

Multivariate Analyse subfossiler Zuwachs-Zeitreihen

Thomas Riemer

Systematika GmbH Heidelberg

Abstract

In den Jahrringbreiten subfossiler Hölzer sind Informationen über klimatische Bedingungen während langer Zeiträume gespeichert. Die statistische Analyse von Zeitreihen des mittleren Zuwachses erlaubt Rückschlüsse auf klimatische Schwankungen und Trends in der Vergangenheit. Dabei sind die langfristigen Änderungen angesichts der heute diskutierten, globalen Klimaänderungen zwar besonders interessant, aber auch methodisch gesehen besonders heikel. Ein neuer Ansatz betrachtet nicht nur die absoluten Durchmesserzuwächse, sondern auch die Korrelationen zwischen verschiedenen regionalen und standörtlichen Herkünften als Indikator für klimatisch bedingten Streß. Für die Berechnung der Korrelationen wurden Maximum-Likelihood-Schätzer hergeleitet und ein SAS-Programm mit AF-Oberfläche entwickelt, welches alle notwendigen Ein- und Ausgabeoperationen, Standardisierungsschritte und optimierten Parameterschätzungen durchführt sowie dem Anwender durch graphische Darstellungen und explorative Techniken eine tiefergehende Analyse der Ergebnisse ermöglicht.

Danksagung

Die Arbeit wurde durch das EG-Umwelt- und Klimaprogramm ADVANCE-10K (ENV4-CT95-0127) gefördert.

Einleitung

Die Beobachtung, daß sich Klima nicht nur über den Raum, sondern auch über die Zeit ändert, ist seit etlichen Jahren Ausgangspunkt für zahlreiche Forschungsaktivitäten. Diese werden insbesondere durch die Erkenntnis angetrieben, daß der Mensch selber in der Lage ist - wenn auch ungewollt - Klima zu beeinflussen. Im Mittelpunkt des Interesses stehen hierbei Erhöhungen in der atmosphärischen CO₂-Konzentration und der damit verbundenen globalen Temperaturerhöhung (HOUGHTON, 1994). Wenn auch die grundlegende Existenz solcher Entwicklungen außer Frage steht, bestehen Meinungsverschiedenheiten über ihre konkreten Auswirkungen sowie ihre praktische Relevanz angesichts der natürlichen Konzentrations- bzw. Temperaturschwankungen. Deren Häufigkeit und Stärke und ist aber auch nur ungefähr oder für die jüngste Vergangenheit bekannt. Die Rekonstruktion klimatischer Variabilität über längere Zeiträume leistet somit neben Politik, Grundlagenforschung, neuen Meßtechniken, Modellrechnungen usw. einen ebenfalls wichtigen Beitrag in der Diskussion über globale Klimaänderungen (JONES et al., 1996).

Unter allen verfügbaren Datenquellen nehmen die Jahresringe der Bäume insofern eine Sonderstellung ein, als sie vergleichsweise einfach zu erheben sind und deshalb schon seit langem, nämlich seit Beginn des 20. Jahrhunderts für Rekonstruktionszwecke eingesetzt worden sind (FRITTS, 1976). Zwischen der Gewinnung von Holzproben und der Klimamodellierung liegen allerdings etliche methodische Schwierigkeiten. Dieser Artikel beschreibt Lösungsansätze für eines der Hauptprobleme, nämlich der Gewinnung verlässlicher Kovariablen für die Rekonstruktion langfristiger Klimaänderungen.

Daten

Im Rahmen des EG-Umwelt- und Klima-Forschungsprojektes „ADVANCE-10K“ wurden zahlreiche Zuwachsreihen aus dem mittel- bis nordeuropäischen Verbreitungsgebiet der Eiche (*Quercus spec.*) zusammengetragen. Für die folgende Untersuchung wurden Holzproben von

„subfossilen“, also nicht mehr rezenten oder lebenden, aber noch nicht fossilen Eichen gewählt. Die später vorgestellten Beispielrechnungen basieren auf einem Teil des insgesamt verfügbaren Datennetzes, ca. 20000 Eichenreihen aus verschiedenen, im Großraum von Nord- und Ostsee liegenden Regionen. Von den einige tausend Jahre abdeckenden Zeitreihen wurde zunächst die letzten 2000 Jahre betrachtet.

Jährliche Durchmesserzuwächse sind als Kovariablen in Klimamodellen ideal,

- wenn sie von Standorten stammen, auf denen ein einziger oder nur wenige Umweltparameter das Wachstum begrenzen,
- wenn nichtklimatische Wachstumsfaktoren wie Alterung, Insektenbefall oder Konkurrenz keine Bedeutung haben oder leicht aus den Daten herauszurechnen sind,
- wenn die Zuwächse für lange Zeiträume bekannt sind, d.h. wenn die Hölzer auf natürlichem Wege langfristig konserviert wurden.

Leider sind die Eigenschaften selten in Kombination anzutreffen, und die Erfüllung aller drei Bedingungen kommt praktisch nicht vor. Abb. 1 verdeutlicht, daß selbst die Zuwächse temperatursensitiver Gebirgsfichten nur bedingt mit Temperaturdaten korreliert werden können.

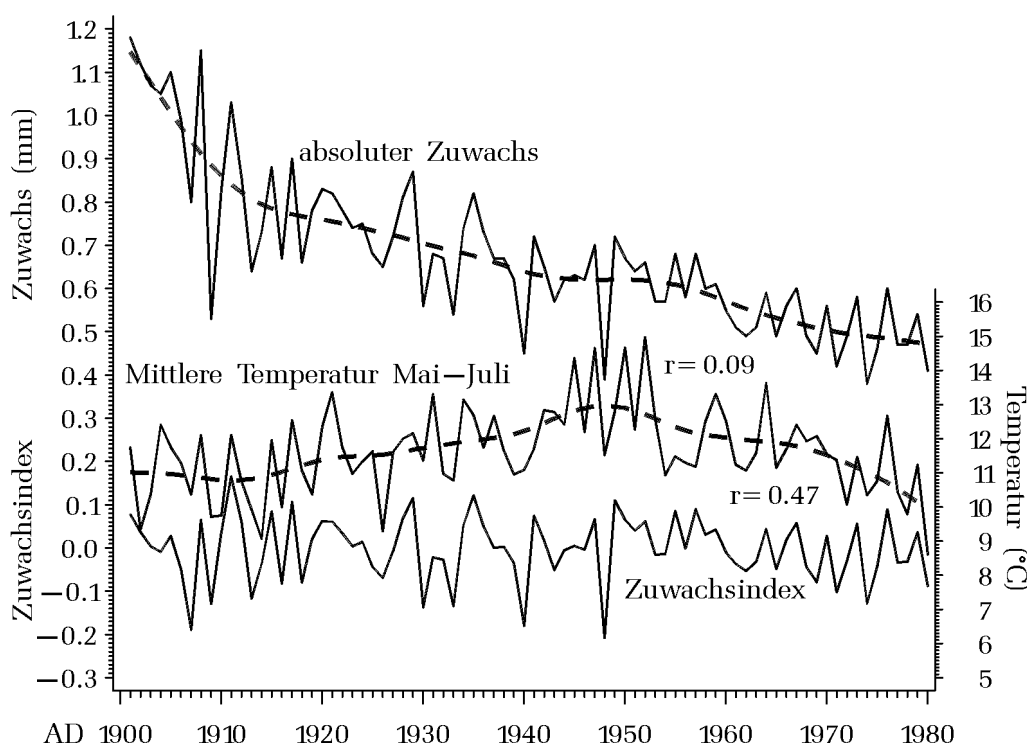


Abb. 1: Korrelation von Zuwachs und Sommertemperatur bei einer Gebirgsfichte (nach LINGG, 1986)

