Quaternary Research 78 (2012) 13–23

Contents lists available at SciVerse ScienceDirect

Quaternary Research

journal homepage: www.elsevier.com/locate/yqres

Spatial reconstruction of semi-quantitative precipitation fields over Africa during the nineteenth century from documentary evidence and gauge data

Sharon E. Nicholson *, Douglas Klotter, Amin K. Dezfuli [1]

Earth, Ocean and Atmospheric Science Department, Florida State University, Tallahassee, FL 32306, USA

ARTICLE INFO

Article history:
Received 8 November 2011
Available online 1 May 2012

Keywords:
Africa
Rainfall
Historical climate
Climatic field reconstruction
Nineteenth century
Climate change

ABSTRACT

The article presents a newly created precipitation data set for the African continent and describes the methodology used in its creation. It is based on a combination of proxy data and rain gauge records. The data set is semi-quantitative, with a "wetness" index of −3 to +3 to describe the quality of the rainy season. It covers the period AD 1801 to 1900 and includes data for 90 geographical regions of the continent. The results underscore a multi-decadal period of aridity early in the nineteenth century.

Introduction

Concerns about global change have increased the need for long-term data sets to evaluate past climate. Of particular interest are multi-decadal and multi-century time series to evaluate the factors governing such low-frequency variability. Because of this interest, there has been a push to develop climatic data sets that cover at least a century or longer. For example, the ICOADS global SST data set has recently been extended back to 1800 AD (Woodruff, 2011), and a reanalysis data set with global circulation parameters now goes back to 1871 AD (Compo et al., 2011). These new data sets provide an unprecedented opportunity to study the temporal stability of the forces driving climate.

The precipitation records for the United States, Europe and Asia extend back two or more centuries, long enough to exploit these historical data sets. In contrast, except for Algeria and South Africa, the modern precipitation record for Africa barely covers a century. In this article we present a newly developed African precipitation data set that extends the record back to 1800 AD, using a combination of gauge records and historical information. The data are semi-quantitative, i.e., "wetness" is expressed as a seven-class index. However, the detail suffices for standard statistical analysis methods, such as linear correlation and spectral analysis, to be applied. The data cover 90 regions of the African continent and are available from the NOAA Paleoclimate Data Center in Boulder, Colorado.

* Corresponding author. Fax: +1 850 644 9642.
E-mail address: snicholson@fsu.edu (S.E. Nicholson).
[1] Current affiliation: Dept. of Earth and Planetary Sciences, Johns Hopkins University, 3400 N. Charles Street, 301 Olin Hall, Baltimore, MD 21218, USA.

One purpose of this article is to present the data set to potential users. With the development of high-resolution paleoclimate indicators for recent centuries (e.g., Halfman and Johnson, 1988; Russell et al., 2007; Shanahan et al., 2008), the data set can provide a useful comparison for many paleoclimate researchers. A second purpose is to demonstrate a methodology for integrating and making use of the diverse indicators available on historical time scales. Much more historical climate information exists and can be exploited in this context.

Information on African precipitation in the 19th century

Gauge records

The African precipitation records for the 19th century are of three types: official records for stations continuing to modern day; regular reporting stations established by colonial governments but generally abandoned with the change of the colonial power structure early in the twentieth century; and temporary, non-governmental observations, usually made at mission stations, explorers' camps, etc. The first type is the most readily usable and this type of record comes mainly from Algeria, Tunisia, Senegal, Namibia, and South Africa. Isolated stations with continuous records also exist in many other countries, such as Sierra Leone, Gambia, Ghana, Nigeria, Cameroon, Kenya, and Zimbabwe. The second type of station record is common in countries such as Togo, Southwest Africa (currently Namibia) and Tanganyika (currently Tanzania), which were under German control. The British, French, Portuguese, Italians and Belgians also left behind extensive meteorological records from their former colonies, but their records continued well into the 20th century.

