwangsläufig ewig damit leben, daß ^{14}C -Messungen aufgrund der säkularen atmosphärischen ^{14}C -Schwankungen in ihrer chronologischen Aussagekraft auf diese Weise nicht nur in numerischer, sondern auch in begrifflicher Hinsicht eng (bzw. enger als in den Standardabweichungen ausgedrückt) begrenzt sind. Die in den Standardabweichungen der ^{14}C -Daten, und vor allem die in längeren Datenserien enthaltenen chronologischen Informationen sollten zugänglich gemacht werden, jedenfalls ist das aufgrund der großen Bedeutung der ^{14}C -Methode für zahlreiche Disziplinen zu fordern, und der Grad, wieweit dies gelingt, hängt empfindlich von der Auswertungsmethode ab.

Resampling - eine Studie mit archäologischen ^{14}C -Daten

Auch aus einem wiederum anderen Blickwinkel heraus betrachtet, wird die störende Realität des von J.H. Ottaway angesprochenen Rätsels offensichtlich. So haben wir bislang noch nicht untersucht, ob J.H. Ottaway tatsächlich recht hat mit dem theoretischen Postulat, nachdem die Breite einer ^{14}C -Verteilung mit zunehmender Datenzahl nur immer größer wird. Theoretisch mag das zu erwarten sein. Verhalten sich ^{14}C -Daten aber auch in der realen Welt auf diese Weise? Zur Klärung dieser Frage werden wir ein Experiment ansetzen. Wir betrachten die Kalibration verschieden großer Stichproben aus einem Satz von insgesamt $N=47$ Daten der ägyptischen 19. Dynastie (Abb. 2).

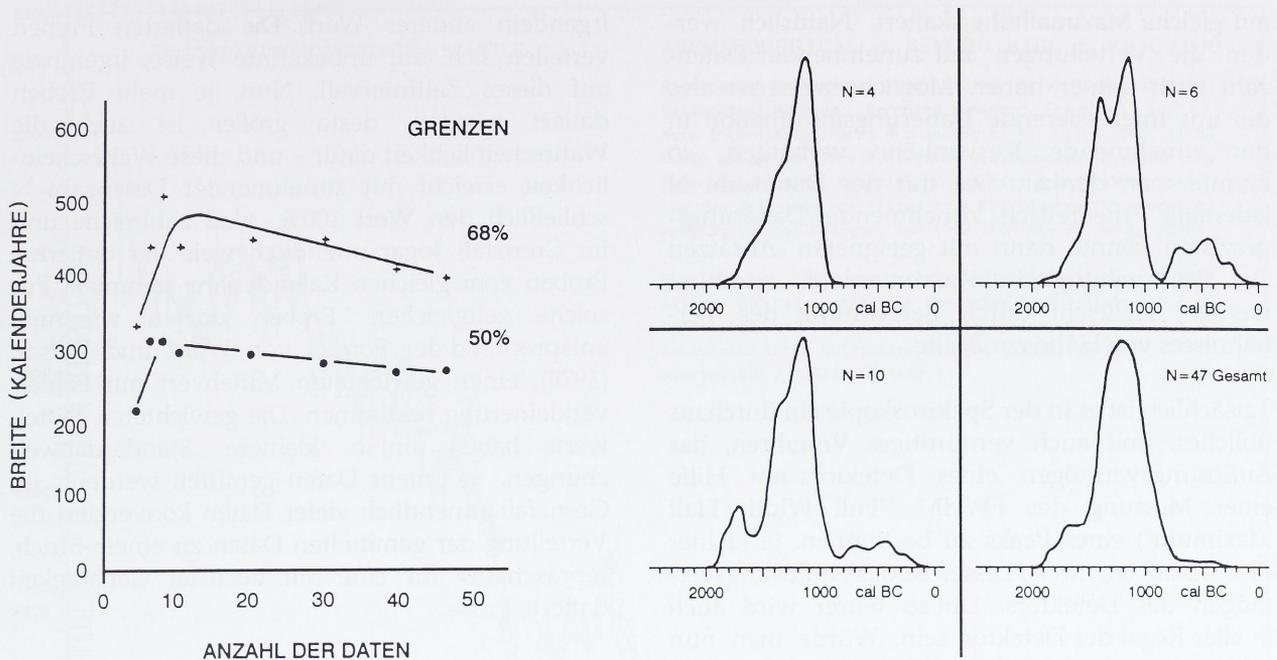


Abb.2 Abhängigkeit der statistischen Grenzen von der Datenzahl. Verschiedene Zufallsstichproben von Daten der ägyptischen 19. Dynastie (WENINGER, 1992)

Die Kalibration kann mit der Strategie des "Resampling" durchgeführt werden, einem Ansatz der jüngeren mathematischen Forschung, dem "Bootstapping" bzw. anderen Simulationsverfahren nahe verwandt (BRUCE 1996). In der folgenden Anwendung ziehen wir aus der vorliegenden Grundgesamtheit von $N = 47$ Daten eine schrittweise immer größer gewählte Stichprobe, mit Zurücklegen bereits gezogener Daten, und kalibrieren diese nach der 2-D-Methode. Abb. 2 zeigt die kalibrierten Wahrscheinlichkeitsverteilungen von Zufallsstichproben mit $N = 4, 6$ und 10 Daten, sowie die kalibrierte Verteilung aller 47 Daten. Zusätzlich berechnet (Abb. 2, links) wurde die Abhängigkeit der berechneten Länge der kalenderzeitlichen Mutungsintervalle, auf 50%- und 68%-Niveau, von der Datenzahl.

Die Ergebnisse sind wie folgt: Für $N=4$ Daten hat das 68%-Intervall die Länge 320 Jahre. Erhöht sich die Datenzahl auf $N=10$, so nimmt die Länge des Intervalls auf 450-500 Jahre zu, und verweilt bei diesem Wert auch bei einer weiteren Erhöhung der Datenzahl auf $N=20$ bis 30 , um bei $N=40$ bis $N=50$ nur langsam auf 400 Jahre abzusinken.

Das Experiment zeigt, daß die Breite der kalenderzeitlichen Wahrscheinlichkeitsverteilung kaum von der Datenzahl abhängt, und wenn überhaupt nur geringfügig kleiner wird. Somit wird die Hypothese von J.H. Ottaway für $N=4$ bis $N=10$ vollauf bestätigt, und sogar im Bereich $N=20$ bis $N=50$ kaum signifikant widerlegt. Wir werden offenbar, sogar bei Anwendung der Methode der 2-D-Dispersionskalibration (womit alle quantifizierbaren Fehler der Kalibration wie Kurvenverlauf, Standardabweichungen der archäologischen Daten und der Kalibrationsdaten berücksichtigt wurden), nicht dafür belohnt, mehr als vier oder sechs Daten gemessen zu haben.

Vertiefen wir also die Betrachtungen. Natürlich haben sich die Verteilungen der Daten mit zunehmender Datenzahl tatsächlich verändert, wenn auch nicht in ihrer Form, sondern in ihrer Gesamtfläche, wie auch in ihrer relativen Amplitude. Das geht allerdings nicht direkt aus Abb. 2 hervor. Um nämlich die Abhängigkeit der Kurvenbreite von der Datenzahl studieren zu können, dem Vorschlag von J.H. Ottaway entsprechend, habe ich die Verteilungen in Abb. 2

auf gleiche Maximalhöhe skaliert. Natürlich werden die Verteilungen mit zunehmender Datenzahl auch immer höher. Möglicherweise ist also die uns interessierende Datierungsinformation in der zunehmende Kurvenhöhe verborgen, so könnte man denken. Die mit der Datenzahl N jedenfalls theoretisch zunehmende Datierungspräzision könnte dann mit geeigneten Ansätzen zur Beschreibung des Kurvenverlaufs extrahiert werden, vielleicht durch Berechnung des Verhältnisses von Höhe zur Breite.

Tatsächlich ist es in der Spektroskopie ein durchaus übliches, und auch vernünftiges Verfahren, das Auflösungsvermögen eines Detektors mit Hilfe einer Messung des FWHM ("Full Width Half Maximum") eines Peaks zu bestimmen. Je kleiner dieses Maß ist, um so besser ist das Auflösungsvermögen des Detektors. Um so teurer wird auch in aller Regel der Detektor sein. Würde man nun auch bei archäologischen ^{14}C -Daten ein solches Maß berechnen, was technisch keine weiteren Schwierigkeiten bereitet, so wäre aber die Verwunderung doch wohl allzugroß, vom Laboratorium als Datierungsergebnis zu erfahren, die Proben der eingesandten Datenserie hätten eine kalibrierte Wahrscheinlichkeitsverteilung mit Breite z.B. 400 Kalenderjahre und Höhe 10 cm (je nach Skalierung) ergeben, und als zusätzliche Information, diese Datierung sei sicherlich genauer als die gleiche Kurvenbreite von 400 Jahren bezogen auf 5 cm Maximalamplitude, nämlich der halben Datenserie. Ein auf diese Weise vermitteltes Datierungsergebnis scheint mir fast ebenso merkwürdig zu sein, wie das von J.H. Ottaway formulierte Rätsel. So stellt sich noch einmal die Frage, wie gelangt man mit Hilfe großer Datenzahlen zu genaueren Ergebnissen?