Die Rohdaten weisen eine Korrelation von 0.09 zu den mittleren Sommertemperaturen auf, während die trendbereinigten Zuwachsindizes mit 0.47 wesentlich stärker korrelieren. Diese Korrelation ist aber bei Einzelbäumen keineswegs die Regel und nur bei sorgfältig gesuchtem Datenmaterial möglich. Das Beispiel zeigt immerhin, daß die Rekonstruktion von kurzfristigen Schwankungen und Extremen bei bestimmten klimatischen Parametern möglich ist (Vgl. auch SCHWEINGRUBER, 1988a). Der längerfristige Trend in den Zuwachsreihen (Abb. 1 oben, gestrichelt) ist aber zum überwiegenden Teil nicht klimatisch bedingt und erschwert die Auswertung erheblich. Weder Ignorieren noch Eliminieren dieser Trends ist

eine Lösung, da man in dem einen Fall rein künstliche Ergebnisse erhielte, im anderen Fall auch die Information über langfristige Temperaturänderungen aus den Ringbreiten lösche. Die Schwierigkeiten sind selbst bei solch ausgesuchtem Datenmaterial erheblich, und sie werden keineswegs geringer, wenn man für Zwecke der großräumigen und langfristigen Klimarekonstruktion mit ausgedehnten Datennetzen wie dem europäischen Eichennetz arbeitet. Dem Vorteil seines Umfangs steht als Nachteil gegenüber, daß in den Zuwächsen überregionale Klimaeinflüsse nicht sehr stark ausgeprägt sind. Es ist von den Herkünften und Standorten der Eichen sowie der zeitlichen Dichte der Belegungen („Wiederholungszahlen“) vergleichsweise heterogen zusammengesetzt. Die Länge der Zeitreihen und die Größe der durch sie abgedeckten Regionen eröffnen aber so weitreichende Auswertungsmöglichkeiten, daß der Forschungsaufwand zur Überwindung von methodischen Schwierigkeiten mehr als lohnend erscheint.

Methoden

Dendrochronologische Datierung

Die einzelnen Eichenholzfunde wurden mit dendrochronologischen Methoden weiterverarbeitet, d.h. ihre Oberfläche präpariert, die Jahrringe vermessen und jahrgenau datiert (SCHWEINGRUBER, 1988b). Eine neue Jahrringreihe gilt als datiert, wenn bei einer bestimmten zeitlichen Zuordnung die Ähnlichkeit zu einer bereits datierten Referenzreihe („Chronologie“) nach statistischen Maßzahlen und gutachterlicher Entscheidung so hoch ist, daß zweifelsfrei feststeht, daß die Jahrringe zeitgleich gebildet wurden. Neu datierte Reihen können ihrerseits in die Chronologie übernommen („dazugemittelt“) werden und so zu ihrer Stabilisierung oder zeitlichen Ausdehnung beitragen. Nach diesem dendrochronologischen Prinzip konnten in Europa Eichenreihen zusammengestellt werden, die je nach Region bis zu ca. 10000 Jahre zurückreichen.

Datenaufbereitung

Die Rohdaten werden zur statistischen Absicherung der Datierung oder zur weiteren dendroklimatologischen Modellbildung in verschiedenen Schritten aufbereitet. Ziele dabei sind die Identifizierung und evtl. Eliminierung von Autokorrelation und Alterungstrends, die Beseitigung der unerwünschten Rechtsschiefe und Kurtosis in den Daten sowie eine Standardisierung der Varianzen. Wichtig ist die rechnerische Betonung der besonders schmalen Jahrringe, die auf absoluter Skala zwar nicht weit vom Durchschnittswert abweichen, denen aber als Indikator für klimatische Extremjahre eine große Bedeutung zukommt. Hier steht eine breite Palette von Transformationen, Filtern und Modellen zur Verfügung, die je nach Datenmaterial und Auswertungszielen individuell angepaßt werden müssen (RIEMER, 1994). Für das Anwendungsbeispiel in diesem Artikel kamen logarithmische Transformationen, variabel-parabolisch gewichtete, gleitende Mittelwerte und negativ-exponentielle Minderengewichtungen der Jugendphasen einer jeden Zuwachsreihe zum Einsatz.

Statistische Methoden der Klimarekonstruktion

Auf diesem Gebiet dominieren klar die linearen Modelle: multiple und multivariate Regressionen, Hauptkomponenten-Transformationen und kanonische Korrelationen. Besonders beliebt sind die „response functions“, schrittweise Regressionen auf die Hauptkomponenten aller betrachteten Regressoren (FRITTS, 1976). Während die Rekonstruktion jährlicher oder kurzfristiger Klimaschwankungen, wie bereits ausgeführt, damit vergleichsweise einfach gelingt, gibt es noch keine zufriedenstellende Möglichkeit, mit solchen Methoden Trends oder längere Zyklen irgendwelcher Umweltvariablen zurückzuverfolgen. Das Grundproblem liegt

hier darin, daß die Trends in den Durchmesserzuwachsen der Einzelbäume recht individuell ausfallen und überwiegend nicht klimatisch bedingt sind. Um diese Schwierigkeit zu umgehen, werden verschiedene Ansätze verfolgt:

