INEQUALITY IN PALEORECORDS

FRANCO BIONDI1,3 AND FARES QEADAN2

1DendroLab, Department of Geography, University of Nevada, Reno, Nevada, USA and School of Earth Sciences, Stanford University, Palo Alto, California, USA
2Department of Mathematics and Statistics, University of Nevada, Reno, Nevada, USA

Abstract. Paleorecords provide information on past environmental variability, and help define ecological reference conditions by means of changes in their characteristics (accumulation rate, geochemical composition, density, etc.). A measure of temporal dissimilarity, which has traditionally been used in dendrochronology and is called “mean sensitivity,” only focuses on first-order time-series lags. In this paper mean sensitivity was extended to all possible lags to derive a mean sensitivity function (MSF). The MSF is equivalent to a one-dimensional form of the paired relative madogram, a tool used in geostatistics to quantify spatial dependence. We then showed that the sum of madograms for all possible time-series lags is encapsulated by a single parameter, the Gini coefficient. This parameter has long been used by econometricians, social scientists, and ecologists as a synthetic, quantitative measure of inequality and diversity. Considering the connection between the MSF and the madogram, and the convenience of summarizing data heterogeneity with a single number, the Gini coefficient is therefore particularly appropriate for succinctly evaluating the diversity of paleorecords. An example of this application is provided by focusing on public domain dendrochronological data for the western conterminous United States.

Key words: data analysis; dendrochronology; environmental variability; Gini coefficient; madogram; mean sensitivity; paleoclimatology; paleoecology; reference conditions; time series; western United States.

INTRODUCTION

Paleorecords provide information on past environmental variability by means of changes in their characteristics (accumulation rate, geochemical composition, density, etc.). For instance, variability over time and/or space in specific features of natural sequences (such as sediments, ice and coral layers, tree rings) is commonly exploited to reconstruct past ecological and climatic fluctuations (Bradley 1999, Egan and Howell 2001). This information is essential for the determination of ecological reference conditions, which in turn guide conservation and restoration of natural landscapes and ecosystems by providing a baseline for measuring the effects of past and present activities (Swetnam et al. 1999, Jackson et al. 2001, Andersen et al. 2004).

Common measures of data heterogeneity, including the coefficient of variation (Bedeian and Mossholder 2000), do not fully describe certain properties, such as time-series continuity, which are typical of paleorecords (Drummond and Wilkinson 2006). Variability in time and space is important not only for reconstruction purposes, but also for comparing specimens to one another in order to identify similarities and differences. This comparison can be used, for instance, to match sequential patterns over multiple timescales for improving accuracy of dates assigned to paleorecords. Such “cross dating” is a mainstay of dendrochronology because of the high replication that is typical of tree-ring samples (Douglass 1941, Stokes and Smiley 1996). Besides visual comparisons, cross-dating accuracy can be tested by means of numerical techniques (Baillie and Pilcher 1973, Holmes 1983, Wigley et al. 1987, Van Deusen 1990, Yamaguchi 1991). Cross dating, which has allowed developing tree-ring chronologies one or two orders of magnitude longer than a tree life span (LaMarche and Harlan 1973, Pilcher et al. 1984, Eronen et al. 2002, Friedrich et al. 2004), occupies such a central role in dendrochronology that its systematic use has been used to define the science itself (Fritts and Swetnam 1989). Recently cross dating has also been applied to other kinds of paleorecords (e.g., Hendy et al. 2003), as the need for replication becomes increasingly recognized (Lough 2004).