Auflösung der Antinomie der Deutung von ^{14}C -Wahrscheinlichkeitsverteilungen

Man kommt einer Auflösung des Rätsels durch folgendes Gedankenexperiment näher. Angenommen, es würden nicht viele, sondern sogar unendlich viele ^{14}C -Daten vorliegen, wenn auch mit jeweils begrenzter Meßgenauigkeit. Wir betrachten also den Grenzfall, um bei dem obigen Beispiel (Abb. 1) zu bleiben, daß alle nur verfügbaren organischen Substanzen aus dem Zeitintervall der 19. Dynastie mit Hilfe der ^{14}C -Methode datiert wurden. Gehen wir ferner davon aus, die 19. Dynastie habe tatsächlich eine Dauer von 200 Jahren (oder

irgendein anderer Wert). Die datierten Proben verteilen sich, auf unbekannte Weise, irgendwie auf dieses Zeitintervall. Nun, je mehr Proben datiert wurden, desto größer ist auch die Wahrscheinlichkeit dafür - und diese Wahrscheinlichkeit erreicht mit zunehmender Datenzahl N schließlich den Wert 100% - daß zahlreiche und im Grenzfall sogar unendlich viele der datierten Proben vom gleichen Kalenderjahr stammen. Für solche zeitgleichen Proben dürfen wir nun, entsprechend der Formel von Ward und Wilson (1978), einen gewichteten Mittelwert mit Fehlerverkleinerung bestimmen. Die gewichteten Mittelwerte haben um so kleinere Standardabweichungen, je mehr Daten gemittelt werden. Im Grenzfall unendlich vieler Daten konvergiert die Verteilung der gemittelten Daten zu einem Strich, repräsentativ für eine mit höchster Genauigkeit datierte Probe.

Zunächst auf der ^{14}C -Skala betrachtet, erhalten wir also, da nun unendlich viele solcher genau datierter Proben unterschiedlichen Alters vorliegen, eine Folge unendlich vieler Striche. Die Probenabfolge ist zwar archäologisch zeitlich begrenzt, weil die Proben einem begrenzten Kalenderintervall von, sagen wir, 200 Jahren entsprechen. Da aber unendlich viele ^{14}C -Daten vorliegen sollen, werden auch unendlich viele Extremwerte aufgetreten sein. Deswegen ist die Abfolge der ^{14}C -Daten unbegrenzt und die Daten streuen entlang der gesamten ^{14}C -Skala von 0 bis unendlich BP. Das mag zwar merkwürdig erscheinen, entspricht aber den Ausführungen von J.H. Ottaway.

Die Abfolge der kalibrierten Daten besitzt nun Häufungsstellen überall dort, wo die Kalibrationskurve flach verläuft. Wird die nun vorliegende Folge von unendlich vielen Strichen kalibriert, so besitzt die Verteilung der Striche auf der Kalenderzeitskala immer Häufungsstellen dort, wo die Kalibrationskurve flach verläuft. Könnte man diese Häufungsstellen graphisch feiner auflösen, so würde man allerdings eine Folge von unterschiedlich dicht mit Strichen belegten Intervallen erkennen, je nach Verlauf der Kalibrationskurve im betreffenden Zeitabschnitt.

Einen Eindruck vom Verhalten der konventionellen und ^{14}C -Daten, und der entsprechenden kalibrierten, gewichteten Mittelwerte, vermitteln Abb. 3 und Abb. 4 für den realen Fall einer zeitlich begrenzten Probengleichverteilung mit Dauer 700 Kalenderjahre. Abb. 3 zeigt den Fall, da die

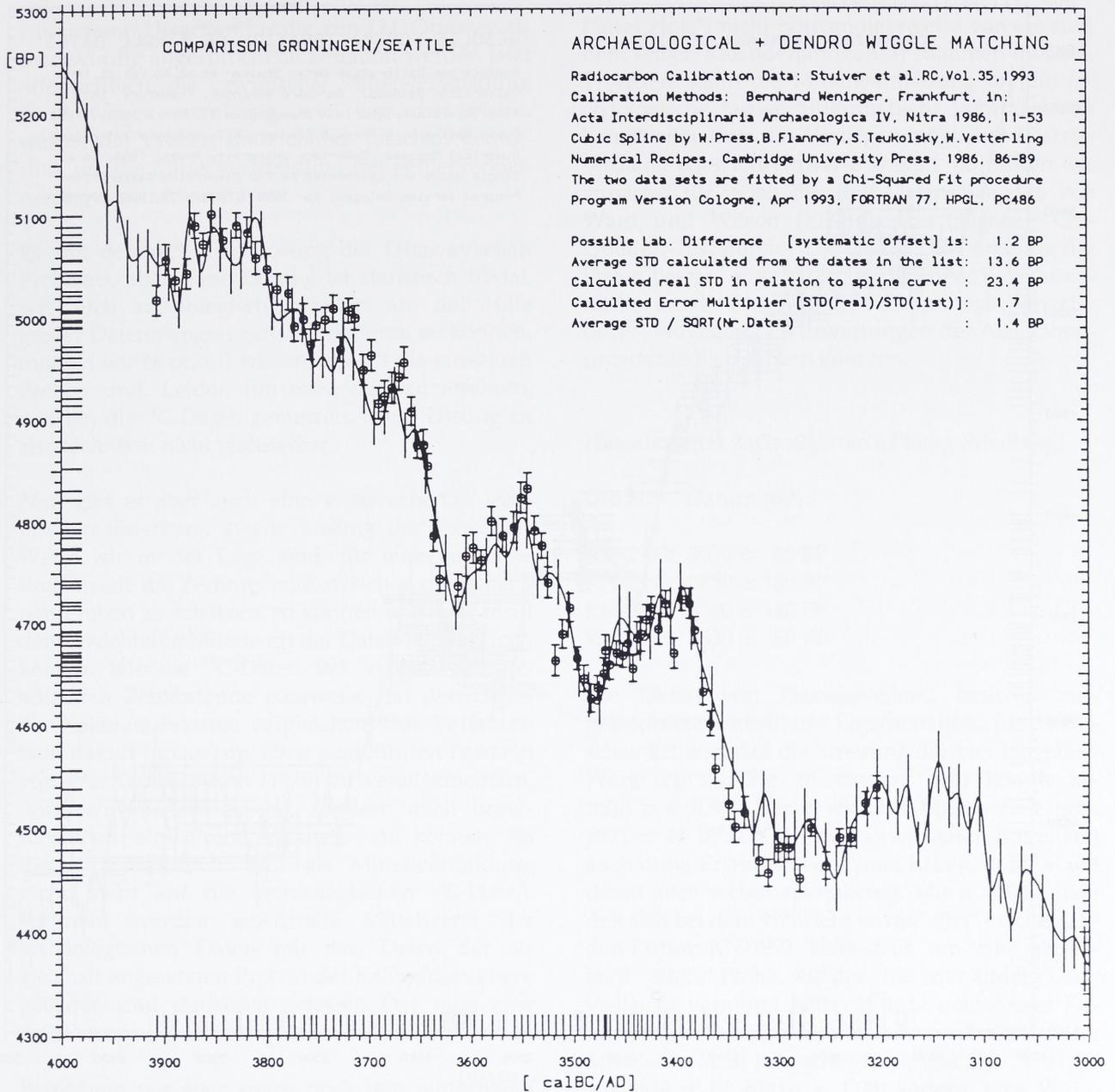


Abb. 3 Messungen der Laboratorien Groningen ($\pm 1 \sigma$ Meßbalken) und Belfast (Spline Kurve) im Zeitbereich 4000 - 3000 calBC. Daten: de Jong, Becker und Mook (1986); Pearson, Pilcher, Baillie, Corbett und Qua (1986). Striche auf der ^{14}C -Skala: BP-Mittelwerte, Groningen Laboratorium. Striche auf der Dendroskala: bekanntes Probenalter 3203-3904 denBC.

Probenreihenfolge bekannt ist. Abb. 4 zeigt den Fall, da die Probenreihenfolge unbekannt ist. In Abb. 4 wurden die Daten einzeln kalibriert, ohne Kenntnis der Probenreihenfolge. Bei den Strichen auf der Dendroskala handelt es sich um die Medianwerte der berechneten Dispersionsverteilungen der einzelnen Daten, die nun auf die Kalibrationskurve hochprojiziert werden können, und dort mit den 68%-Grenzen der Ausgangsdaten und

der kalibrierten Verteilungen versehen sind. Die Abfolge der Striche in Abb. 4 zeigt die erwähnten, artifiziellen Häufungsstellen auf der Dendroskala. Die Abfolge der Striche auf der ^{14}C -Skala belegt die Existenz entsprechender Häufungsstellen der unkalibrierten Daten.