- Nichtparametrische Methoden verwenden reduzierte Informationen wie Spitzenwerte (Extremjahre), Anomalien (Depressionsphasen) oder klassierte Werte (+/- Abfolgen), um dadurch möglichst altersunabhängige Indizes für Umwelteinflüsse zu erhalten. Das Problem dabei ist, daß die große Masse des verfügbaren Datenmaterials eher schwach mit Klima korreliert und daß deshalb jeder Informationsverlust vermieden werden muß. Ferner nimmt man sich mit derartigen Methoden in der Regel die Möglichkeit, Zuwachs und Klima in eine metrische Beziehung zu setzen.
- Eingeschränkte Trendbereinigungen versuchen auf verschiedenem Wege, im Zweifelsfall lieber zuviel als zuwenig Alterstrend in den Zeitreihen zu belassen. Die Konsequenz dabei ist allerdings, daß bei eventuell gefundenen „Ergebnissen“ nie klar ist, ob es sich nicht um rein rechnerische Folgen dieser unvollständigen Trendbereinigung handelt.
- Die Bildung von Altersklassen ermöglicht eine Auswertung, in der immer nur altersmäßig vergleichbare Proben zusammen verrechnet werden. Sie berücksichtigt aber nicht, daß auch im gleichen Alter je nach Standort unterschiedliche mittlere Zuwächse geleistet werden.
- Regionale Trendmodelle gehen von der vereinfachenden Annahme aus, daß Bäume aus der gleichen Region einen einheitlichen Alterstrend haben, der über die Zeit konstant bleibt, und daß sich der konkrete Zuwachs aus Überlagerung dieses konstanten Trends und des variablen Klimas ergibt. Dann ließe sich der Trend aus allen Proben einer Region schätzen und aus den Zeitreihen herausrechnen (z.B. mit einem linearen Modell). Klar ist jedoch, daß dies nicht ganz der Realität entspricht. Es bleibt die Unsicherheit, wieviel von evtl. erhaltenen Ergebnissen auf derartig vereinfachende Annahmen zurückzuführen ist (JONES, et al., 1996).

Idee: Korrelationen statt Mittelwerte

Ausgangspunkt für diese neue Strategie ist die Beobachtung, daß zwischen regional gebildeten Eichenchronologien auch über große Entfernungen, d.h. Tausende von Kilometern, signifikante Korrelationen existieren (LEUSCHNER et al., in Vorbereitung). Damit ist klar, daß das Wachstum der Bäume auch durch überregionale Klimafaktoren beeinflusst wird. Derartige Korrelationen scheinen aber nicht konstant zu verlaufen. Variable Korrelationen können zum einen durch heterogene Zusammensetzungen aus unterschiedlich vielen oder unterschiedlich sensitiven Bäumen verursacht werden, zum anderen durch variablen Streß, also durch sich änderndes Klima. Wenn es gelingt, die ersten beiden Ursachen auszuschalten, wären also höhere Korrelationen ein direktes Indiz für ein ungünstigeres Klima, insbesondere dann, wenn diese Erhöhung von durchschnittlich geringeren Zuwächsen begleitet wird. Durch gezielten Paarvergleich, z.B. von Bäumen aus gemäßigten Regionen mit solchen aus besonders kalten bzw. trockenen Regionen, könnte man die Art der Klimaänderung weiter eingrenzen und im günstigsten Falle, unter Zuhilfenahme von rezenten Zuwachsdaten und Klimadaten für die Modellanpassung, sogar quantifizieren. Zumindest lassen sich hohe bzw. niedrige Korrelationen einfach und direkt interpretieren, so daß man auch ohne Klimadaten zu naheliegenden Hypothesen gelangt. Ferner sind altersabhängige Trends in den Korrelationen bei den Einzelbäumen weit weniger dominant als die Trends in der mittleren Ringbreite, so daß auch die Chancen größer sind, langfristige Klimaänderungen bzw. Verschiebungen von Klimazonen zu rekonstruieren. Der Trend in der Ringbreite spielt für die Korrelationsberechnung keine nennenswerte Rolle und kann vollständig eliminiert werden, ohne daß Langzeit-Entwicklungen in den Korrelationen verloren gehen. Das Problem bei diesem Ansatz bleibt die hetero-

gene Zusammensetzung der Probenkollektive. Variablen Mischungsanteilen unterschiedlicher Standorttypen kann man durch statistische Analysen begegnen, die im weitesten Sinne zu den Clusterverfahren gehören (RIEMER, 1994). Wie das Problem variabler Probenzahlen zu lösen ist, wird im folgenden beschrieben.

Varianzkomponenten im „Signal+Noise“-Modell

Die einfachste Weg, überregionalen Klimastreß anhand von Zuwachs-Zeitreihen zu quantifizieren, besteht in der gleitenden Berechnungen von Kreuzkorrelationen zwischen den regional gemittelten Zuwachsreihen. Die empirischen Korrelationen zwischen diesen Chronologien hängen aber nicht nur von dem klimatisch bedingten Streß ab, sondern auch von dem Verhältnis von individueller Varianz der Bäume zur Varianz der gemeinsamen Komponente sowie den Belegungen (Wiederholungszahlen) der Chronologien (Abb. 2).

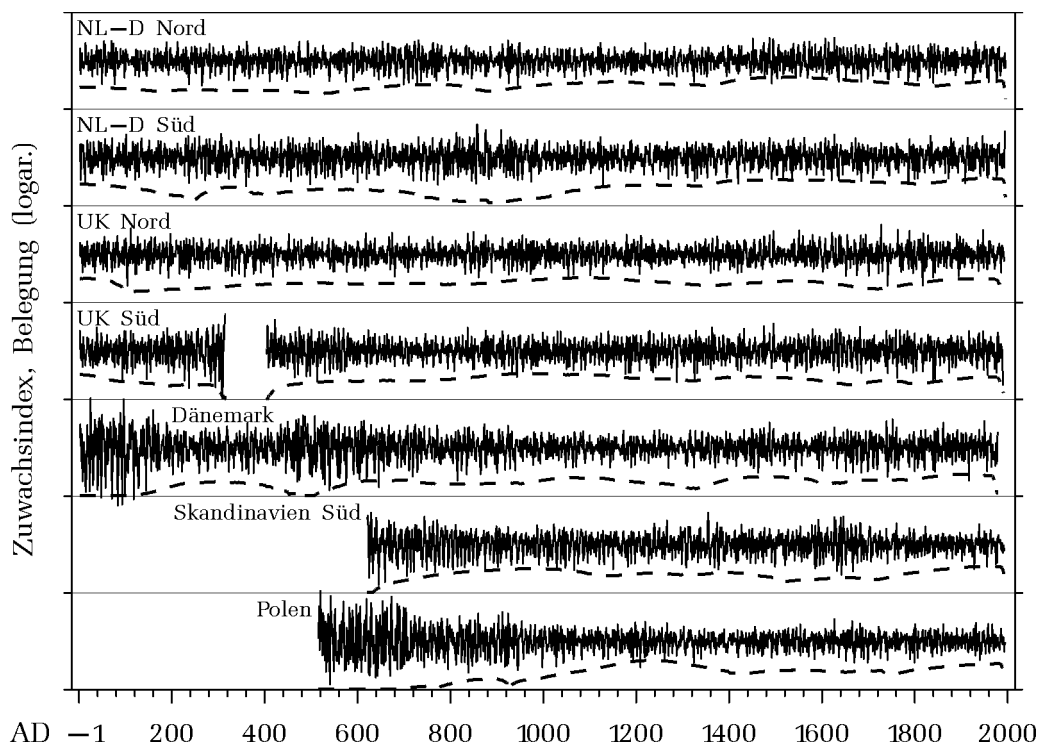


Abb. 2: Abnahme der Varianz mit steigender Belegung (gestrichelt) bei 7 europäischen Jahrringchronologien

Der einfache Ansatz, Korrelationen zu berechnen, führt also zu künstlichen Ergebnissen, in denen natürliche Umwelteinflüsse und rechnerische Scheineffekte vermischt sind. Zwar immer noch vereinfacht, aber schon deutlich realistischer lassen sich die Zusammenhänge im „Signal+Noise“-Modell erfassen.