The problem of quantifying the degree of variation in a data set using a single number emerged early in the development of modern dendrochronology. In the very first issue of the journal Ecology, Douglass (1920) proposed a parameter, called “mean sensitivity,” to synthetically measure the year-to-year (or high-frequency) variability of a tree-ring series. Mean sensitivity, as proposed by Douglass, was supposed to be an indicator, independent of ring size, of the climatic signal present in a tree-ring series (Douglass 1928:104). After nine decades since it was first proposed, mean sensitivity continues to attract interest in the scientific community.
(see for example Falcon-Lang 2005; Cook and Pederson, in press), although its statistical properties were found to add relatively little to the information given by the standard deviation and first-order autocorrelation (Strackee and Jansma 1992). To date, no study has focused on the fact that mean sensitivity only refers to adjacent values in a time series, hence the extension of mean sensitivity from one-year lags to all possible time-series lags has not yet been proposed. As an analogy, time-series autocorrelation is often reported using only one-year (or “first-order”) lags, but when all possible lags are considered, the resulting combination of coefficients has become known as the ‘autocorrelation function’ (Box and Jenkins 1976). In this paper we argue that mean sensitivity also needs to be extended to all possible lags, thereby generating a mean sensitivity function (MSF). We then show that the MSF is equivalent to a special type of madogram, a tool used in geostatistics to quantify dependence between spatially distributed data (Pannatier 1996, Buddenbaum et al. 2005). Finally, we demonstrate that madograms calculated for all possible lags of a time series can be combined into a single parameter, the Gini coefficient (Gini 1912). Long used by econometricians to measure inequality (Sen 1973, Milanovic 1997), the Gini coefficient (G) is a common statistical index of diversity in social (Anand 1982) and demographic (Shkolnikov et al. 2003) studies, and it is also a “splitting rule” for automated classification algorithms (Breiman et al. 1984). Its role in ecology as a synthetic, quantitative measure of heterogeneity was first recognized with regard to plant-size distributions (Weiner and Solbrig 1984, Weiner 1985), and G has since become a common tool in plant-population studies (Knox et al. 1989, Damgaard and Weiner 2000). Considering the connection between the MSF and the madogram, and the convenience to summarize all-lag variability with a single number, the Gini coefficient is therefore particularly appropriate for succinctly evaluating the diversity of paleoenvironmental records. We provide an example of this application by focusing on public-domain tree-ring records for the western United States.

**MATERIALS AND METHODS**

**Mean sensitivity**

In Douglass’ words, mean sensitivity “may be described as the difference between each two successive rings divided by their mean” (Douglass 1920:29). We could then write its formula as

$$MS = \frac{2}{n-1} \sum_{i=2}^{n} \frac{|w_i - w_{i-1}|}{w_i + w_{i-1}}$$

with \(w\) = ring width, \(n\) = length of the tree-ring series, \(t\) = 1, 2, ..., \(n\) = year in the tree-ring series. In Douglass’ days computing power was a rare commodity, hence for practical purposes mean sensitivity was calculated by taking "the sum of all the changes in 10 years without regard to sign and divide by the sum of the 10 years’ growth" (Douglass 1928:30). This way of computing mean sensitivity was formalized by Schulman (1956:20), as the “average absolute first difference expressed as a ratio of the mean ring-width,” so that Schulman’s formula can be rewritten as

$$MS = \frac{n}{n-1} \sum_{i=1}^{n} w_i - w_{i-1}$$

Schulman correctly pointed out that mean sensitivity is influenced by the presence of a trend, and that as long as none is present, Eqs. 1 and 2 give very similar results. Douglass had already indicated that mean sensitivity could be affected by non-climatic factors, such as injuries (Douglass 1928:104). For these reasons, mean sensitivity is nowadays computed for tree-ring chronologies, which are typically obtained after removing biological trends and other peculiarities (such as deformations caused by wounds or scars), and represent a combination of multiple, standardized ring-width index series of the same species and site (Cook and Kairiukstis 1990, NRC 2006). In his classic book, Fritts (1976:258), reinstated Douglass’ formula (Eq. 1) for computing mean sensitivity, and that formula continues to be used by current dendrochronological software, such as COFECHA (Holmes 1983).

Given that ring width cannot be negative, mean sensitivity (MS) reaches a theoretical minimum of 0 when all rings have exactly the same width, and a theoretical maximum of 2 when a ring is locally absent every other year. Because these theoretical limits are not found in nature, values of MS for tree-ring chronologies normally vary from 0.1 to 0.6 (Fritts and Shatz 1975). The presence of two (or more) adjacent missing rings makes the ratio in Eq. 1 undefined, so in this case the ratio is assumed to be missing. While this can happen in individual specimens, it is extremely unlikely that two adjacent rings are locally missing in all samples used to develop a tree-ring chronology. As a relatively extreme case, Biondi et al. (2003) reported that two consecutive locally absent rings affected 8 out of 63 *Pinus hartwegii* samples following the 1913 eruption of Volcán de Colima, Mexico.

**Gini coefficient**

The Gini coefficient (G) for discrete data is given by the sum of absolute differences between all pairs of observations, scaled by mean and sample size (Weiner and Solbrig 1984). In mathematical notation,

$$G = \frac{1}{2n^2 \hat{\sigma}} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|$$

$$= \frac{1}{2n} \sum_{i=1}^{n} x_i \sum_{j=1}^{n} |x_i - x_j|.$$
This definition is found in Sen (1973), but it is slightly different from Gini’s own formula (Gini 1912:36), which was
\[ G = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|. \] (4)

It is interesting to note that multiplying Eq. 3 by \( n/(n-1) \) was suggested as a way to calculate unbiased sample estimates of \( G \) (Weiner and Solbrig 1984, Weiner 1985) when, in reality, Gini’s intention was to obtain a measure of “mean difference between quantities” that did not involve computing the mean of such quantities. In economics, the quantity defined by Eq. 3 is also known as half of the relative mean difference (Xu 2004).