Wir führen das Gedankenexperiment nun zum grausamen Ende, und bilden (illegalerweise) den

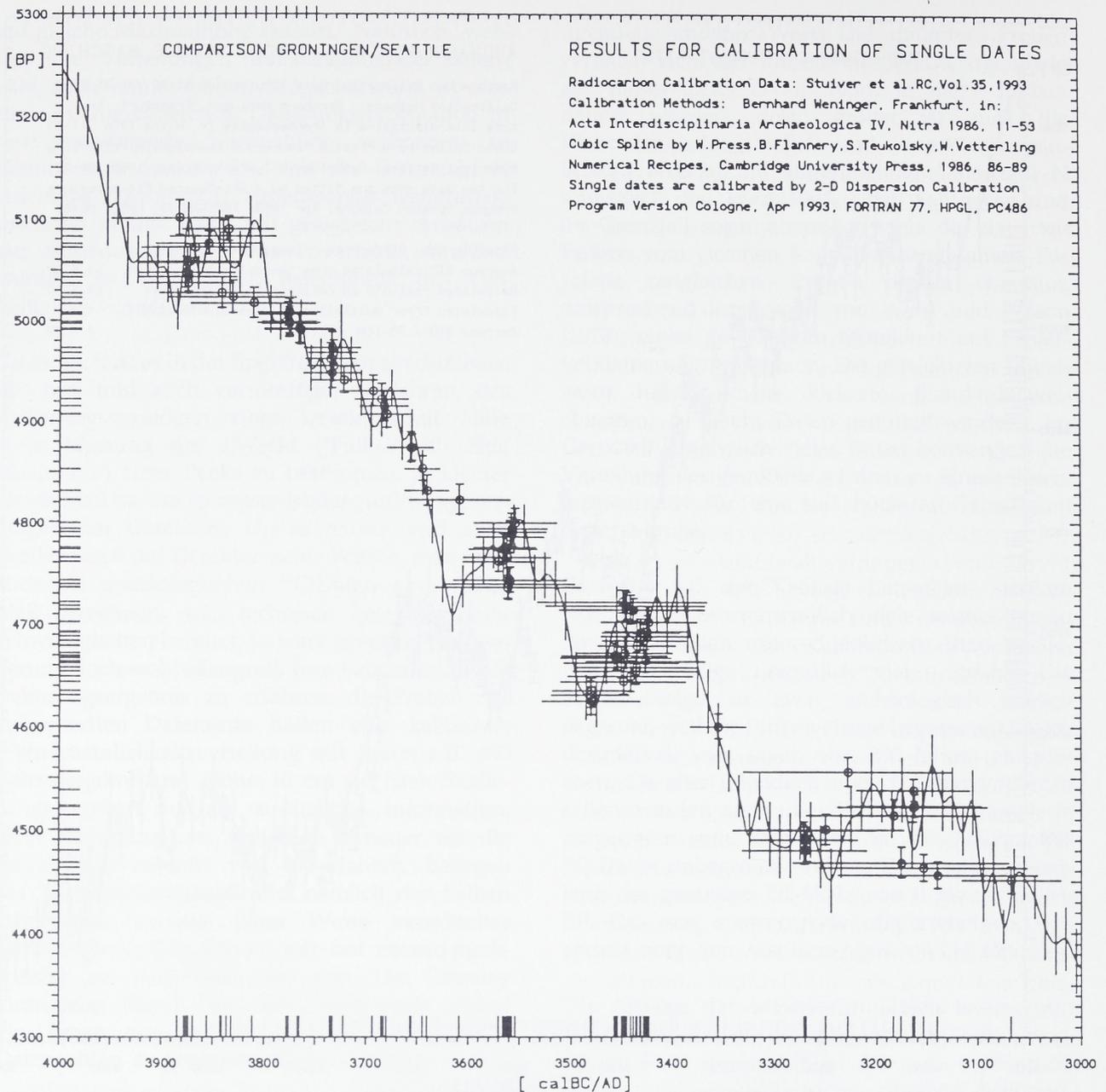


Abb. 4 ^{14}C -Daten wie Abb. 3, in Unkenntnis des Probenalters. Striche auf der ^{14}C -Skala: BP-Mittelwerte, Laboratorium Groningen. Striche auf der Dendroskala: Medienwerte der cal-Einzeldaten. Kreuze in der Kalibrationskurve, senkrechte Linien: $\pm 1 \sigma$ BP-Meßbalken, waagrechte Linien: $\pm 1 \sigma$ cal-BC Intervalle.

gewichteten Mittelwert aller, also unendlich vieler Striche (^{14}C -Mittelwerte mit minimalen Standardabweichungen). Das Ergebnis ist ein singulärer, hochgradig schwarzer Strich auf der ^{14}C -Skala. Passend dazu gibt es auf der Dendroskala nicht nur einen hochgradig schwarzen Strich, sondern (wegen der mehrfachen Ablesungen) eine kurze Abfolge von besonders geschwärzten Intervallen, die durch Leerstellen getrennt sind.

Damit haben wir nun zuletzt den Fehler gemacht, die Dauer der 19. Dynastie von 200 Jahren auf 0 zu reduzieren. Eigentlich hätte die Mittelwertbildung dort aufhören müssen, da die Intervalle auf der Dendroskala exakt die Dauer der 19. Dynastie abdecken. Das gleiche Gedankenexperiment kann man übrigens auch mit dem Verfahren der punktweise Dispersionskalibration nach Ottaway (1973) ansetzen, mit gleichen Ergebnissen.

Die tieferen Ursachen für die von J.H. Ottaway als merkwürdig angesprochene Situation werden jetzt offensichtlich: die gewünschte Mittelwertbildung ist deswegen unzulässig, weil wir *nicht wissen*, welche der Proben hinreichend "gleichalt genug" sind, um die Mittelwertbildung mit Fehlerverkleinerung ansetzen zu dürfen.

Es gibt demnach eine Lösung des Ottaway'schen Problems, aber diese Lösung ist statistisch trivial, wie auch archäologisch zirkulär: um mit Hilfe großer Datenmengen genauer datieren zu können, müßten wir (a priori) wissen, wie alt die einzelnen Proben sind. Leider, um eben das zu erfahren, wurden die ^{14}C -Daten gemessen. Diese Lösung ist also technisch nicht realisierbar.

Nun gibt es aber auch eine entsprechende technisch realisierbare, zweite Lösung des Problems. Wenn wir in der Lage sind, für eine gegebene Probenserie die Zeitintervalle zwischen den einzelnen Proben zu schätzen, so können wir zwar nicht den gewichteten Mittelwert der Daten bilden. Doch können wir die ^{14}C -Daten bei Vorgabe der geschätzten Zeitabstände paarweise mit den Daten der Kalibrationskurve vergleichen. Das Verfahren läuft darauf hinaus, die oben angeführten Formeln von Ward und Wilson (1978) zu verallgemeinern, um nicht nur zeitgleiche, sondern auch unterschiedlich alte Proben mitteln zu können. In diesem Fall bezieht sich die Mittelwertbildung nicht mehr auf die archäologischen ^{14}C -Daten. Vielmehr werden gewichtete Mittelwerte der archäologischen Daten mit den Daten der als gleichalt angesetzten Proben der Kalibrationskurve gebildet, und statistisch getestet. Das setzt eine Annahme zur Besiedlungsdauer voraus.

Betrachten wir aber zuvor noch den einfacheren Fall, daß die Proben der drei mittleneolithischen Fundorte je gleichalt sind (Besiedlungsdauer jeweils 0 Jahre).

Statistische Auswertung der Daten aus Hasselsweiler 2, Inden 3, und Inden 1 Annahme: Besiedlungsdauer = 0 Jahre...

Es ist, wie oben erwähnt, auch in diesem Fall strenggenommen nicht zulässig, einen fehlerverkleinerten gewichteten Mittelwert nach Ward und Wilson (1978) zu bilden. Der Grund ist, daß die Proben schon aus archäobotanischen Gründen

("altes Holz") nicht notwendigerweise von ein und dem selben Kalenderjahr(zehnt) stammen müssen. Trotzdem kann die Mittelwertbildung für die archäologische Auswertung nützliche Informationen bereitstellen, was das Vorliegen möglicher Ausreisserdaten oder Extremwerte angeht. So führen wir zunächst für jeden der drei Datensätze den von Ward und Wilson (1978) vorgeschlagenen Chi-Quadrat-Test durch¹. Das Ziel ist, die Streuung der Daten besser einschätzen zu können. Es geht zunächst um die Frage, ob die Proben möglicherweise doch - entgegen den Erwartungen der Ausgräber - unterschiedlich alt sein könnten.