0033-5894/$ – see front matter © 2012 University of Washington. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.yqres.2012.03.012

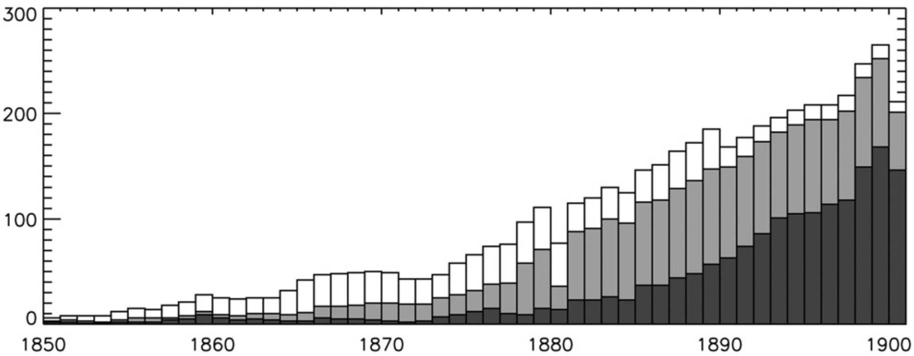

Figure 1. Number of African rainfall stations in the gauge archive during the period 1850 to 1900. Light shading, stations in South Africa; no shading, stations in Algeria; dark shading, stations in remaining countries in Africa.

A useful inventory of African precipitation data is that of Supan (1898), who summarizes all available African gauge data up to the time of his publication and also gives the source of the data for each station. Many of the records made their way into scientific journals. Particularly useful sources are *Petermann's Geographische Mitteilungen* and *Meteorologische Zeitschrift*, as well as the geographical journals of several national geographic societies. A more extensive list of sources of nineteenth century African precipitation records is included in Nicholson (2001).

Figure 1 shows the total number of African stations available from 1850 to 1899 AD. By far the largest number of records is for South Africa and Algeria, so these are shown separately. Records become plentiful for Algeria in the 1860s AD and for South Africa in the 1880s AD. For most other African countries, records commence in the late 1880s or 1890s. Figure 2 shows the location of stations available during one or more years of the nineteenth century. Clearly, the coastal areas are best represented. Some of the notable long-term station records are listed in Table 1.

Documentary data

Documentary data consist of verbal descriptions of conditions related to climate. The sources include settlers' diaries, explorers' journals, reports of scientific expeditions, historical chronicles documenting the reign of centuries of kings, oral tradition, missionary reports, and government records. The diaries, journals and reports are tremendous sources of meteorological information. In some cases the material is merely descriptive, e.g., references to famine or drought, the character of the year's rainy season, or the state of a lake. However, many of the explorers, such as the German Heinrich Barth (1859), kept day-by-day records of the weather, often including even temperature measurements.

The documentary data reports provide transient and temporary records. Sources of more continuous and homogeneous reporting are the European settlers in various African countries and the numerous mission stations, particularly those in the Anglophone countries. The missionary records are particularly important because annual entries are made, the observations are diligently recorded, and many mission stations endured for decades in the same location. These sources generally give descriptive information on such things as famine, drought, the conditions of crops, unusual weather, and the status of vegetation and waterways.

Several authors have produced multi-decadal climate chronologies for Africa from missionary records. These include chronologies for the Kalahari (Nash, 1996; Nash and Endfield, 2002a,b), Lesotho (Grab and Nash, 2010; Nash and Grab, 2010), and South Africa (Vogel, 1989; Kelso and Vogel, 2007). Vogel (1989) validated her records for time periods with overlapping gauge data, supporting the accuracy and credibility of the missionary reports. Support for the validity of the documentary records also comes from correlations with El Niño and the Southern Oscillation (Kelso and Vogel, 2007; Nash and Endfield, 2008). However, the missionary records in Africa have not been fully exploited in terms of climate information. Much more is available, particularly for East Africa.