Application example

To demonstrate how data diversity can be summarized by the Gini coefficient, we used a subset of all tree-ring chronologies available from the International Tree-Ring Data Bank (ITRDB; Grissino-Mayer and Fritts 1997; available online).4 We limited the analysis to chronologies longer than 100 years that began in or before 1880, ended in or after 1960, and were located between 30–50°N latitude and 100–125° W longitude. Additional tree-ring chronologies from single-leaf pine (Pinus monophylla) collections were provided by ongoing research projects conducted by the DendroLab (F. Biondi and S. Strachan, 2005 and 2007 unpublished manuscripts). By essentially selecting the western portion of the conterminous United States we limited the analysis to a region where a large number of well-replicated chronologies are readily available.

The entire data set was used to compute the Gini coefficient by species using the common period 1880–1960. A subset of chronologies longer than 400 years (that is, reaching back to at least 1560) was used to compute Gini coefficients by 100-year periods (1561–1660, 1661–1760, 1761–1860, 1861–1960). Differences between species and centuries were tested using the bootstrap procedure described by Dixon et al. (1987). Confidence intervals for the population Gini coefficient were first calculated using the percentile, bias-corrected, and accelerated bootstrapping procedures (see Efron and Tibshirani 1993), where each bootstrap resample is a random selection, with replacement, of the observations. However, we found no advantage of using one method over another, since numerical results agreed up to the third decimal digit (sometime the fourth one) when testing differences between species. Therefore, results were only reported for the percentile method. Data analysis was performed using a combination of software packages, including R (R Development Core Team 2006), SAS (SAS Institute 2003), and GMT (Wessel and Smith 2002).

Results

Mean sensitivity function as a special madogram

By extending the computation of mean sensitivity (Eq. 1) to all possible lags of a time series \((i, j)\), one obtains a mean sensitivity function (MSF), defined as
\[ MS_l = \frac{2}{n - l} \sum_{j=l+1}^{n} \frac{|x_i - x_{i-l}|}{i + i-l} \] (5)
with MS\(_l\) = mean sensitivity for lag \(l\). This expression also represents a way to describe time-series continuity based on the mean absolute deviation between pairs of sample values scaled by their mean. Eq. 5 is equivalent to a one-dimensional, pairwise relative madogram, which is one of the experimental measures proposed in geostatistics to quantify spatial variability (Deutsch and Journel 1998). A demonstration of this relationship is as follows. The (semi)madogram is commonly defined as half of the average absolute difference between pairs of points separated by a given vector. In mathematical notation, it is the following (Deutsch and Journel 1998: Eq. III.8):
\[ \gamma_M(h) = \frac{1}{2N(h)} \sum_{j=1}^{N(h)} |x_j - y_j| \] (6)
with \( \gamma_M(h) \) = (semi)madogram value for separation vector \( h \); \( N(h) \) = total number of pairs separated by vector \( h \); \( x_j \) and \( y_j \) = one of the pairs of values separated by vector \( h \). In a pairwise relative (semi)variogram \( \gamma_{PR}(h) \), each pair is normalized by its squared average, as follows (Deutsch and Journel 1998: Eq. III.6):
\[ \gamma_{PR}(h) = \frac{1}{2N(h)} \sum_{j=1}^{N(h)} \frac{(x_j - y_j)^2}{(x_j + y_j)/2}. \] (7)
By analogy, in a pairwise relative (semi)madogram \( \gamma_{PRM}(h) \), each pair would be normalized by its absolute average, as follows:
\[ \gamma_{PRM}(h) = \frac{1}{2N(h)} \sum_{j=1}^{N(h)} \frac{|x_j - y_j|}{(x_j + y_j)/2}. \] (8)
Assuming that all time-series values are nonnegative, Eq. 8 becomes
\[ \gamma_{PRM}(h) = \frac{1}{N(h)} \sum_{j=1}^{N(h)} \frac{|x_j - y_j|}{x_j + y_j}. \] (9)
In a time series, the separation vector \( h \) is equivalent to the lag \( l \) between sequential values, and the number of pairs for a given separation vector is then equal to the length of the time series minus the lag, that is
\[ N(h) = n - l. \] (10)
For a time series, one can then write