Der Einfachheit halber gehen wir von stationären, normalverteilten Zeitreihen ohne Autokorrelation aus, was in der Praxis mit geeigneten Aufbereitungsverfahren hinreichend genau erreicht werden kann (RIEMER, 1994). Ferner genügt zunächst die Betrachtung von nur zwei Regionen oder allgemein „Klassen“ K_x und K_y mit n_x bzw. n_y Einzelreihen x_j , $j=1, \dots, n_x$ und y_j , $j=1, \dots, n_y$. Beginnjahr und Länge der Einzelreihen sind in der Regel verschieden; ihre individuellen zeitlichen Indextmengen seien T_{xj} , $j=1, \dots, n_x$ bzw. T_{yj} , $j=1, \dots, n_y$. Laut Modell

bestehen die Zeitreihen in jedem Jahr t aus einem klassentypischen Signal s_x bzw. s_y und individuellen Abweichungen e_{xj} bzw. e_{yj} :

$$x_{jt} = s_{xt} + e_{xjt}, \quad j=1, \dots, n_x, \quad t \in T_{xj} \quad \text{und} \quad y_{jt} = s_{yt} + e_{yjt}, \quad j=1, \dots, n_y, \quad t \in T_{yj}$$

Die Abweichungen sind normalverteilt mit Erwartungswert 0 und unabhängig von allen anderen Variablen im Modell. Ihre Varianz ergibt sich aus einer konstanten „Noise-“ oder Restvarianz σ_n^2 und einer bekannten Gewichtung w :

$$e_{xjt} \sim N\left(0, \frac{\sigma_{nx}^2}{w_{xjt}}\right), \quad e_{yjt} \sim N\left(0, \frac{\sigma_{ny}^2}{w_{yjt}}\right)$$

Im typischen Fall ist $w_{xjt}=w_{yjt}=1$, kann aber auch in einzelnen Jahren niedriger angesetzt werden, um z.B. unsichere Werte in einer Zeitreihe minderzugewichten. In anderen Fällen können Gewichte $w>1$ eine schon erfolgte Aggregation von w Einzelreihen repräsentieren. Die Signalreihen s_x und s_y sind nicht direkt beobachtbar; sie werden wiederum als Folgen identisch normalverteilter Zufallsvariablen ohne Autokorrelation modelliert. Somit bilden s_{xt} und s_{yt} unabhängige, bivariat normalverteilte Zufallsvektoren. Ihr Erwartungswert ist wegen der Standardisierung der Daten ohne Bedeutung und wird gleich 0 gesetzt:

$$\begin{pmatrix} s_{xt} \\ s_{yt} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{sx}^2 & r_t \cdot \sigma_{sx} \cdot \sigma_{sy} \\ r_t \cdot \sigma_{sx} \cdot \sigma_{sy} & \sigma_{sy}^2 \end{pmatrix}\right)$$

Im Mittelpunkt des Interesses stehen die variablen Korrelationen r_t zwischen den unbekanntem Signalreihen. Diese werden ihrerseits durch die regionalen Chronologien, d.h. gewichtete Mittelwertreihen \tilde{s}_x , \tilde{s}_y prognostiziert. Wenn J_{xt} bzw. J_{yt} die Indexmengen der Reihen aus K_x , K_y bezeichnen, die im Jahr t Werte aufweisen, dann:

$$\tilde{s}_{xt} = \bar{x}_{t,t} = \frac{\sum_{j \in J_{xt}} (w_{xjt} \cdot x_{jt})}{\sum_{j \in J_{xt}} (w_{xjt})} = s_{xt} + \frac{\sum_{j \in J_{xt}} (w_{xjt} \cdot e_{xjt})}{\sum_{j \in J_{xt}} (w_{xjt})} \quad (\tilde{s}_{yt} \text{ analog})$$

Diese Schätzung ist wiederum normalverteilt mit Erwartungswert 0 und reduzierter Varianz:

$$\text{Var}(\tilde{s}_{xt}) = \text{Var}(s_{xt}) + \frac{\left(\sum_{j \in J_{xt}} w_{xjt}^2 \cdot \text{Var}(e_{xjt})\right)}{\left(\sum_{j \in J_{xt}} w_{xjt}\right)^2} = \sigma_{sx}^2 + \frac{\sigma_{nx}^2}{\sum_{j \in J_{xt}} w_{xjt}}$$

($\text{Var}(\tilde{s}_{yt})$ analog) Bezeichnet man die jährliche Gewichtssumme $\sum_{j \in J_{xt}} w_{xjt}$ mit $w_{x,t}$ und das

Verhältnis von Signal- und Restvarianz mit $\psi_x = \frac{\sigma_{sx}^2}{\sigma_{nx}^2}$, dann erhält man varianzstandardisierte Chronologien aus:

$$x_t^* = \frac{\bar{x}_{t,t}}{\sigma_{nx} \cdot \sqrt{\psi_x + 1/w_{x,t}}} \quad (y_t^* \text{ analog})$$

wobei für σ_{nx} und ψ_x geeignete Schätzwerte eingesetzt werden müssen. Da die s_x und e_{xj} nicht direkt beobachtbar sind, können die Varianzkomponenten auch nicht mit einfachen Schätzern ermittelt werden. Hier bietet sich vielmehr das Maximum-Likelihood-Prinzip an. Da Korrelationen zwischen den x_{jt} nur innerhalb des gleichen Jahres t auftreten, ist die gesamte Likelihood innerhalb der Klasse K_x schon durch die jährlichen Kovarianzmatrizen Σ_t festgelegt. Seien $v_{1,t}, \dots, v_{n(t),t}$ die Gewichte der $n(t)$ Reihen aus K_x , die Werte im Jahr t aufweisen, d.h. $\{w_{xjt}, j \in J_{xt}\} = \{v_{i,t}, i=1, \dots, n(t)\}$, dann:

$$\Sigma_t = \begin{bmatrix} \sigma_{xs}^2 + \sigma_{xr}^2/v_{1,t} & \sigma_{xs}^2 & \cdot & \cdot & \cdot & \sigma_{xs}^2 \\ \sigma_{xs}^2 & \sigma_{xs}^2 + \sigma_{xr}^2/v_{2,t} & \cdot & \cdot & \cdot & \vdots \\ \vdots & \cdot & \cdot & \cdot & \cdot & \sigma_{xs}^2 \\ \sigma_{xs}^2 & \cdot & \cdot & \cdot & \sigma_{xs}^2 & \sigma_{xs}^2 + \sigma_{xr}^2/v_{n(t),t} \end{bmatrix}$$