Other proxy indicators on the historical time scale

Other proxy indicators of historical climate conditions include lake depth and/or extent, river discharge and tree rings. These present a challenge in interpretation because of the complex response function to climate variables, including the integration of conditions over more than one year. The height of a lake surface, for example, is influenced by rainfall, temperature and wind, as well as other variables that modulate evaporation (Nicholson et al., 2000). The composition of a sediment core reflects several factors, including characteristics of the inflow, the net balance of precipitation and evaporation, and even wind (Johnson et al., 2002). The growth rings of trees respond to temperature, precipitation, and other factors controlling moisture availability, including orographic effects (Nicolini et al., 2010; Touchan et al., 2011).

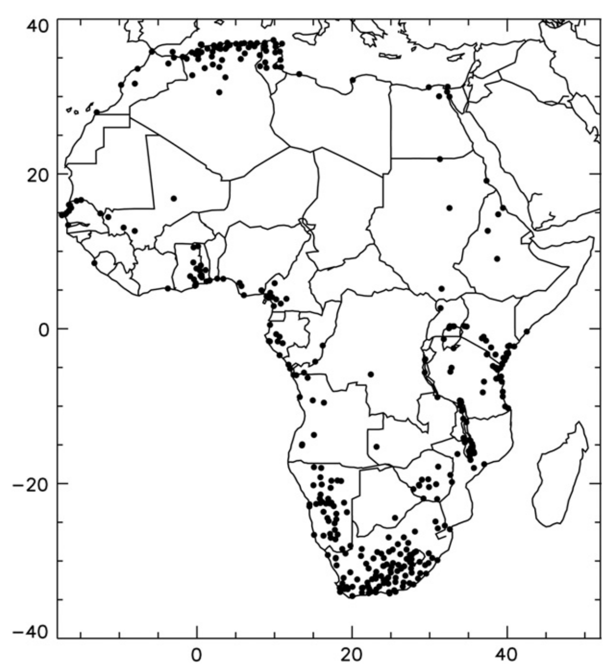

Figure 2. Map of locations with one or more years of data in the 19th century.

Table 1
Notable African stations with long precipitation records and first year of regular coverage.

Station	First year
Algiers, Algeria	1837
Constantine, Algeria	1838
Cape Town, South Africa	1838
Oran, Algeria	1841
Biskra, Algeria	1845
St. Louis, Senegal	1851
Luanda, Angola	1858
Alexandria, Egypt	1859
Port Said, Egypt	1866
Cairo, Egypt	1870
Mogador, Morocco	1873
Zanzibar, Tanzania	1874
Tunis, Tunisia	1873
Freetown, Sierra Leone	1875
Mombasa, Kenya	1875
Tripoli, Libya	1877
Omaruru, Namibia	1882
Aburi, Ghana	1883
Bismarckburg, Togo	1883
Banjul, Gambia	1884
Rehoboth, Namibia	1884
Douala, Cameroon	1885
Walvis Bay, Namibia	1885
Accra, Ghana	1886
Lagos, Nigeria	1886
Massawa, Eritrea	1886
Yaounda, Cameroon	1889

For Africa, tree-ring analysis is limited by the fact that there are few locations where the trees have annual rings. Most of the dendroclimatological studies stem from Morocco, Tunisia, and Algeria (Touchan et al., 2008, 2011). However, some work has been done in Namibia (e.g., Trouet et al., 2001), Tanzania (Fichtler et al., 2004) and Zimbabwe (e.g., Therrell et al., 2006). Unfortunately, the few tree-ring chronologies available to us either did not consistently correlate well with annual rainfall or they covered locations/time periods with adequate gauge data. For that reason tree-ring data were not incorporated into the archive.

For the African lakes a wealth of documentary information exists, enough to produce multi-century chronologies for ten lakes in equatorial and southern Africa (Nicholson, 2001). Unfortunately, those chronologies were of limited use in producing the annual resolution data set described in this article. Our prior work (Nicholson et al., 2000; Nicholson and Yin, 2001; Yin and Nicholson, 2002) demonstrates the difficulties and complexities, which are described in the Methodology section.

Some lacustrine hydrologic records do have annual resolution. In such cases the sediment displays varves, or two distinct layers per year, reflecting seasonal conditions. An excellent case in point is Lake Bosumtwi, in West Africa (Shanahan et al., 2008). Data for this lake have been included in this archive, as have some data points for Lakes Victoria, Tanganyika, Rukwa, Chad, Malawi, Chilwa, Stefanie, Naivasha, Albert, and Turkana. Further detail on the lacustrine records utilized is presented in the Methodology section.