4 [http://www.ncdc.noaa.gov/treering.html]
\[ n/C_0 \times \frac{X_n}{C_0} i = 1, j = x_i/C_0 y_i/C_21_0 y_i/C_21_0: \]

It is easy to recognize that the \((x_i, y_i)\) pair of Eq. 6, when considered for a time series, as in Eq. 11, is equivalent to the \((i_t, i_t/C_0 l)\) pair of Eq. 5, which defines the mean sensitivity function (MSF). We can then formalize the connection between the MSF and the madogram as follows:

\[ \gamma_{\text{PRM}(l)} = \frac{1}{n} \sum_{i=1}^{n-1} \frac{|x_i - y_i|}{x_i + y_i} \text{ if } x_i \geq 0 \text{ and } y_i \geq 0. \]  

(11)

If mean sensitivity had been computed using squared rather than absolute differences, the mean sensitivity function would have been equivalent to a one-dimensional, pairwise relative variogram (Srivastava and Parker 1989).

**Gini coefficient as a sum of madograms for all possible lags**

If we consider nonnegative time-series data, the Gini coefficient combines into a single number the madograms for all possible lags, as follows:

\[ \text{MS}_l = 2 \gamma_{\text{PRM}(l)}. \]  

(12)

<table>
<thead>
<tr>
<th>Species</th>
<th>Code</th>
<th>N†</th>
<th>Min.–Max. (°W)</th>
<th>Mean ± SD</th>
<th>Min.–Max. (°N)</th>
<th>Mean ± SD</th>
<th>First year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinus ponderosa</td>
<td>PIPO</td>
<td>199</td>
<td>100.0–122.7</td>
<td>111.6 ± 6.5</td>
<td>30.1–49.6</td>
<td>39.2 ± 4.8</td>
<td>1260–1880</td>
</tr>
<tr>
<td>Pseudotsuga menziesii</td>
<td>PSME</td>
<td>157</td>
<td>104.4–124.0</td>
<td>112.5 ± 0.1</td>
<td>30.4–49.6</td>
<td>40.1 ± 5.9</td>
<td>–136–1867</td>
</tr>
<tr>
<td>Pinus edulis</td>
<td>PIED</td>
<td>62</td>
<td>103.3–114.2</td>
<td>108.4 ± 2.5</td>
<td>32.2–40.8</td>
<td>36.1 ± 2.0</td>
<td>1270–1837</td>
</tr>
<tr>
<td>Picea engelmannii</td>
<td>PCEN</td>
<td>40</td>
<td>105.0–120.7</td>
<td>111.6 ± 5.3</td>
<td>31.9–49.9</td>
<td>42.8 ± 4.8</td>
<td>1224–1828</td>
</tr>
<tr>
<td>Pinus monophylla</td>
<td>PIMO</td>
<td>31</td>
<td>113.3–119.7</td>
<td>115.9 ± 1.8</td>
<td>37.6–41.1</td>
<td>39.2 ± 0.8</td>
<td>958–1831</td>
</tr>
<tr>
<td>Pinus flexilis</td>
<td>PIFL</td>
<td>30</td>
<td>105.0–119.4</td>
<td>111.4 ± 4.3</td>
<td>33.8–49.6</td>
<td>41.6 ± 4.0</td>
<td>–77–1508</td>
</tr>
<tr>
<td>Pinus jeffreyi</td>
<td>PJJE</td>
<td>27</td>
<td>115.5–123.0</td>
<td>119.1 ± 2.2</td>
<td>31.0–42.0</td>
<td>36.9 ± 3.1</td>
<td>1304–1880</td>
</tr>
<tr>
<td>Juniperus occidentalis</td>
<td>JUOC</td>
<td>24</td>
<td>116.8–121.8</td>
<td>119.8 ± 1.2</td>
<td>36.9–44.9</td>
<td>41.5 ± 2.3</td>
<td>–420–1654</td>
</tr>
<tr>
<td>Other (33 total)</td>
<td></td>
<td>196</td>
<td>100.4–123.6</td>
<td>116.3 ± 5.9</td>
<td>31.0–49.9</td>
<td>39.6 ± 5.3</td>
<td>–6000–1880</td>
</tr>
</tbody>
</table>

Note: First year = AD year when each chronology began; last year = AD year when each chronology ended.
† N = number of chronologies. Species are listed by decreasing N; when N < 20, species were grouped under “Other” (in this category, N per species varied from 1 to 18).
‡ Length = no. years in each chronology.
§ Gini coefficient computed according to Eq. 3 for the years 1880–1960.