$= \sigma_{xs}^2 \cdot \mathbf{1}_{n(t)} \cdot \mathbf{1}_{n(t)}^T + \sigma_{xr}^2 \cdot D_t^{-1}$, wobei $D_t = \text{diag}(v_{1,t}, \dots, v_{n(t),t})$. Die Log-Likelihood ohne konstante Summanden und den Faktor $-1/2$ hat die Gestalt

$$l^*(\sigma_{xs}^2, \sigma_{xn}^2; X) = \sum_{t \in T_x} \log |\Sigma_t| + \text{tr}(\Sigma_t^{-1} \cdot \bar{x}_t \cdot \bar{x}_t^T),$$

worin X die Matrix aller Zeitreihen aus K_x und \bar{x}_t der Spaltenvektor der Reihenwerte im Jahr t ist. Umformung ergibt

$$\Sigma_t^{-1} = \frac{1}{\sigma_{xn}^2} \cdot \left(D_t - \frac{D_t \cdot \mathbf{1}_{n(t)} \cdot \mathbf{1}_{n(t)}^T \cdot D_t}{\sigma_{xn}^2 / \sigma_{xs}^2 + \mathbf{1}_{n(t)}^T \cdot D_t \cdot \mathbf{1}_{n(t)}} \right)$$

und:

$$|\Sigma_t| = (\sigma_{xn}^2)^{n(t)} \cdot |D_t^{-1}| \cdot \left(1 + (\sigma_{xs}^2 / \sigma_{xn}^2) \cdot \mathbf{1}_{n(t)}^T \cdot D_t^{-1} \cdot \mathbf{1}_{n(t)} \right)$$

Das verbleibende Minimierungsproblem führt mit einfacher Analysis zu den beiden Maximum-Likelihood-Gleichungen

$$\sum_{t \in T_x} \left((\hat{\sigma}_{xn}^2 + w_{x,t} \cdot \hat{\sigma}_{xs}^2) \times \left(\left(\sum_{j \in J_{xt}} (w_{xjt} \cdot x_{jt}^2) \right) - n(t) \cdot \hat{\sigma}_{xn}^2 \right) \right) = \sum_{t \in T_x} \left(\hat{\sigma}_{xs}^2 \times \left(\sum_{j \in J_{xt}} (w_{xjt} \cdot x_{jt}) \right)^2 \right)$$

und:

$$\sum_{t \in T_x} (w_{x,t} \cdot (\hat{\sigma}_{xn}^2 + w_{x,t} \cdot \hat{\sigma}_{xs}^2)) = \sum_{t \in T_x} \left(\left(\sum_{j \in J_{xt}} (w_{xjt} \cdot x_{jt}) \right)^2 \right)$$

Dabei handelt es sich also um ein quadratisches Gleichungssystem mit zwei Unbekannten. In nicht ausgearteten Fällen führt genau eine reelle Lösung zur Maximierung der Likelihood. Dann haben die Varianzkomponentenschätzer die explizite Darstellung:

$$\hat{\sigma}_{xn}^2 = \frac{\sqrt{b^2 - 4 \cdot a \cdot c} - b}{2 \cdot a}, \quad \hat{\sigma}_{xs}^2 = \frac{g - w_{x..} \cdot \hat{\sigma}_{xn}^2}{q}, \quad \text{wobei}$$

$$n = \sum_{t \in T_x} n(t), \quad w_{x..} = \sum_{t \in T_x} w_{x,t}, \quad p = \sum_{t \in T_x} w_{x,t} \cdot n(t), \quad q = \sum_{t \in T_x} (w_{x,t})^2,$$

$$f = \sum_{t \in T_x} \sum_{j \in J_{xt}} w_{xjt} \cdot x_{jt}^2, \quad g = \sum_{t \in T_x} \left(\sum_{j \in J_{xt}} w_{xjt} \cdot x_{jt} \right)^2, \quad h = \sum_{t \in T_x} \left(w_{x,t} \cdot \sum_{j \in J_{xt}} w_{xjt} \cdot x_{jt}^2 \right),$$

$$a = n \cdot q - w_{x..} \cdot p, \quad b = p \cdot g + w_{x..} \cdot h - q \cdot f - w_{x..} \cdot g \quad \text{und} \quad c = g \cdot (g - h).$$

Gleitende Korrelationen

Mit Hilfe der geschätzten Varianzkomponenten lassen sich also die Chronologien – unabhängig von der Belegung – auf eine Standardnormalverteilung transformieren. Die Informationen über die Signal-Korrelationen r_t sind dann in den quadratischen Abweichungen $(x_t^* - y_t^*)^2$ enthalten. Um die Verteilung dieser Quadrate zu bestimmen, benötigt man noch die Kovarianz der standardisierten Chronologien:

$$\text{Kov}(x_t^*, y_t^*) = \frac{\text{Kov}(s_{xt}, s_{yt})}{\sigma_{xn} \cdot \sigma_{yn} \cdot \sqrt{(\psi_x + 1/w_{x,t}) \cdot (\psi_y + 1/w_{y,t})}} = r_t \times \sqrt{\frac{\psi_x \cdot \psi_y}{(\psi_x + 1/w_{x,t}) \cdot (\psi_y + 1/w_{y,t})}}$$

Die Kovarianz bzw. Korrelation hat also die Gestalt $r_t \cdot \theta_t$, wobei θ_t als „effektiver Belegungsfaktor“ des Klassenvergleiches das geometrische Mittel der effektiven Belegungsfaktoren innerhalb der Klassen ist. Diese heißen auch „Expressed Population Signal“ oder EPS. Sie streben mit steigenden Wiederholungszahlen gegen 1, und zwar um so schneller, je geringer die Noise-Varianz im Vergleich zur Signalvarianz ist. Das EPS wird in der Dendroklimatologie normalerweise über einfache heuristische, varianz- oder korrelationsanalytische Näherungsformeln geschätzt, (COOK und KAIRIUKSTIS, 1989), kann aber, wie oben ausgeführt, wesentlich effektiver und präziser von den Varianzkomponenten aus den Maximum-Likelihood-Gleichungen abgeleitet werden.