Hasselsweiler 2 (Großgartach, Planig-Friedberg)

Lab.Nr. Datum (BP)

KN-2992: 5930 ± 65 BP

KN-2993: 5850 ± 150 BP

KN-2994: 5840 ± 140 BP

KN-2995: 5830 ± 60 BP

Die Daten von Hasselsweiler 2 besitzen ausgezeichnete statistische Eigenschaften: die Wahrscheinlichkeit, daß die Streuung der vier einzelnen Werte rein zufällig entstanden sein könnte, beträgt $p = 0.86$. Der gewichtete Mittelwert ergebe 5871 ± 41 BP. Da wir den gewichteten Mittelwert auch ohne Erlaubnis berechnet haben, können wir damit auch weiter spekulieren. Mit $p = 0.86$ handelt sich bei dem vielleicht etwas "älter" erscheinenden Datum KN-2992 keinesfalls um eine signifikant "ältere" Probe, wie der eine oder andere Leser vielleicht vermutet hätte. Würde man dieses Datum trotzdem außer acht lassen, so würde sich der Mittelwert auch nur geringfügig von 5871 ± 41 BP auf 5834 ± 52 BP ($p = 1.00$) ändern. Eine Wahrscheinlichkeit von 100% ist auch kaum besser als 86%. Wir kommen somit zum Ergebnis, daß es sich ohne weiteres um innerhalb weniger Jahrzehnte zeitgleiche Proben handeln *könnte* - der archäologischen Hypothese der Ausgräber entsprechend. So wird man jedenfalls die Alternativhypothese, demnach die Proben signifikant unterschiedlich alt wären, aus statistischer Sicht problemlos verwerfen.

Die gleichen, nahezu idealen statistischen Eigenschaften besitzen auch die Daten von Inden 3:

Inden 3 (älteres Rössen II)

KN-2649: 5630 ± 145 BP
 KN-2650: 5830 ± 75 BP
 KN-2651: 5870 ± 60 BP
 KN-2652: 5820 ± 60 BP
 KN-2653: 5830 ± 65 BP

Wir werfen jetzt vollends alle noch vorhandenen Skrupel beiseite, den gewichteten Mittelwert zu berechnen. Die Daten von Inden 3 haben einen Mittelwert von 5829 ± 31 BP, und die Wahrscheinlichkeit dafür, daß die Streuung der Daten ein Zufallsprodukt ist, beträgt $p = 0.80$. Das ist zufriedenstellend. Läßt man interessehalber das möglicherweise zu jung erscheinende Datum KN-2649 weg, so ändert sich der Mittelwert nur geringfügig von 5829 ± 31 BP auf 5839 ± 32 BP ($p = 0.99$). Wir haben demnach keinen Grund, irgendeines der Daten bei der weiteren Auswertung wegzulassen. So wird es auch keine Rolle spielen, wenn wir hier anhand rein statistischer Überlegungen entscheiden, das Datum KN-2649 bei der späteren Auswertung doch mitzunehmen.

Aus statistischer Sicht etwas problematischer sind die Daten von Inden 1. Zunächst ist festzuhalten, daß die anhand der Labordokumente zusammengestellte Tabelle zwei ¹⁴C-Daten enthält, die bei Biermann nicht aufgeführt sind. Es handelt sich um die Daten KN-I.523: 6010 ± 285 BP und KN-I.524: 6180 ± 50 BP, die Biermann wohl wegen ihres zu hoch erscheinenden Alters von vornherein außer acht gelassen hat.

Inden 1 (jüngeres Rössen II)

KN-I.190: 5840 ± 45 BP
 KN-I.330: 5790 ± 50 BP
 KN-I.345: 5700 ± 75 BP
 KN-I.517: 5820 ± 45 BP
 KN-I.520: 5750 ± 50 BP
 KN-I.521: 5900 ± 40 BP
 KN-I.523: 6010 ± 285 BP
 KN-I.524: 6180 ± 50 BP
 KN-I.525: 5880 ± 65 BP
 KN-I.526: 5720 ± 90 BP
 KN-I.527: 5800 ± 95 BP

Biermann hatte in der Tat gute Gründe, diese Daten wegzulassen, denn der gewichtete Mittelwert von 5862 ± 17 BP hat eine berechnete

Wahrscheinlichkeit $p = 0.00$. Entfernt man die Probe KN-I.524 mit dem Argument, es handele sich um einen unglaublichen Extremwert, so besitzt der neue Mittelwert 5820 ± 18 BP eine Wahrscheinlichkeit $p = 0.39$. Läßt man dazu noch den in der Ausreißer-Rangordnung nachfolgenden Wert für KN-I.521 weg, so liegt die Wahrscheinlichkeit zufriedenstellend bei $p = 0.84$ und die Serie besitzt einen statistisch bereinigten, gewichteten Mittelwert von 5800 ± 20 BP.

Schließlich zeigt folgende Zusammenstellung der statistisch bereinigten gewichteten Mittelwerte, daß die ¹⁴C-Daten aller drei Siedlungen identisch sind:

Siedlung	Mittelwert BP	Anzahl Daten (BP-identisch)	Ausreißer (BP-unterschiedlich)
Hasselweiler 2	5871 ± 41 BP	4	0
Inden 3	5829 ± 31 BP	5	0
Inden 1	5800 ± 20 BP	11	2

Die Wahrscheinlichkeit, daß die 20 Daten, welche den Test von Ward und Wilson (1978) bestanden haben, einen identischen ¹⁴C-Gehalt haben, beträgt $p = 0.71$. Es liegen nur zwei Proben vor, welche bei der statistischen Auswertung besonders auffallen. Die erste Probe KN-I.521: 5900 ± 40 BP entstammt der Grube 710d von Inden 1. Die zweite Probe KN-I.524: 6180 ± 50 BP ist ebenfalls etwas zu "alt" geraten - und wurde ebenfalls recht genau gemessen. Diese Probe entstammt einer Siedlungsgrube (Stelle 985) in Inden 1, welche von den Bauten XXVII und XXVIII überschritten wird. So könnte es sich in beiden Fällen um "altes Holz" bzw. innere Jahrringe aus größeren Holzbalken handeln. Damit könnten 20 der insgesamt vorliegenden 22 Proben (also immerhin 91% !) im Rahmen weniger Jahrzehnte gleichalt sein. Nur - wir können uns dessen nicht sicher sein. Weiterführende Aussagen über das tatsächliche Probenalter sind, wie erwähnt, mit Hilfe derartiger statistischer Tests nicht möglich. Dazu bedarf es weiterer Untersuchungen.

Monte Carlo Wiggle Matching

Die Idee zum Monte Carlo Wiggle Matching entstand vor einigen Jahren, als ich dabei war, die Methode des Archäologischen Wiggle Matching mit Hilfe der ¹⁴C-Daten aus Ägypten zu testen. Die wichtigste Voraussetzung des AWM ist eine

quantitative zeitliche Gliederung der datierten Proben, die man für die ägyptische Chronologie bequem der mehrbändigen Ausgabe der "Cambridge Ancient History" (CAH) entnehmen kann. Nachdem die datierten Proben zeitlich gegliedert sind, werden die entsprechenden ^{14}C -Daten einem FORTRAN-Rechenprogramm mitgeteilt (<AWM>), wobei die Daten von jung nach alt in eine Datentabelle einzugeben sind. Die jüngste Probe erhält den (kalenderzeitlichen) Abstand $t=0$. Damit wird das zu bestimmende Endjahr der Serie definiert. Die älteren Proben erhalten zeitliche Abstände, bezogen auf diesen Nullpunkt.

Wie sich zeigte, stimmen die Ergebnisse des historischen Wiggle Matching gut, mit Abweichungen im Rahmen einiger Jahre bis Jahrzehnte, mit der vorgegebenen CAH-Chronologie überein, für alle Dynastien I-XX. Allerdings sind die Standardabweichungen der meisten ^{14}C -Daten an den ägyptischen Proben recht groß (± 50 bis ± 100 BP) und da die "wiggles" der Kalibrationskurve ähnliche Amplituden besitzen, gelingt es nicht, über eine generelle Bestätigung der "mittleren" ägyptischen Chronologie hinauszukommen.

Immerhin konnte mit Hilfe des AWM ermittelt werden, daß die gezielte ^{14}C -Messung von kurzlebigen Proben mit gesicherter Zuordnung zu den beiden vorletzten Pharaonen der I. Dynastie es aller Voraussicht nach ermöglichen müßte, eine hochauflösende Chronologie (± 10 Kalenderjahre) der I. Dynastie zu erarbeiten (WENINGER 1986).

Bei diesen Versuchen ist mir nun wiederholt der Fehler unterlaufen, eine Probenserie zunächst falsch zu konzipieren oder sogar, im Extremfall, eine korrekte Serie zeitlich invers in die Datentabelle einzugeben. Dabei wurde die jüngste Probe fälschlich als älteste Probe gerechnet, die älteste Probe wurde zur jüngsten, und so fort, und auf diese Weise habe ich wiederholt die statistisch besten Einpassungen von schlecht gegliederten und zuweilen sogar von zeitlich rückwärts laufenden Datenserien ermittelt.

Da die AWM-Ergebnisse trotzdem durchweg gut mit der später ermittelten, korrigierten AWM-Einpassung übereinstimmen, habe ich zur Kontrolle weitere Versuche mit Dendroserien (jahrgenau bekanntes Endjahr, genau bekanntes Zeitintervall) durchgeführt, mit dem gleichen Ergebnis.

Der allgemeine Befund ist a) daß AWM-Einpassungen kaum von der Probenreihenfolge abhängen. Sie sind b) gegenüber Vertauschungen einzelner Proben in der Serie robust. Das Ergebnis kann c) stark von der Länge des kalenderzeitlichen Gesamtintervalls abhängen. Das muß d) aber nicht der Fall sein, und es gibt zahlreiche Beispiele dafür, daß auch bei einer immer weiteren Verlängerung des Gesamtintervalls einzelne ^{14}C -datierte Phasen durchaus stabile AWM-Endjahre besitzen (MANNING UND WENINGER 1992; WENINGER 1992).