Using Eq. 16, the numerator of Eq. 15 becomes

\[ 2\left(\sum_{i=1}^{n-1} |x_i - x_{i-1}| + \sum_{i=1}^{n} |x_i - x_n|\right) \]

which can then be rearranged into the following formula:

\[ 2\left(\sum_{i=1}^{n} |x_i - x_{i-1}| + \sum_{i=1}^{n} |x_i - x_n|\right) \]

From Eqs. 15 and 18 it follows that

\[ G = \frac{1}{2n} \sum_{i=1}^{n} |x_i - x_{i-1}| + \sum_{i=1}^{n} |x_i - x_n| \]


<table>
<thead>
<tr>
<th>Species</th>
<th>Code</th>
<th>N†</th>
<th>Min.–Max. (°W)</th>
<th>Mean ± SD</th>
<th>Min.–Max. (°N)</th>
<th>Mean ± SD</th>
<th>First year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudotsuga menziesii</td>
<td>PSME</td>
<td>73</td>
<td>104.4–124.0</td>
<td>112.4 ± 6.3</td>
<td>31.9–48.9</td>
<td>40.9 ± 5.4</td>
<td>–136–1557</td>
</tr>
<tr>
<td>Pinus ponderosa</td>
<td>PIPO</td>
<td>51</td>
<td>103.3–122.7</td>
<td>114.4 ± 7.0</td>
<td>31.2–49.6</td>
<td>40.6 ± 4.3</td>
<td>1260–1557</td>
</tr>
<tr>
<td>Pinus flexilis</td>
<td>PIFL</td>
<td>30</td>
<td>105.0–119.4</td>
<td>111.4 ± 4.3</td>
<td>33.8–49.6</td>
<td>41.6 ± 4.0</td>
<td>–77–1508</td>
</tr>
<tr>
<td>Pinus monophylla</td>
<td>PIMO</td>
<td>27</td>
<td>113.3–119.7</td>
<td>115.9 ± 1.8</td>
<td>37.9–41.1</td>
<td>39.9 ± 0.7</td>
<td>958–1558</td>
</tr>
<tr>
<td>Juniperus occidentalis</td>
<td>JUOC</td>
<td>23</td>
<td>116.8–121.8</td>
<td>119.7 ± 1.2</td>
<td>37.4–44.9</td>
<td>41.5 ± 2.4</td>
<td>–420–1548</td>
</tr>
<tr>
<td>Pinus edulis</td>
<td>PIED</td>
<td>23</td>
<td>105.2–112.1</td>
<td>109.1 ± 2.1</td>
<td>34.2–40.8</td>
<td>37.3 ± 2.0</td>
<td>1270–1560</td>
</tr>
<tr>
<td>Other (20 total)</td>
<td></td>
<td>114</td>
<td>105.0–123.6</td>
<td>115.7 ± 5.6</td>
<td>31.0–49.6</td>
<td>39.6 ± 4.5</td>
<td>–6000–1560</td>
</tr>
</tbody>
</table>

Note: First year = AD year when each chronology began; last year = AD year when each chronology ended.
† N = number of chronologies. Species are listed by decreasing N; when N < 20, species were grouped under “Other” (in this category, N per species varied from 1 to 18).
‡ Length = no. years in each chronology.
§ Gini coefficient computed according to Eq. 3 for the years 1561–1960.
TABLE 1. Extended.

<table>
<thead>
<tr>
<th>Last year</th>
<th>Length(\ddagger)</th>
<th>Gini (1880–1960)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.–Max.</td>
<td>Mean ± SD</td>
<td>Min.–Max.</td>
</tr>
<tr>
<td>1963–2004</td>
<td>1982 ± 11</td>
<td>111–736</td>
</tr>
<tr>
<td>1960–2001</td>
<td>1977 ± 11</td>
<td>106–2129</td>
</tr>
<tr>
<td>1964–2002</td>
<td>1977 ± 10</td>
<td>162–709</td>
</tr>
<tr>
<td>1971–1999</td>
<td>1985 ± 5</td>
<td>156–769</td>
</tr>
<tr>
<td>1964–2004</td>
<td>1991 ± 13</td>
<td>146–1043</td>
</tr>
<tr>
<td>1964–2000</td>
<td>1984 ± 12</td>
<td>457–2076</td>
</tr>
<tr>
<td>1964–1996</td>
<td>1986 ± 8</td>
<td>111–663</td>
</tr>
<tr>
<td>1964–2001</td>
<td>1988 ± 9</td>
<td>327–2420</td>
</tr>
<tr>
<td>1960–2003</td>
<td>1986 ± 9</td>
<td>103–7980</td>
</tr>
</tbody>
</table>

\[ G = (|x_1 - x_2| + |x_2 - x_3| + \cdots + |x_n - x_1| + |x_1 - x_3| + \cdots + |x_n - x_n|) / \left( \frac{n}{n} \sum_{i=1}^{n} x_i \right). \] 