The record of the Nile flow has even higher resolution, with minimum and maximum levels recorded in most years going back to the 7th century (Toussoun, 1925). Lamb (1966) suggested that its summer minimum reflects the equatorial rainfall regime and that its annual flood represents rainfall in the source region of the Blue Nile in Ethiopia. This is a reasonable approximation. However, the distinction is not so clear-cut, with a very abnormal summer minimum influencing the subsequent flood and vice versa. Rainfall in the regions traversed by the Nile also affects its flow. Nevertheless, the Nile record provided useful data for most years of the 19th century.

Methodology

Development of a methodology

The development of the historical data set described herein is an outgrowth of a methodology developed by the first author over several decades of research. A basic methodology of historical climate reconstruction for Africa and the associated difficulties are described in Nicholson (1979, 1981a, 1996). Early work produced a broad, highly generalized sketch of the long-term climate history for key sectors, such as the Sahel, and spatial anomaly patterns of periods of unusual rainfall conditions over large sectors of Africa (e.g., Nicholson, 1978, 1980). Later works detailed long-term fluctuations of lakes mainly in East Africa (e.g., Nicholson, 1998a,b, 1999).

A serious difficulty in exploiting Africa's historical precipitation record is that the information is largely fragmentary, especially prior to the 1890s, as well as qualitative. A methodology was developed to utilize these diverse bits of information to construct semi-quantitative indicators of precipitation for 90 regions of the continent (Nicholson, 2001). The data set described in the current article represents a refinement and extension of the methodology developed therein.

The key to the climate reconstruction is the use of regions that are homogeneous with respect to interannual variability (Nicholson, 1986). These regions are shown in Figure 3. A basic assumption is that information pertaining to any location within the region can be used to produce a precipitation time series representing the region. Secondly, because so much of the available material is descriptive rather than quantitative, a seven-class system is used to describe the "wetness" of the season. These range from −3 to +3, with negative numbers indicating below normal rainfall, positive numbers indicating above normal rainfall and a zero indicating normal conditions.

Gauge data

The gauge data were comparatively easy to utilize. The first step was to ascertain geographical coordinates for the location of the observations. The second was to find a modern analog station, in order to determine an appropriate long-term mean for the 19th

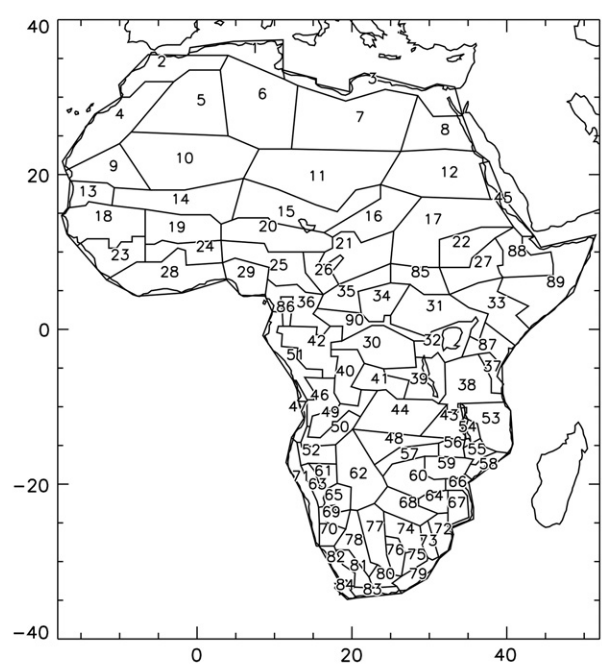

Figure 3. Map of 90 regions utilized in the data set.

https://doi.org/10.1016/j.yqres.2012.03.012 Published online by Cambridge University Press

Table 2
Example of documentary entries for the Niger Bend. Included, from top to bottom, are the region number, wetness index, geographical location, year, and documentary entry.