Es fällt insbesondere auf, daß die Ergebnisse des Wiggle Matching auch im Extremfall, für zeitlich rückwärts laufende Probenserien, im Rahmen weniger Kalenderjahrzehnte korrekt ausfallen. Eine Erklärung für diese Beobachtung ist darin zu suchen, daß die dem Wiggle Matching zugrundegelegten numerischen Methoden schon vom Ansatz her dafür ausgelegt sind, robuste Ergebnisse zu liefern. Bei dem im Programm <AWM> angesetzten Algorithmus (Methode der kleinsten Fehlerquadrate), jedenfalls, wird der Versuch unternommen, eine Einpassung der Daten zu finden, derart, daß die einzelnen Daten der Serie möglichst symmetrisch zur Kalibrationskurve liegen. Dabei wird das statistische Gewicht von ^{14}C -Daten oberhalb der Kurve durch das Gewicht der Daten unter der Kurve ausgeglichen.

Stellt man sich die einzupassende Datenserie als eine entlang der Dendro-Zeitachse gestreckte Ellipse vor, die mit zuweilen stark streuenden ^{14}C -Daten ausgefüllt ist, so wird diese Ellipse mit Hilfe des AWM auf der Kalibrationskurve zentriert eingepaßt. So spielt es keine große Rolle für das Ergebnis, ob die Datenserie nun vorwärts oder rückwärts laufend eingepaßt wird. Zwar besitzen die Einpassungen für rückwärts laufenden Serien geradezu "katastrophale" statistische Maße, und derartige Einpassungen sind aus statistischer Sicht abzulehnen. Doch erweist sich das berechnete Endjahr der Serie selbst als überaus robust gegenüber der gewählten Probenreihenfolge.

Im Unterschied zur Probenreihenfolge spielt die zeitliche Ausdehnung der Ellipse in Richtung der Dendro-Zeitskala eine gewichtige Rolle. Wird die Ellipse gedehnt oder gestreckt, so bleibt das berechnete Kalenderalter von Proben im mittleren Bereich der Ellipse zwar zumeist erhalten, wie gewünscht, und dem Zentralwertsatz der Wahrscheinlichkeitstheorie folgend. Je länger die Ellipse ist, desto mehr

rutschen seine beiden Enden aber hin zu jüngeren bzw. älteren Kalenderjahren. Der wichtigste Parameter beim AWM erscheint demnach vor allem die Gesamtdauer der Serie zu sein, nicht aber die genaue Probenreihenfolge.

Beim Monte Carlo Wiggle Matching wird diese Beobachtung technisch nach folgendem Algorithmus umgesetzt:

Do 2 Für Intervalle 0-500 Kalenderjahre, in Schritten 10 Jahre.

Wähle ein kalenderzeitliches Intervall für die mögliche Siedlungsdauer.

Do 1 Für beliebige Probenreihenfolgen, $M=1$ bis 1000 Versuche.

Plaziere die Proben per Zufallsgenerator irgendwo in diesem Intervall. Liegen N Daten vor, so können die Proben z.B. in gleichmäßigen Abständen von $20/N$ Jahren auf das Intervall aufgeteilt werden.

-Berechne die beste Einpassung der Datenserie in die Kalibrationskurve.

-Berechnetes Endjahr speichern.

-Berechnetes Maß der Einpassung speichern (Chi-Quadrat-Wert).

1 Fortfahren mit Schritt 1

(andere Probenreihenfolge)

2 Fortfahren mit Schritt 2

(Verlängerung des Intervalls)

Als Ergebnis kann ein Diagramm erstellt werden, in dem die statistische Streuung der berechneten Endjahre der Serien in Abhängigkeit von der Länge des gewählten kalenderzeitlichen Intervalls dargestellt wird. Da zu jedem berechneten Endjahr dessen statistische Wertigkeit (Chi-Quadrat) gespeichert wird, können die einzelnen Endjahre entweder akzeptiert oder verworfen werden. Anschließend werden die Verteilungen der Serien-Endjahre nach archäologischen Kriterien ausgewertet.

Monte Carlo Wiggle Matching: Inden 1

Bei der Eingabe wurden alle ^{14}C -Daten berücksichtigt, mit Ausnahme der zu alt erscheinenden

Datierung KN-I.524 (6180 ± 50 BP). Die Ergebnisse sind in Abb. 5 dargestellt. Sie lassen sich wie folgt nachvollziehen.

Die Proben werden mit einem Zufallsgenerator jeweils 9000 mal gleichmäßig auf zunehmend große Intervalle der Länge 18, 36, 54...bis maximal 180 Jahre gestreut (Rechenzeit: 40 Stunden mit 486-Prozessor). Abb. 5 zeigt die Verteilungen der berechneten Endjahre für jedes dieser Intervalle. Die Spalte mit Überschrift "Length (yrs)" zeigt das jeweilige Intervall an. Zur Orientierung wird eine Gaußkurve durch die Verteilung der Endjahre gelegt, und zwar auch dann, wenn die Verteilung multimodal ist. Dessen Median und 68%-Mutungsintervall wird in der Spalte "Average Fit-Yr" numerisch ausgegeben. Sollte die Verteilung multimodal sein, wie z.B. bei den Intervallen mit Dauer 16 bis 48 Jahre der Fall ist, so wird man das notierte 68%-Sicherheitsintervall verdoppeln, und die Auswertung auf 95%-Niveau konzipieren. Die Verteilungen der Endjahre sind, der Existenz von Schwingungen der Kalibrationskurve entsprechend, teilweise quantisiert ("Lock-In-Effekt").

Nun wird die mögliche Dauer der Besiedlung von Inden 1 von Biermann (siehe oben) auf 100 bis 150 Jahre geschätzt, was 4-5 Hausgenerationen entspricht. Für diese Intervalle liegen die Endjahre im Bereich 4700 bis 4530 calBC. Dabei ist zu betonen, daß keine Permutation der datierten Holzkohleproben, einer Zufallsstichprobe der Größe $N=9000$ entnommen, eine Datierung vor 4700 calBC ergeben hat. Einige wenige Ablesungen vor 4700 calBC erhält man lediglich für Intervalle mit Länge kürzer als 54 Jahre. Diese kurzen Intervalle können zwar anhand von archäologischen Erwägungen zur Gesamtdauer der Siedlungssequenz ausgeschlossen werden, aber man sollte trotzdem damit rechnen, daß die eine oder andere Holzkohle aus Inden 1 in die Anfangsphase um oder kurz nach 4700 calBC datiert.

Monte Carlo Wiggle Matching: Hasselsweiler 2

Bei der Eingabe wurden alle Daten berücksichtigt. Es wurden pro Intervall jeweils 1000 Iterationen der Probenreihenfolge getestet. Die Ergebnisse für Intervalle der Länge 6 bis 60 Jahre sind in Abb. 6 dargestellt. Da Biermann für Hasselsweiler 2 eine Besiedlungsdauer von 1-2 Hausgenerationen schätzt, und Intervalle der Dauer 20-40 Jahre

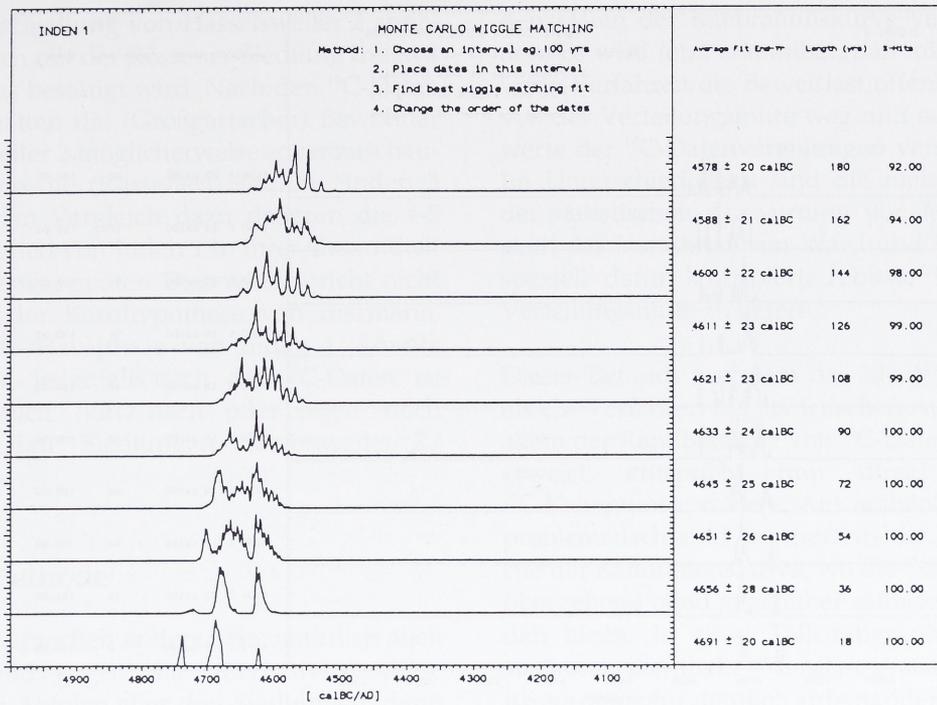


Abb. 5 Monte Carlo Wiggle Matching für die ¹⁴C-Daten von Inden 1. Simulierte Siedlungsdauer 18 bis 180 Jahre, in Schritten von 18 Jahren. Dargestellt sind die Verteilungen der Endjahre der eingepaßten Datenserien, mit Vertauschungen der Probenreihenfolge nach einem Zufallsgenerator.