Similarly, for \( l = 2 \), it follows that
\[ |x_1 - x_3| + |x_2 - x_4| + \cdots + |x_{n-2} - x_n| = 2(n-2)\gamma_{M(2)}. \]

Using the same procedure, for \( l = n - 2 \) it is possible to write
\[ |x_1 - x_{n-1}| + |x_2 - x_n| = 4\gamma_{M(n-2)} \]

and for \( l = n - 1 \) one finds
\[ |x_1 - x_n| = 2\gamma_{M(n-1)}. \]

Considering Eqs. 22–25, the Gini coefficient given in Eq. 19 can then be expressed by
\[ G = [2(n-1)\gamma_{M(1)} + 2(n-2)\gamma_{M(2)} + \cdots + 4\gamma_{M(n-2)}] + \frac{2\gamma_{M(n-1)} / \left( \frac{n}{n} \sum_{i=1}^{n} x_i \right)}{2 \sum_{i=1}^{n-1} [(n-1)\gamma_{M(i)}] / \left( \frac{n}{n} \sum_{i=1}^{n} x_i \right)}. \]

which proves Eq. 13.

The relation found in Eq. 13 can also be verified using matrix notation. If we consider an \( n \)-element vector, \( \mathbf{x} \), then

\[ |x_1 - x_2| + |x_2 - x_3| + \cdots + |x_{n-1} - x_n| = 2(n-1)\mathbf{y}_{M(1)}. \]

TABLE 2. Extended.

<table>
<thead>
<tr>
<th>Last year</th>
<th>Length(\ddagger)</th>
<th>Gini (1561–1960)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.–Max.</td>
<td>Mean ± SD</td>
<td>Min.–Max.</td>
</tr>
<tr>
<td>1962–2001</td>
<td>1979 ± 12</td>
<td>424–2129</td>
</tr>
<tr>
<td>1963–1998</td>
<td>1982 ± 11</td>
<td>412–736</td>
</tr>
<tr>
<td>1964–2000</td>
<td>1984 ± 12</td>
<td>457–2076</td>
</tr>
<tr>
<td>1976–2004</td>
<td>1994 ± 11</td>
<td>426–1043</td>
</tr>
<tr>
<td>1964–2001</td>
<td>1988 ± 10</td>
<td>435–2420</td>
</tr>
<tr>
<td>1964–2002</td>
<td>1976 ± 10</td>
<td>405–709</td>
</tr>
<tr>
<td>1960–2000</td>
<td>1984 ± 9</td>
<td>429–7980</td>
</tr>
</tbody>
</table>
The double partial sums found in Eq. 3 can be stored in matrix $A$ as follows:

$$A = \begin{bmatrix}
|x_1 - x_1| & |x_2 - x_1| & |x_3 - x_1| & |x_4 - x_1| & \cdots & |x_n - x_1| \\
|x_1 - x_2| & |x_2 - x_2| & |x_3 - x_2| & |x_4 - x_2| & \cdots & |x_n - x_2| \\
|x_1 - x_3| & |x_2 - x_3| & |x_3 - x_3| & |x_4 - x_3| & \cdots & |x_n - x_3| \\
|x_1 - x_4| & |x_2 - x_4| & |x_3 - x_4| & |x_4 - x_4| & \cdots & |x_n - x_4| \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
|x_1 - x_n| & |x_2 - x_n| & |x_3 - x_n| & |x_4 - x_n| & \cdots & |x_n - x_n|
\end{bmatrix}$$

One can see that $A_{xy}$ is a zero-diagonal, symmetric square matrix, where the sub-matrix sum $\xi(l)$ of its $l$th upper (or lower) diagonal equals the sum of Eq. 20, i.e.,

$$\xi(l) = |x_1 - x_{1+l}| + |x_2 - x_{2+l}| + \cdots + |x_{n-l} - x_n|$$

$$= \sum_{j=1}^{n-l} |x_i - x_{i+j}|.$$  \hspace{1cm} (27)

From Eqs. 27 and 28 it also follows that the calculation of the Gini parameter according to Eq. 3 loses one more degree of freedom than the one performed using Eq. 13.