Die Quadrate $(x_t^* - y_t^*)^2$ haben offensichtlich bis auf einen konstanten Faktor eine χ^2 -Verteilung mit einem Freiheitsgrad. Wegen $E(x_t^* - y_t^*)^2 = 2 - 2 \cdot r_t \cdot \theta_t$ ist $\hat{r}_t = \frac{1 - \frac{1}{2}(x_t^* - y_t^*)^2}{\theta_t}$

ein erwartungstreuer Schätzer für r_t . In der Praxis wird dies allerdings nur approximativ gelten, da die Größen auf der rechten Seite von den unbekanntem Varianzkomponenten abhängen, für die ihrerseits Schätzwerte einzusetzen sind. Die Schätzung hat weiterhin den Nachteil, daß sie nicht auf das Intervall $[-1,1]$ beschränkt ist. Vielmehr können alle Werte zwischen $-\infty$ und $1/\theta_t$ angenommen werden. Die χ^2_1 -Verteilung besitzt zudem eine große Schiefe, so daß der mittlere quadratische Fehler (MSE) der Schätzung in einzelnen Jahren sehr groß ist. Alle diese Nachteile werden sinnvollerweise dadurch abgeschwächt bzw. aufgehoben, daß die Interpretation der Schätzer nicht getrennt in den Einzeljahren, sondern in zeitlich geglätteter oder gefilterter Form erfolgt (Abb. 3). Ferner können die Zeitreihen der \hat{r}_t wieder als Eingangsdaten für weiterführende Modelle dienen, so daß ihre approximative Erwartungstreue und bekannte Verteilung zunächst wichtigere Eigenschaften sind als ein kleiner MSE.

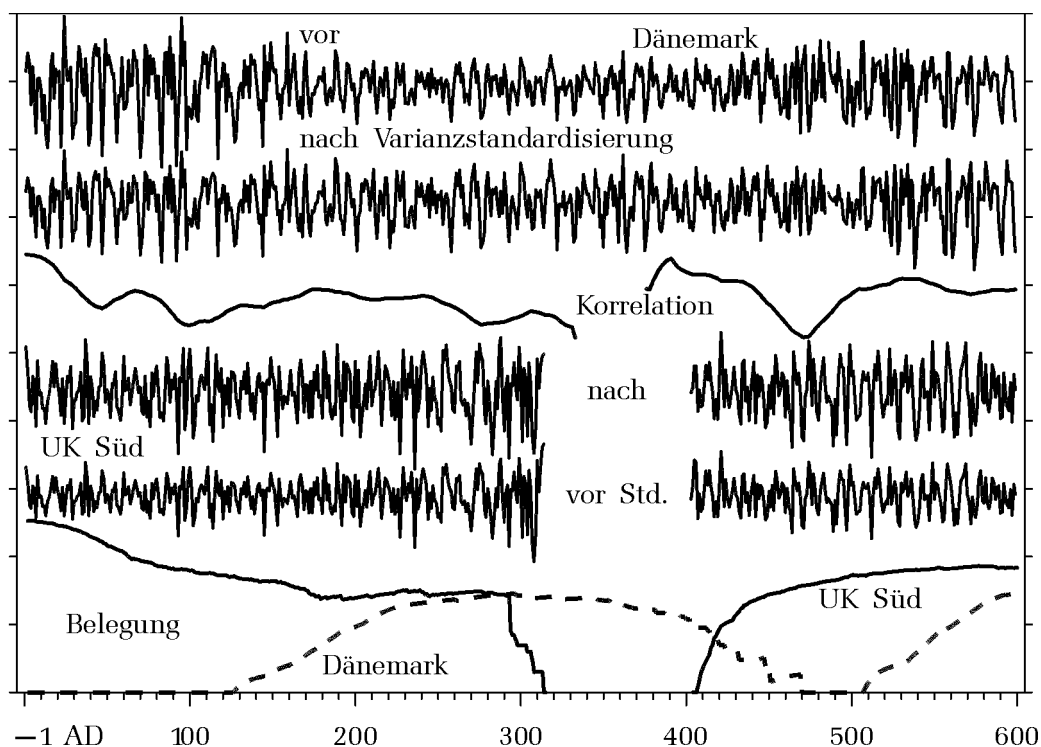


Abb. 3: Gleitende Berechnung von Kreuzkorrelationen nach Varianzstandardisierung

Die Chronologien ohne Standardisierung in Abb. 3 (außen) stellen einen Ausschnitt aus Abb. 2 dar. Die Std. führt zu einer etwa konstanten Ausschlagsstärke in den Kurven (innen), die nicht mehr von der Anzahl Einzelproben (unten) abhängt. Nach der Std. können aus den Varianzen zwischen den Chronologien mittels eines Korrekturfaktors jährlich Kreuzkorrelationen geschätzt werden, die man sinnvollerweise in geglätteter Form darstellt (Mitte).

Hauptkomponenten auf Korrelationsbasis

Sollen die Korrelationen zwischen mehreren Regionen nicht nur ausgerechnet, sondern auch gemeinsam graphisch dargestellt und in ihrer zeitlichen Dynamik interpretiert werden, führt das den Betrachter schon bei vergleichsweise wenigen Chronologien an seine Grenzen, sind doch bei m Chronologien schon $m \times (m-1)/2$ Paarvergleiche (1, 3, 6, 10, 15, 21 usw.) möglich. Hier kann eine Extraktion der Hauptkomponenten (HK) aus den Korrelationsmatrizen wertvolle Hilfestellung geben. Das Prinzip besteht darin, für die Regionen neue Koordinaten im ein-, zwei- oder dreidimensionalen Raum zu konstruieren, so daß die Abstände zwischen den Koordinaten so gut wie möglich die tatsächlichen Abstände repräsentieren, wobei der tatsächliche quadratische Abstand gerade als 1 abzüglich Korrelation definiert ist. Dann könnten, statt der $m \times (m-1)/2$ Korrelationen einfach die m Koordinaten („Scores“) der ersten HK dargestellt werden, ohne viel Information zu verlieren. Die Interpretation fällt hier auch bei mehreren Chronologien leicht, weil „weit weg“ auf der Graphik auch „verschieden im Zuwachs“ bedeutet und umgekehrt (Abb. 6)

Bei einer solchen Transformation ist zu beachten, daß verschiedene Anordnungen der Scores die gleichen Korrelationen repräsentieren können. Deshalb muß man geeignete Nebenbedingungen oder Standardisierungsvorschriften aufstellen, so daß eine leichte und durchgängige Interpretation der Scores möglich ist. Für zeitkontinuierliche Darstellungen („Kurven“) muß zudem gewährleistet sein, daß kleine Veränderungen in den Korrelationen auch nur kleine Verschiebungen in den Scores zur Folge haben. Eine sinnvolle Nebenbedingung lautet z.B.:

„Der Mittelwert aller m Koordinaten beträgt 0.“ Ferner ist die Kovarianzkomponente „Varianz des Mittelwertes aller Chronologien“ für die Distanzbetrachtung ohne Bedeutung, so daß ihre Elimination keinen Nachteil und nur den Vorteil hat, den Rang der Korrelationsmatrix um 1 zu verringern. So wird sichergestellt, daß die gesamte Information in maximal $m-1$ Hauptkomponenten enthalten ist. Praktisch heißt dies, die HK nicht direkt aus der Korrelationsmatrix R_t , sondern aus der projizierten Variante $Q_t = P_m \times R_t \times P_m$ zu extrahieren, wobei

$P_m = I_m - \frac{1_m \cdot 1_m^T}{m}$ die einfache Projektionsmatrix der Mittelwertstandardisierung und m die

Anzahl von Chronologien im Jahr t ist. Dann erfolgt eine Eigenwertbestimmung und die Zerlegung $Q_t = C_t \cdot D_t \cdot C_t^T$, wobei C_t die Matrix der spaltenweisen Eigenvektoren und D_t die Diagonalmatrix der absteigend sortierten Eigenwerte von Q_t ist. Die gesuchten Scores ergeben sich dann aus $D_t^{1/2} \cdot C_t^T$, worin die ersten $m-1$ Zeilen den HK und die m Spalten den Chronologien zuzuordnen sind. Abb. 6 zeigt beispielhaft die Scores der ersten HK von 7 Chronologien über ca. 1000 Jahre.