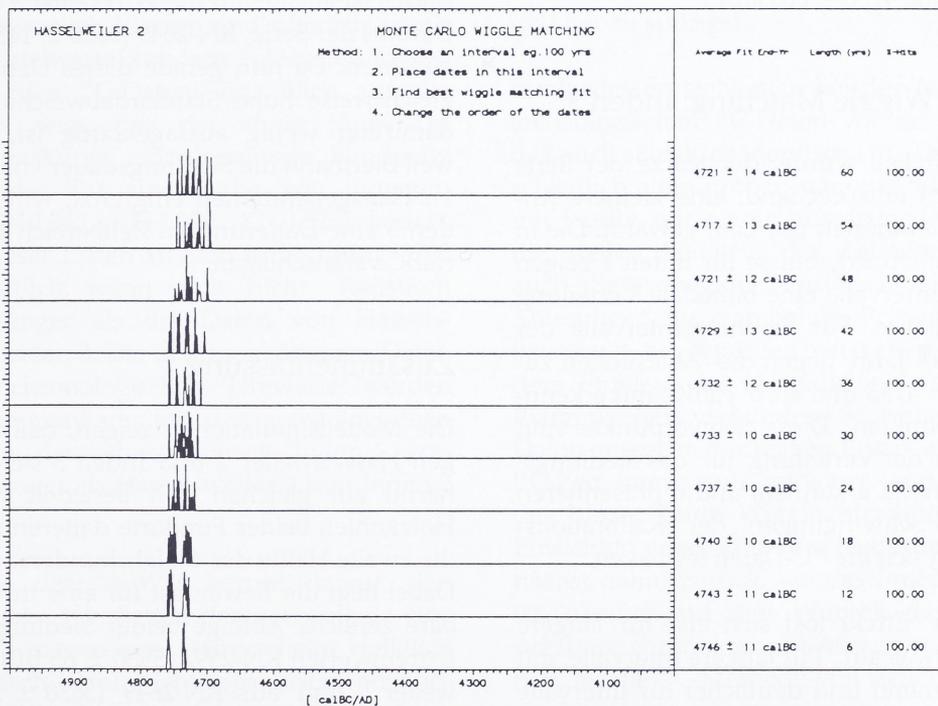


Abb. 6 Monte Carlo Wiggle Matching für die ¹⁴C-Daten von Hasselweiler 2. Intervalle 6 bis 60 Jahre, in Schritten von 6 Jahren.

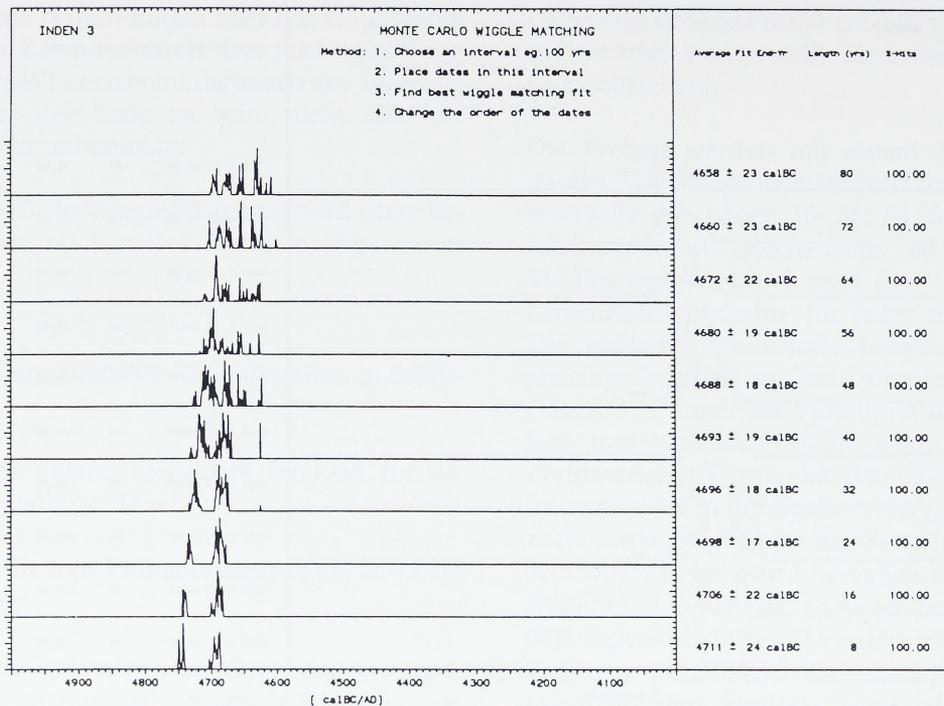


Abb. 7 Monte Carlo Wiggle Matching für die ^{14}C -Daten von Inden 3. Intervalle 8 bis 80 Jahre, in Schritten von 8 Jahren.

vorgibt, erhält man als Monte-Carlo-Ergebnis eine Datierungsspanne 4750-4700 calBC.

Monte Carlo Wiggle Matching: Inden 3

Auch in diesem Fall wurde, der Kürze der Serie ($N=5$ ^{14}C -Daten) entsprechend, eine kleinere Anzahl von 1000 Iterationen pro Lauf gewählt. Die in Abb. 7 dargestellten Ergebnisse für Inden 1 zeigen für die kurzen Intervalle eine bimodale Verteilung der Serien-Endjahre. Für Siedlungsintervalle der Dauer 8 bis 80 Jahre liegen die Ablesungen zunächst zwischen 4755 und 4670 calBC, mit erkennbaren Schwerpunkten. Diese Schwerpunkte sind auch innerhalb der Verteilung für das Siedlungsintervall 48 Jahre zu erkennen, und repräsentieren entsprechende Schwingungen der Kalibrationskurve, an denen sich die ^{14}C -Daten festhaken.

Dieser "Lock-In" Effekt löst sich erst für längere Zeitintervalle etwas auf. Für längere Intervalle, mit 32 Jahren beginnend und deutlicher für Intervalle der Dauer 40 bis 56 Jahre, treten bevorzugt Ablesungen auf dem nächst jüngeren "Quantenzustand" 4620 calBC der Kalibrationskurve auf. Die mit der Siedlungsdauer weiter zunehmende

Häufigkeit der jungen Ablesungen (4670-4600 calBC) ist alleine auf die Wirksamkeit des jüngsten Datums der Serie, KN-2649 (5630 ± 145 BP) zurückzuführen. Da nun gerade dieses Datum eine vergleichsweise hohe Standardabweichung hat, und damit nur wenig aussagekräftig ist, aber auch weil Biermann die Siedlungsdauer von Inden 3 auf 1-2 Hausgenerationen eingrenzt, wird man für Inden 3 eine Datierung im Zeitbereich 4755 bis 4670 calBC veranschlagen.

Zusammenfassung

Die Modellsimulationen zeigen, daß die Siedlungen Hasselsweiler 2 und Inden 3 sicherlich annähernd zur gleichen Zeit besiedelt wurden. Die Holzkohlen beider Fundorte datieren durchweg in die zweite Hälfte des 48. Jahrhunderts vor Christus. Dabei liegt die Beweislast für eine trotzdem denkbare zeitliche Abfolge beider Siedlungen auf den Extremwerten KN-2992 (5930 ± 65 BP) für Hasselsweiler 2 und auf KN-2649 (5630 ± 145 BP) für Inden 3.

Wir kommen damit zum Ergebnis, daß die zentrale Aussage von Biermann, demnach die

Großgartacher Siedlung von Hasselsweiler 2 annähernd zeitgleich mit der Rössener Siedlung Inden 3 wäre, durchaus bestätigt wird. Nach den ^{14}C -Daten zu urteilen, hätten die (Großgartacher) Bewohner von Hasselsweiler 2 möglicherweise sogar zuschauen können, wie die (Rössener) Siedlung Inden 3 gebaut wird. Im Vergleich dazu datieren die 4-5 Hausgenerationen von Inden 1 in ihrer Gesamtheit anscheinend etwas später. Dies widerspricht nicht zwangsläufig der Kernhypothese von Biermann, denn die erste Hausphase von Inden 1 könnte ohne weiteres, jedenfalls nach den ^{14}C -Daten zu urteilen, zeitlich kurz nach oder sogar noch parallel zu den Siedlungen Hasselsweiler 2 / Inden 3 liegen.