**Application example**

A total of 766 tree-ring chronologies (Fig. 1A) satisfied our selection requirements. Only 31 chronologies included year 2000, as most chronologies in the International Tree-Ring Data Bank (ITRDB, see footnote 4; Grissino-Mayer and Fritts 1997) were archived in previous years. A total of 41 tree species were included in the data set, with a prevalence of ponderosa pine ($Pinus ponderosa$ [PIPO], 199 chronologies) and Douglas-fir ($Pseudotsuga menziesii$ [PSME], 157 chronologies). Together with southwestern pinyon pine ($Pinus edulis$ [PIED], 62 chronologies), these three species comprised about 55% of the selected time series.

In terms of spatial distribution, PIPO and PSME records were the most widespread; among tree species that accounted for more than 20 chronologies, single-leaf pinyon ($Pinus monophylla$ [PIMO]) had the smallest range of geographical coordinates (Table 1).

The subset of 341 chronologies with a 1560–1960 common period (Fig. 1B and Table 2) included a total of 26 species, with a prevalence of Douglas-fir (PSME, 73 chronologies) and ponderosa pine (PIPO, 51 chronologies), as in the entire 766-series set. Together with limber pine ($Pinus flexilis$ [PIFL], 30 chronologies), these three species comprised about 45% of the selected time series (Table 2). In terms of spatial distribution, PSME and PIPO were still the most widespread species, although their numbers were about one third of what they were in the entire set. From two sample $t$ tests on mean longitude and latitude, this reduction in number of records affected the distribution of PSME time series.
(P > 0.29) less than that of PIPO (0.01 < P < 0.04). On the opposite end, PIFL chronologies were exactly the same in the 766-series set as in the 341-series subset, and there was a difference of only one chronology for *Juniperus occidentalis* (JUOC). Coordinates of pinyon chronologies showed no significant difference (P > 0.92) for PIMO, which went from 31 to 27 chronologies between the entire set and the subset, but some difference (P > 0.03) emerged for PIED, whose chronologies decreased in number by about two thirds (Tables 1 and 2).

For the 1880–1960 period, *Pinus edulis* chronologies were characterized by the greatest diversity among all other species, as shown by the Gini coefficient (Table 1). The map of interpolated Gini values (Fig. 1C) presented a general increase from north to south, with a pronounced peak in the American Southwest, which was most likely related to the greater presence of *Pinus*.
edulis time series in that region. Differences between species appeared to decrease when Gini coefficients were computed for a much longer period, from 1561 to 1960 (Table 2). Since Gini values in Tables 1 and 2 were computed separately for each chronology, we more rigorously tested differences between species by first pooling all 1880–1960 data by species, and then computing bootstrapped 95% confidence intervals as suggested by Dixon et al. (1987).

Tree-ring diversity varied between species (Fig. 2), with *Picea engelmannii* (PCEN; Engelmann spruce) having the lowest value, and *Pinus edulis* (PIED; southwestern pinyon) showing the highest one. Between these two extremes, two groups of species were significantly different from each other. The higher group consisted of ponderosa pine (*Pinus ponderosa*; PIPO) and Douglas fir (*Pseudotsuga menziesii*, PSME). The lower group included single-leaf pinyon (*Pinus monophylla*, PIMO), Jeffrey pine (*Pinus jeffreyi*, PIJE), limber pine (*Pinus flexilis*, PIFL), and the combination of all other 33 species represented in the data set. Western juniper (*Juniperus occidentalis*, JUOC) could be considered a transition species between the two groups. Also, some differences existed within the two groups, for instance PIMO had significantly higher diversity than PIFL, but overall it was clear that tree-ring chronologies were separable into these broader categories (Fig. 2).

Interpolated maps of Gini coefficients computed at each site for the four 100-year intervals did not reveal large changes in spatial patterns (Fig. 3). We more rigorously tested differences between historical periods by first pooling the 341 chronologies by 100-year intervals (a total of 34 100 values for each time period), and then computing bootstrapped 95% confidence intervals as suggested by Dixon et al. (1987). Each period was different from the others: tree-ring diversity was highest in the most distant past (1561–1660), remained fairly constant in the intermediate periods (from 1661 to 1860), and increased again in the most recent interval (1861–1960; Fig. 4).

**DISCUSSION**

The connection between a measure of temporal variability (mean sensitivity for lag l, MS_l) and geostatistical measures of data continuity is not entirely new. The application of variography (or variogram analysis) to time-series data has been proposed as a promising tool for investigating processes with nonstationary mean and variance (Haslett 1997). That is because the variogram only requires intrinsic stationarity, whereas the autocorrelation function requires second-order stationarity (Cressie 1993). Most paleoecological records are temporally autocorrelated, and the degree of temporal dependence usually decreases for increasing lags (Drummond and Wilkinson 2006). While
a temporal correlogram, or autocorrelation function, measures similarity in time-series data and tends to decrease with increasing lag, a temporal variogram, or semivariance function, measures dissimilarity in time-series data and tends to increase with increasing lag (Isaaks and Srivastava 1989). For a stationary time-series process with null mean and unit variance, the theoretical correlogram is the mirror image of the theoretical variogram (Davis 1986). In addition, linear and exponential variogram models (Isaaks and Srivastava 1989), when used in a time-series context, allow for temporal dependence at infinite lags (see also Ma 2004).