Software

Realisierung mit SAS®

Die beschriebenen Methoden wurden in Form einer SAS-Applikation programmiert. Die Einbettung in das Softwaresystem SAS wurde gegenüber einer eigenständigen Realisierung bevorzugt, um eine kurze Entwicklungszeit einhalten zu können. Die Rechenzeit war zweit-rangig, zumal sich schnell zeigte, daß auch größere Datenmengen mit der SAS-Lösung in hinreichend kurzer Zeit bewältigt werden können. Das im Anschluß beschriebene Anwendungsbeispiel benötigte z.B. auf einem nicht mehr allerneuesten Pentium-PC für 20000 Zeilen weniger als 15 Minuten. Die Anwendung hat eine graphische AF-Bedienungsfläche und läuft rein maus- bzw. menügesteuert (Abb. 4).

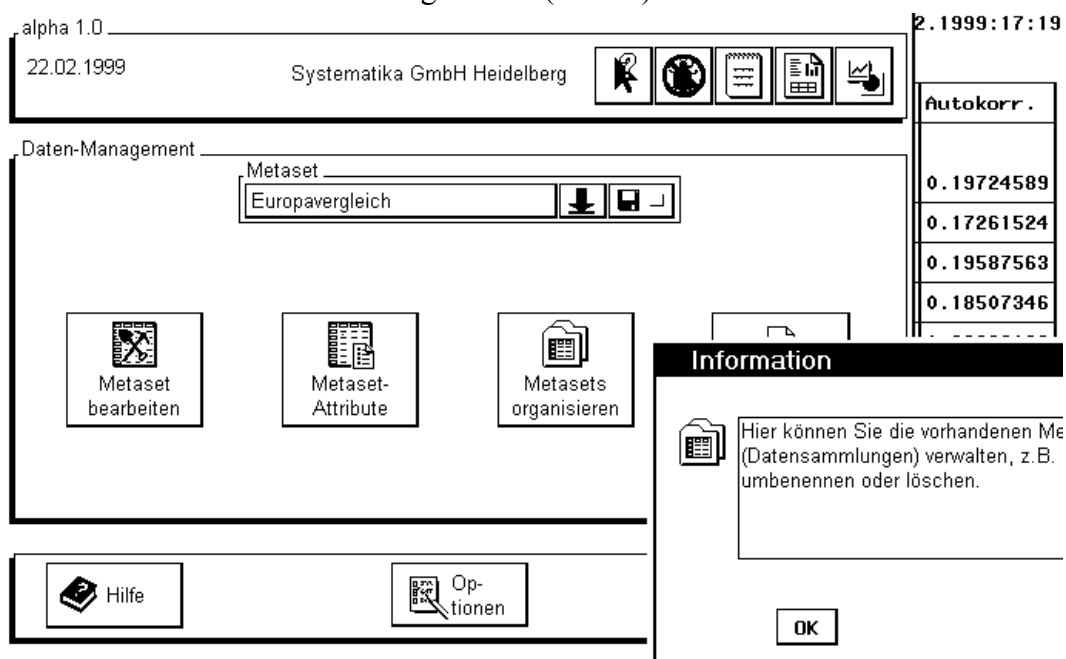


Abb. 4: AF-Oberfläche der SAS-Applikation (Beispiel: Menü der Datenverwaltung)

Leistungsumfang

- Import von Zeitreihen im dendrochronologischen Standardformat
- Import als Einbettung oder Verknüpfung, so daß die Datenpflege wahlweise im SAS oder im ursprünglichen System erfolgt
- Einfache Einbeziehung von SAS-Datasets
- Zusammenstellung von externen und SAS-Dateien zu virtuellen Datasets oder „Metasets“
- Durchgängige Berücksichtigung gewichteter Statistiken
- Performance-Optimierung durch Verfügbarhaltung von Zwischenergebnissen
- Ergebnisausgaben als Tabellen, Graphiken, Datasets oder exportierte Zeitreihen
- Einheitliche Menüsteuerung
- Weitgehende Automatisierung
- Optionale Speicherungsmöglichkeit für jede erfolgte Einstellung
- Ausführliche Hilfe zu Frames und Objekten (Bedienungselementen).

Anwendungsbeispiel

Als Beispiel soll eine Korrelationsanalyse von Eichenchronologien aus 7 europäischen Regionen dienen (Abb. 2): Niederlande plus Deutschland, getrennt nach Nord und Süd, Großbritannien plus Irland, ebenfalls nord-südgeteilt, Dänemark, Südnorwegen plus Südschweden sowie Polen. Wenn mit dieser Untersuchung auch nur ein erster Schritt in Richtung einer präziseren Klimarekonstruktion gemacht ist, demonstriert sie doch einige der Möglichkeiten, die mit den oben beschriebenen Methoden hinzugewonnen wurden.

Abb. 5 zeigt die wichtigsten der über alle Regionen aggregierten Statistiken zu den trendbereinigten Zeitreihen. Nach den Angaben zum Datenumfang, den Häufigkeiten und zur zeitlichen Ausdehnung kommen verschiedene Zerlegungen der Gesamtvarianz, gefolgt von weiteren Lage-, Streuungs- und Korrelationsmaßen. Bemerkenswert ist, daß auch in diesem großräumigen Vergleich die mittlere Kreuzkorrelation zwischen den Chronologien sehr hoch ist (0.759), auch wenn die beiden deutschen Chronologien aufgrund ihrer Ähnlichkeit und hohen Probenzahl diese Statistik weit nach oben bewegt haben.

alpha 1.0
22.02.1999 Systematika GmbH Heidelberg

Basis-Statistiken

Metaset Europavergleich	Anzahl Dateien 1	Aggreg. Reihen 19955	Einzelreihen 19877.1
Ältestes Jahr 2 AD	Jüngstes Jahr 1995 AD	Zeitspanne 1994	Anzahl Werte 2361048
Gesamtvar. 0.04567	V. in Klassen 0.041523	Quervarianz 0.037017	Summe Gew. 2135867
Signalvar. 0.0086	Restvarianz 0.030509	Klassen 7	Mittl. Gewicht 0.904627
SNR 800.4388	Quervar. in Kl. 0.03287	Korrel. in Kl. 0.21990	Längsvarianz 0.041519
EPS 0.99875	zwi. Klassen 0.004146	zwi. Klassen 0.75894	Autokorrel. 0.19728
Mittelwert -0.00297	Std.-Abweich. 0.213705	Minimum -2.41	Maximum 1.36