Kritik der Methode

Die ^{14}C -Daten sprechen andererseits natürlich auch nicht zwingend gegen die Alternativhypothese einer direkten Abfolge aller drei Siedlungen, denn die gewichteten ^{14}C -Mittelwerte der drei Siedlungen sind nun einmal statistisch identisch, und die Kalibrationskurve ist im relevanten Zeitbereich einigermaßen flach. So muß es erstaunen, wenn es der Methode des "Monte Carlo Wiggle Matching" (MCWM) trotzdem gelingt, einen gewissen zeitlichen Abstand zwischen den Siedlungen offenzulegen. Verantwortlich hierfür sind offenbar weniger die ^{14}C -Mittelwerte, sondern Feinheiten in den Verteilungen der ^{14}C -Daten. Ein Blick auf die Datentabellen zeigt, daß der etwas jünger in die Kalibrationskurve einzuhängende Fundplatz Inden 1 in der Tat eine Reihe von jüngeren ^{14}C -Daten (KN-I.345, KN-I.520, KN-I.526) besitzt. Die Werte dieser Daten streuen in Richtung 5700 BP, also deutlich, wenn auch nicht statistisch signifikant, jünger als die Daten von Hasselsweiler 2 und Inden 3. Die in diesen jüngeren Daten enthaltenen chronologischen Hinweise werden erst im Zusammenhang mit den archäologischen Überlegungen deutlicher, wonach Inden 1 voraussichtlich länger als Hasselsweiler 2 und Inden 3 besiedelt war.

Die Methode des MCWM beruht darauf, den datierten Proben von Seiten der Archäologie eine zeitliche Tiefe zuzuweisen. Das erscheint vielleicht unproblematisch, wenn nicht sogar selbstverständlich zu sein. Wie oben angeführt, verallgemeinert das MCWM die Mittelwertbildung nach Ward und Wilson (1978) dahingehend, daß die Daten einer Serie nicht mehr untereinander, sondern mit

den Daten der Kalibrationskurve verglichen werden. Es wird jetzt erkennbar, daß mit dem Monte-Carlo-Verfahren die Beweislast offenbar bevorzugt von der Verteilungsmitte weg und auf die Extremwerte der ^{14}C -Datenverteilungen verschoben wird. Im Unterschied dazu sind die meisten Konzepte der statistischen Auswertung von Meßdaten, wie auch das Verfahren von Ward und Wilson (1978), speziell dafür konzipiert, robuste Maße für die Verteilungsmitte zu liefern.

Dieser Befund, nachdem das MCWM sich zuletzt als ein Verfahren der statistischen Auswertung vor allem der Randbereiche von ^{14}C -Datenverteilungen erweist, entspricht nun direkt auch dem ^{14}C -Kalibrationsproblem. Aus archäologischer Sicht problematisch sind ja einerseits die flachen Bereiche der Kalibrationskurve, wo die ^{14}C -Methode vorübergehend blind gegenüber zeitlichen Unterschieden bleibt. In einer Tellstratigraphie kann man statistisch identische ^{14}C -Alter, je nach Verlauf der Kurve, auch für deutlich aufeinanderfolgende Siedlungsphasen erhalten. Daneben treten Sprünge der Kalibrationskurve auf, die sich auf *annähernd* zeitgleiche Proben (mit Altersunterschieden in den Jahrringen von Holzkohlen 20-100 Kalenderjahre) so auswirken, daß die ^{14}C -Daten an Proben aus ein und der gleichen Schicht plötzlich anfangen, ohne erkennbaren archäologischen Grund ebenfalls hin und her zu springen.

Infolgedessen steht man bei der Auswertung von archäologischen ^{14}C -Daten vor der Schwierigkeit, daß auch statistisch identische ^{14}C -Daten von unterschiedlich alten Proben stammen können, und damit häufig nur wenig aussagen. Umgekehrt sind die steilen Bereiche der Kalibrationskurve, wie auch die wenigen Fenster möglicher genauer Ablesungen, die man bei der Probenentnahme also bevorzugt zu beproben versuchen wird, eng mit dem Problem der Identifikation von statistischen Extremwerten verbunden. So steht die Logik der Daten entgegen der naiven Erwartung, wonach die ^{14}C -Zeit immer nur vorwärts läuft. Die Methode des Monte Carlo Wiggle Matching nimmt diese Einsichten ernst, und verschiebt die Beweislast zunächst dahin zurück, wo das Problem entstanden ist, nämlich auf den Wunsch des Archäologen, präzisere chronologische Aussagen treffen zu können, als die ^{14}C -Methode von sich aus zu liefern in der Lage ist.

Auf jeden Fall wird man, unabhängig von der verwendeten Kalibrationsmethode, im Einzelfall den

recht hohen Standardabweichungen vieler archäologischer ^{14}C -Daten gegenüber skeptisch bleiben. Diese Zurückhaltung ist zwar nicht direkt der jeweiligen Kalibrationsmethode, und auch nicht immer den ^{14}C -Messungen anzulasten, doch bleiben die Ergebnisse der ^{14}C -Kalibration weiterhin primär eine Funktion der ^{14}C -Meßgenauigkeit wie auch der Präzision der Interlaboreichung. Betrachten wir dazu erneut die Tabelle A mit den gewichteten Mittelwerten der ^{14}C -Daten der drei Siedlungen, so wird erkennbar, daß die fehlerverkleinerte Standardabweichung von ± 20 BP für die Daten von Inden 1 ($N=11$) einerseits längst schon in der Größenordnung der zeitlichen Streuung der datierten Holzkohlen bzw. der überlieferten Jahrringe liegt. Dieser Fundort ist, in Relation zu den chronologischen Eigenschaften der datierten Proben, (erfreulicherweise) fast schon überdatiert. Andererseits liegt die Standardabweichung von ± 20 BP auch schon gefährlich nahe an der potentiellen Grenzgenauigkeit der Moderneichung des Kölner-Laboratoriums. Wie in Manning und Weninger (1992) ausgeführt wird, besitzen die Standardabweichungen der Kölner ^{14}C -Messungen Error-Multiplier zwischen 1.2 und 1.4. Die Daten sind aller Voraussicht nach im Rahmen von ± 10 BP korrekt geeicht, bezogen auf die Laboratorien Seattle und Belfast (STUIVER UND PEARSON 1986). Diese zusätzlichen Fehlerquellen, die in den Fallstudien zunächst nur aus Bequemlichkeit nicht berücksichtigt wurden, verblassen offenbar im Vergleich zu den archäologischen Fehlerquellen. Diese sind teilweise biologischen Ursprungs ("altes Holz") und kaum korrigierbar, aber trotzdem abzuschätzen.

Aus archäologischer Sicht problematischer, im Vergleich zu den laborspezifischen Fehlerquellen, erscheint nämlich die Grundannahme der dargelegten Simulationen zu sein, nach der die Proben eine zeitliche Gleichverteilung repräsentieren. Diese Annahme ist keinesfalls selbstverständlich, und dient auch nur zur Vereinfachung der Berechnungen. So wäre es durchaus denkbar, daß die datierten Proben zeitliche Schwerpunkte besitzen. Wir erwarten aus folgenden Gründen einen Schwerpunkt der Daten vor allem in Richtung "zu alt", und das natürlich zunächst aus dem einfachen Grund, weil die Besiedlung stattfindet, nachdem die Häuser gebaut wurden. Gerade in der Anfangsphase einer Siedlung, also beim Bau neuer Häuser, besteht ein beträchtlicher Bedarf an frischem Bauholz, das erst später verbrannt und als Holzkohle manchmal in situ abgelagert wird. Ein mögliches

Beispiel wäre das Datum KN-I.517 (5820 ± 45 BP) aus dem Pfostenloch von Haus III in Inden 1. Ferner besitzen die gefällten Bäume Jahrringe mit quadratisch variierender Breite, so daß deren ^{14}C -Gehalt ebenfalls nicht gut mit der Modellannahme einer Gleichverteilung übereinstimmen wird. Darüber hinaus zeigen dendrotypologische Studien, daß die verfügbare Menge an Hölzern aus Rodungen zur Anlage der ersten Wirtschaftsflächen häufig ausreicht, um abgelagert zu werden, woraus dann in späteren Siedlungsphasen Häuser gleichen Dendroalters gebaut werden. Im Kontext solcher Möglichkeiten betrachtet, würden nicht einmal jahrgenaue Dendrodaten an den überlieferten Holzkohlen ausreichen, um die von Biermann gestellte Frage eindeutig zu beantworten. Was die ^{14}C -Methode angeht, so hängen die Resultate jedenfalls, wie aufgezeigt wurde, zuletzt von einigen wenigen ^{14}C -Messungen ab, wobei es sich bei den beobachteten (wenigen) Extremwerten nicht um radiometrische "Ausreißer" handeln muß.

Ausblick

Die dargelegte Technik des (archäologischen) *Proben-Resampling*, und das heißt in der vorgelegten Studie die theoretische Simulation einer Vielzahl möglicher Holzkohlesequenzen, erscheint trotz dieser Vorbehalte einigermaßen vielversprechend zu sein. Jedenfalls läßt sich diese Technik weiter verallgemeinern.