Tree-ring chronologies were used to illustrate the computation and testing of the Gini coefficient as an all-lag measure of diversity in nonnegative time series. We used 766 tree-ring chronologies to test for species differences in tree-ring variability using the common period 1880–1960, and we found that Gini coefficients were significantly higher for single-leaf pinyon (*Pinus edulis*), and significantly lower for Engelmann spruce (*Picea engelmannii*) than for other species in the data set. Statistically significant differences between species were consistent with prior dendroecological research, showing how shade-tolerant, forest-interior species from mesic environments (such as spruce) usually grow at more constant rates than shade-intolerant, open-grown species from xeric environments (such as pinyon), so that the former are characterized by “complacent” rings, and the latter by “sensitive” ones (Fritts 1976). On the other hand, the significant changes we uncovered in tree-ring diversity over time, using a subset of 341 tree-ring chronologies with common period 1561–1960, have not been previously reported. Given the number of tree species, geographical areas, and ecosystems represented in the data set, it is likely that such temporal changes are due to changes in environmental conditions, hence more focused research should be conducted to determine potential mechanisms for the fluctuations in tree-ring variability observed over the western United States.

Bootstrapped confidence intervals were computed using the percentile method. This was adequate because the Gini resampling distribution in our study was approximately Gaussian, with mean equal to the sample Gini coefficient (Fig. 5). Since the bootstrap percentile estimator follows the assumption that the distribution of the sample Gini coefficient (\( G \)) is symmetric about \( G \), the 95\% confidence intervals of the percentile bootstrap method are invariant under any monotonic transformation, and therefore adequate for our purposes (Efron and Tibshirani 1993, Manly 1998, Chernick 1999). Further research should expand the results we presented to allow for negative values in the time series.

![Histogram of 10,000 bootstrap replications of the sample Gini coefficient for the period 1661–1760, where the vertical dashed line indicates \( \hat{G} \), which is the Gini coefficient from the original sample, and \( \hat{G}^* \) is the mean Gini coefficient from the bootstrapped samples. Also given is \( E \), the expected difference between bootstrapped sample Gini and original sample Gini coefficients. The efficiency of the percentile method is due to the tight fit between the histogram and the Gaussian distribution (black line).](image)
By incorporating all possible madograms, hence representing time-series continuity at all possible lags, the Gini coefficient is well suited to represent paleodata diversity. All studies that focus on robust indicators of changing ecological state at multiple temporal scales (such as Parrish and Spicer 1990, Labandeira et al. 2002, Wilf et al. 2006) could benefit from this innovative method. This approach has been described here for equally spaced time-series data, so it is immediately applicable to all such records, including those derived from tree rings, corals, ice cores, and laminated sediments (especially varves) in lacustrine, tidal, and marine environments. Other paleorecords, such as those derived from non-laminated stratigraphic sequences and fossil assemblages, could be suitable for this method after interpolation of the geochronological timescale into equally spaced intervals. As shown in the application to tree-ring data, having a robust, replicable, and comprehensive index of data variability helps with formally testing hypotheses, uncovering significant patterns, and identifying underlying mechanisms. It is interesting that sensitivity, diversity, heterogeneity, variability, inequality, and their opposites (complacency, similarity, homogeneity, continuity, equality), are terms used in the scientific literature, but usually confined within the boundary of specific disciplines. Consider for instance the term “sensitivity,” which is only found in dendrochronology, and the term “inequality,” which is typical of economics. The widespread use of the Gini coefficient in various fields of science, and its novel meaning we have presented in this article, provides an opportunity for bridging and communicating across disciplines.

ACKNOWLEDGMENTS

This work was conceived during the first author’s doctoral studies at the University of Arizona and benefited highly from discussions with Don Myers. Research was partially supported by NSF grant ATM-0503722. F. Biondi was also funded, in discussions with Don Myers. Research was partially supported by a Guest Professorship from the Chair of Forest Ecology, Swiss Federal Institute of Technology, Zurich, Switzerland. We are grateful to three anonymous reviewers for their comments and suggestions.

LITERATURE CITED