%A 14
%F −3
%K Niger Bend
%D 1853/1854
%T Famine and drought in the Niger Bend (c. 17 N 3 W) (Barth, 1859).

%A 14
%F
%K Niger Bend
%D 1854c
%T It was, near Tombouctoo, rainy on the fifteenth of January; Barth states that rain and thunder is not unusual along this quarter of the river toward the end of January and the beginning of February.

%A 14
%F −2
%K Niger Bend
%D 1864/1865
%T Great drought in Oualata and the Soudan (Mali) (Marty, 1921).

%A 14
%F −3
%K Niger Bend
%D 1865/1866
%T Great drought and famine at Tichitt, Chinguetti, Oualata, Araouane, and Tombouctoo (Monteil, 1939).

%A 14
%F +2
%K Niger Bend
%D 1866
%T This year of abundant precipitation "ended a long period of drought which affected the regions of the Sahara between Adrar in the West and Tombouctoo in the East" (Plote, 1974).

%A 14
%F +2
%K Niger Bend
%D 1880
%T After Lenz's passage in this year, floods of the Niger were always good, until at least 1897, and the region became very prosperous (Plote, 1974). In 1880 the floods were exceptionally good, filling Lake Faguibine and the daounas, bringing fertility to a region that had been left sterile by many years of drought. This would imply that the 1870s were reasonably dry.

century station. Then quality control was carried out. For short-term records, quality control was limited to searching for outliers. This was done via comparison with the records of the analog station, identifying any historical data that fell outside the extremes in the modern station record.

The gauge data were then converted into the seven-class system, to obtain a "wetness" value for each year. Several conversion schemes were tested on modern station records, their applicability assessed by examining the resultant frequency distribution of annual anomalies. The chosen conversion distinguished the classes based on standard deviations from the long-term mean. A wetness index of zero corresponds to annual values within 0.25 standard deviations of the mean. Index values of −1/+1 are assigned to annual totals between −0.25/+0.25 and −0.75/+0.75 standard deviations. Values of −2/+2 are indicated for anomalies falling between −0.75/+0.75 and −1.25/+1.25 standard deviations. More extreme departures from the mean are classed as −/+3.

An issue that arose is how to define a "year" for stations with a rainy season that extends across two calendar years (e.g., from October to the following March). When an analysis is restricted to such rainfall regimes, annual data are typically calculated for what is termed the "rainfall year" or the "agricultural year." Such a regime prevails over most of Southern Africa. This approach is problematic when the analysis sector also includes rainfall stations with the rainy season confined to a single calendar year. This is the case for most of North Africa. Therefore we tested the impact of using calendar- versus rainfall-year for various analyses and found that the impact was relatively small. We also correlated calendar-year totals with rainfall-year totals. Correlations were extremely high for the cases with the rainy season extending across two calendar years. The explanation is that most of the interannual variability is contributed by only a few months. In Southern Africa these months generally fall between January and April. Hence, in evaluating interannual variability over Africa calendar-year totals can be used without difficulty.

Documentary information

Each documentary entry was first assigned to one of the 90 regions shown in Figure 3, based on its location. Then the entry was subjectively fit into one of the seven "wetness" classes, with the presence of key descriptors being used to distinguish the various categories. For example, a dry year might be classified as −1, but a −2 categorization requires explicit mention of a drought. An entry is classed as −3, severe drought, only if human consequences such as famine or migration are mentioned. In this scheme, a zero denotes normal conditions and +1 to +3 indicate a range from good rains to anomalously wet then very wet. The last category occurs infrequently and often indicates a flood situation.

Table 2 gives an example of documentary information for the Niger Bend region of West Africa, which includes Timbuktu. Each entry in the archive was put into this format. It includes the rainfall region the entry fits into the "wetness" class, the general geographical region, the year to which the entry pertains, and the entry itself. Unfortunately, in some cases a wetness class cannot be ascertained. In these cases, the entry is retained in the archive but not utilized for reconstructing precipitation. When feasible, geographical coordinates are given within the descriptive section. In some cases, an interpretation is given within the text of the entry. For example, if a famine is discussed by the source, an annotation indicating that the famine appeared to be related to drought or to conditions in a given month might also be included.