So wurde oben das Problem angesprochen, daß derzeit noch keine allgemein akzeptierte Methode (bzw. mathematischer Formalismus) zur Bestimmung der \pm -Fehlers des dendrochronologischen Wiggle Matching existiert. Diese Frage läßt sich nun rein experimentell, wie zu erkennen ist, mit Hilfe des *Daten-Resampling* klären. Dazu wird man die ^{14}C -Daten einer "floating" Dendroserie mit Hilfe eines Gauß'schen Zufallsgenerators nach Maßgabe der vorliegenden Standardabweichungen stochastisch variieren, zum Beispiel 1000 mal, und dabei die statistische Streuung der berechneten Endjahre beobachten. Daraus läßt sich auf experimentellem Weg, eine entsprechende Rechnerleistung vorausgesetzt, eine 68%-Standardabweichung (oder 95%-Intervall) zur Annahme des Best-fit-Endjahres ermitteln.

Mit Hilfe einer Kombination von stochastischen Techniken des *Proben- und Daten-Resampling*

schließlich, läßt sich auch die Wirkung anderer physikalischer und statistischer Parameter, die beim 'Wiggle Matching' von Interesse sind, ermitteln. So lassen sich insbesondere auch Fehler simulieren, die im Dendrolaboratorium beim mechanischen Abtrennen der sowohl vertikal wie auch horizontal ausgebildeten Jahrringe auftreten können. Solche Fehlerquellen, die vergleichsweise klein und deswegen nur schwer zu fassen sind (\pm einige Jahre), spielen für die Eichung von ^{14}C -Hochpräzisionslaboratorien, und damit zusammenhängend auch für die Untersuchung möglicher regionaler Feinstrukturen des natürlichen Kohlen-

stoffkreislaufs eine zunehmend wichtige Rolle.

Anmerkung

¹ Die Berechnungen der Chi-Quadrat-Wahrscheinlichkeiten für gewichtete Mittelwerte wurden mit dem von S. Robinson (Geological Survey, Munlo Park, Californien) erstellten Rechenprogramm <Statave> durchgeführt. An dieser Stelle möchte ich Steven Robinson sehr herzlich für die Bereitstellung des Programms danken.

Tabelle der ^{14}C -Daten

Fundplatz Hasselsweiler

Der Fundort liegt in der Gemeinde Titz, Kreis Düren, Nordrhein-Westfalen. Datiert wurden vier Proben, die im Rahmen einer systematischen Grabung in den Jahren 1977-1978 vom Grabungstechniker M. Wolters geborgen wurden.

Literatur: Bonner Jahrbücher 1979, 1979. Untersuchungen zur neolithischen Besiedlung der Aldenhovener Platte IX, 386, Abb. 13.

KN-2992: 5930 \pm 65 BP

Fundstelle HW 2 A 594-9

Archäologische Zuweisung: Großgartacher Kultur

Die Probe umfaßt Holzkohle der Laborqualität A ("sehr gut", große Brocken) mit Eingangsgewicht 18 g. Die Probe entstammt einer Holzkohleschicht auf verziegeltem, in situ gebrannten Grubenboden mit Rotlehm. Zwischen Probe und Grubenboden liegt eine dünne, fast schwarze Lage aus tonigem Löß. Die Probe wurde während der Grabung geschlämmt. Die Holzkohlepartikel wurden während der Grabung ausgelesen und von Wurzeln befreit.

KN-2993: 5850 \pm 150 BP

Fundstelle HW 2 A 165-19

Archäologische Zuweisung: "Großgartacher Kultur/ Stufe Planig-Friedberg der Rössener Kultur"

Datiert wurde verkohltes Getreide der Qualität B (gut erkennbare Kohlenstoffpartikel, gemischt mit Bodenanteilen) mit Eingangsgewicht 83 g. Die Schicht mit verkohltem Material lag unmittelbar auf, und teilweise vermengt mit einer Schicht aus Rotlehmbröckeln, die den unteren Teil der veriegelten Grube bilden.

KN-2994: 5840 \pm 140 BP

Fundstelle HW 2 A 165-54

Es handelt sich um verkohlte Getreidekörner der Qualität B.

Die Schicht der Probe liegt über der den Boden bedeckenden Schicht.

KN-2995: 5830 \pm 60 BP

Fundstelle 165-56

Archäologische Zuweisung: "Großgartacher Kultur/ Stufe Planig-Friedberg der Rössener Kultur".

Bei der Fundstelle 165-56 handelt es sich um eine Siedlungsgrube der Breite 230 cm, mit erhaltener Tiefe 40 cm. Die Grube lag unter einer 50-60 cm tiefen Pflugzone. Die Probe der Qualität B hatte ein Eingangsgewicht 93 g, und umfaßt verkohlte Getreidekörner vermischt mit Holzkohlelagen. Es handelt sich um eine Linse auf der Grubensohle, vermischt mit Rotlehm.

Fundplatz Inden 3

Der Fundplatz liegt im Kreis Düren, bei Jülich, Nordrhein-Westfalen. Die Proben entstammen einer Notbergung des Jahres 1968, die unter der Leitung von R. Kuper und I. Schröter durchgeführt wurde. Die Proben wurden von J. Eckert und P. Stehli im Jahr 1969 zur Datierung eingereicht.

KN-2649: 5630 ± 145 BP
Grube 1,
Archäologische Zuweisung: "jüngeres Rössen"

Holzkohle mit Gewicht 10 g.

KN-2650: 5830 ± 75 BP
Grube 2
Zuweisung: älteres Rössen

Holzkohle mit Gewicht 45 g, aus dem gesamten Bereich der Grube 2 zusammengestellt.

KN-2651: 5870 ± 60 BP
Obere Fundschicht Quadrant e, Teilgrube 3
Zuweisung: älteres Rössen

Fundplatz Inden 1

Ausgräber R. Kuper, geborgen 27. 6. 1965

KN-I.190: 5840 ± 45 BP
Stelle 6, Profil A-B
Zuweisung: "älteres Rössen" (?)

Die Probe mit Eingangsgewicht 194 g entstammt einer Grubenfüllung mit Holzkohle und einzelnen Getreidekörnern.

KN-I.330: 5790 ± 50 BP
Pfosten aus Bau XXIII, Stelle 741c, Profil E-F.

Es handelt sich um Holzkohle mit Gewicht 122 g

KN-I.345: 5700 ± 75 BP
Grube Stelle 1175d, Schichten Ia und Ib
Zuweisung: Rössen

Aus einer holzkohlereichen Fundschicht des Grubenkomplexes wurden in 50-70 cm unter der Oberfläche insgesamt 200 g Holzkohle entnommen.

KN-2652: 5820 ± 60 BP
Grube 75, Teilgrube 5
Zuordnung: Rössen

Die Probe entstammt einer holzkohlereichen Fundschicht in 150-170 cm Tiefe unter der Oberfläche in Teilgrube 5 eines Grubenkomplexes.

KN-2653: 5830 ± 65 BP
Grube 17, Quadrant d
Zuweisung: jüngeres Rössen
Holzkohle 80 g

Etwa 50-90 cm unter der Oberfläche in der Fundschicht der Grube.

Die Probe mit Gewicht 305 g wurde aus Holzkohlen der Schichten Ia und Ib zusammengestellt.

KN-I.517: 5820 ± 45 BP
Pfostenloch Stelle 41 des Hauses III
Zuweisung: Rössen

Die Holzkohleprobe mit Gewicht 416 g wurde einem Pfostenloch 30-60 cm unter der Oberfläche entnommen.

KN-I.520: 5750 ± 50 BP
Grube Stelle 493A
Zuweisung: Rössen.

Die Holzkohle mit Gewicht 407 g wurde der Grube Stelle 493a 30-90 cm unter der Oberfläche entnommen. Diese Grube ist "archäologisch mit Grube 307 verbunden".

KN-I.521: 5900 ± 40 BP
Grube Stelle 710d
Zuweisung: Rössen

Es handelt sich um Holzkohle entnommen 90 cm unter der Oberfläche.

KN-I.524: 6180 ± 50 BP
Grube Stelle 985
Zuweisung: Rössen

Sehr saubere Holzkohle mit Gewicht 15 g, 80-100 cm unter der Oberfläche entnommen. Die Grube Stelle 985 wird überschritten von den Bauten XXVII und XXVIII.

KN-I.525: 5880 ± 65 BP
Grube Stelle 1245A, Schicht 1a
Zuweisung: Rössen

Sehr saubere Holzkohle mit Gewicht 18 g, 80-120 cm unter der Oberfläche.

KN-I.526: 5720 ± 90 BP
Grube 1245 A
Zuweisung: Rössen

Saubere Holzkohle mit Gewicht 1 g, entnommen 70-90 cm unter der Oberfläche.

KN-I.527: 5800 ± 95 BP
Grube Stelle 1277
Zuweisung: Rössen

Sehr saubere Holzkohle mit Gewicht 11 g.

KN-I.523: 6010 ± 285
Stelle 880 und Stelle 980
Wandgräben von Bau XXVIII
Zuweisung: Rössen

Mischprobe aus verschiedenen Stellen mit wenig Holzkohle, viel Löß und gebrannter Lehm, 40-100 cm unter der Oberfläche.