Hilfe Liste Neube-rechn. Zurück

Abb. 5: Statistische Maßzahlen für die Daten des Anwendungsbeispiels

Von allen berechneten Korrelationen seien hier nur die Scores der 1. Hauptkomponente für den Zeitraum 1000 AD bis heute wiedergegeben (Abb. 6). Wie oben erläutert, stehen weit auseinanderliegende Kurven für niedrige bzw. negative Korrelationen, während hohe Korrelationen auch zu eng beieinanderliegenden Kurven führen. Da eine Hauptkomponente nur einen Teil der Information aus den vollständigen Korrelationsmatrizen wiedergeben kann, ist hier allerdings keine genaue Rückrechnung in Korrelationen möglich. Einige Aussagen lassen sich aber direkt aus der Abb. herleiten:

- Die Korrelationen unterscheiden sich zwischen den Regionen.
- Es kristallisieren sich ca. 4 Cluster heraus: Großbritannien/Irland (gestrichelt), Niederlande/Deutschland (durchgezogen), Skandinavien (kurz-lang) und Polen (gepunktet). Dänemark (kurz-kurz-lang) vermittelt zwischen den letzten drei Clustern.
- Die Korrelationen verändern sich mit der Zeit.
- Die größten Unterschiede gibt es zwischen Großbritannien („UK“) und Kontinentaleuropa. Es kommt in bestimmten Zeitabschnitten zu Annäherungen, vor allem um 1400 und 1650 herum. Nach 1400 ist die Clusterstruktur vor allem durch den Abstand UK-Festland geprägt.
- Auch zwischen den übrigen Regionen gibt es Phasen erhöhter Korrelation, z.B. Dänemark-Skandinavien um 1200, UK-Süd und Dänemark um 1700.
- Ab 1400 korrelieren die skandinavischen und die südlichen Eichen fast durchgängig stärker.

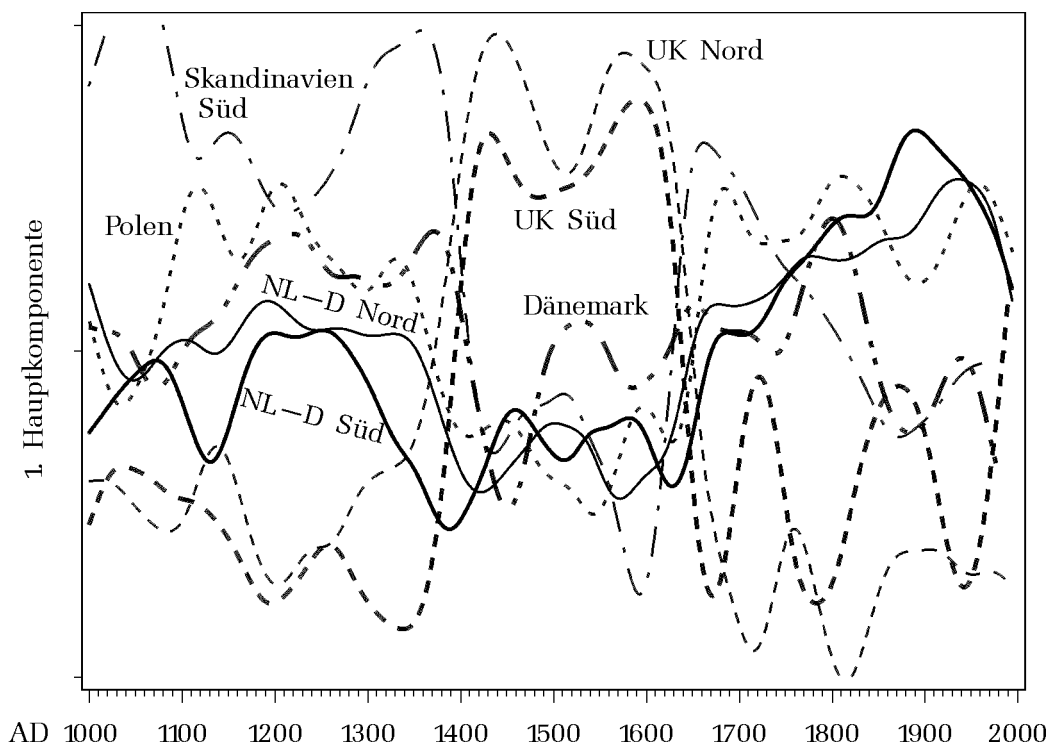


Abb. 6: Scores der ersten Hauptkomponente aus den Korrelationen zwischen 7 Jahrringchronologien

Die besonders auffallenden Änderungen in dem Bereich 1350-1400 fallen zeitlich ungefähr mit dem vermuteten Ende des mittelalterlichen Klimaoptimums zusammen. Die im folgenden höheren Korrelationen z.B. zwischen den temperatursensitiven, südschandinavischen und den

unterschiedlich geprägten, deutsch-niederländischen Eichen stützen deshalb die Hypothese von niedrigeren Temperaturen nach dieser Zeit.

Literatur

- COOK, E.R. und L.A. KAIRIUKSTIS [Hrsg.], 1989: Methods of Dendrochronology. Applications in the Environm. Sciences. Kluwer, Dordrecht, 394 S.
- FRITTS, H.C., 1976: Tree Rings and Climate. Academic Press, London, 567 S.
- HOUGHTON J., 1994: Global Warming: The Complete Briefing. Lion Publishing, Oxford, 192 S.
- JONES, P.D., R.S. BRADLEY und J. JOUZEL [Hrsg.], 1996: Climatic variations and forcing mechanisms of the last 2000 years. Springer, Berlin, 649 S.
- LINGG, W., 1986: Dendroökologische Studien an Nadelbäumen im alpinen Trockental Wallis. Berichte der Eidg. Anstalt für das forstl. Versuchswesen, Birmensdorf, Nr. 287, 81 S.
- RIEMER, T., 1994: Über die Varianz von Jahrringbreiten. (Dissertation) Berichte des Forschungszentrums Waldökosysteme A121, Göttingen, 375 S.
- SCHWEINGRUBER, F.H., 1988a: Climatic Information for the past hundred Years in Width and Density of Conifer Growth Rings. Lecture Notes in Earth Sciences 16, Springer, Berlin, S. 35-55
- SCHWEINGRUBER, F.H., 1988b: Tree Rings. Basics and Applications of Dendrochronology. Kluwer, Dordrecht, 276 S.