Limited validation of the documentary entries was carried out by comparing time series based on those entries with gauge data available for the same time and region. Few regions had enough data to carry out such a validation. An example is given in Nicholson (2001). Vogel (1989), Nash and Endfield (2002a), and Kelso and Vogel (2007) also carried out extensive validation of their reconstructions based on missionary records and found them to be very credible.

Incorporating hydrological information

The hydrologic information is difficult to interpret in terms of precipitation. Three specific issues need to be dealt with: the geographical region or regions represented by the indicator, the nature of the relationship between the indicator and precipitation,

Table 3
Algorithms for combining historical and gauge data, based on the number of entries of both. G denotes gauge records in the region in question. H represents historical entries for the region in question. If more than one gauge or historical entry is available, the average for the region and year (μG or μH) is used.

Gauge→ / Historical ↓	0	1	2	3	4+
0		G	μG	μG	μG
1	H	2/3 H + 1/3 G	1/2 H + 1/2 μG	1/3 H + 2/3 μG	μG
2	μH	μH	1/3 μH + 2/3 μG	1/3 μH + 2/3 μG	μG
3	μH	μH	2/3 μH + 1/3 μG	1/2 μH + 1/2 μG	μG
4	μH	μH	2/3 μH + 1/3 μG	1/2 μH + 1/2 μG	μG
5+	μH	μH	3/4 μH + 1/4 μG	1/2 μH + 1/2 μG	μG

Table 4
Appropriate substitute regions for rainfall regions 1 to 90 (see Fig. 3 for location). If the region number is not listed, no other region provides an acceptable substitute. The first column of the table is the region for which substitutes are needed. Subsequent columns are the regions that can be used as a replacement for the missing one.

Region #	Acceptable substitute regions									
1	6									
5	6									
6	1	5								
9	10	13	14							
10	9									
11	14	15	16							
12	15	16	17							
13	9	14	18							
14	9	11	13	15	18	19	20			
15	11	12	14	16	20					
16	11	12	15	17	20	21				
17	12	16	21	22						
18	13	14	19	23						
19	14	18	20	23						
20	14	15	16	19	21	26				
21	16	17	20	26						
22	17									
23	18	19	24	28						
24	23	25	28	29						
25	24	28	29	36						
26	20	21								
27	31									
28	23	24	25	29						
29	24	25	28	36						
31	27	32	33	87						
32	31	38	39	87						
33	31	87	88	89						
36	25	29	86							
37	38	87								
38	32	37	39	43	44	53	87			
39	32	38	44							
40	41									
41	40									
43	38	44	48	53	54	56				
44	38	39	43	48						
46	47	51								
47	46	51								
48	43	44	56	57	62					
49	50									
50	49									
51	46	47								
52	61	62	63	71						
53	38	43	54							
54	43	53	55	56						
55	54	56	58	59						
56	43	48	54	55	57	59				
57	48	56	59	60	62					
58	55									
59	55	56	57	60	66					
60	57	59	62	64	66	68				
61	52	62	63	65						
62	48	52	57	60	61	65	68	69	77	78
63	52	61	65	69	71					
64	60	66	67	68	72					
65	61	62	63	69	78					
66	59	60	64	67	72					
67	64	66								
68	60	62	64	72	73	74	77			
69	62	63	65	70	71	78				
70	69	71	78	82						
71	52	63	69	70						
72	64	66	68	73						
73	68	72	74	75						
74	68	73	75	76	77					
75	73	74	76	79	80					
76	74	75	77	79	80					
77	62	68	74	76	78	80				
78	62	65	69	70	77	80	81			
79	75	76	80							
80	75	76	77	78	79	81	83			
81	78	80	82	83						
82	70	81	83	84						
83	80	81	82							
84	82									
86	36									
87	31	32	33	37	38					
88	33	89								
89	33	88								

and conversion of the values of the indicator to the seven-point semi-quantitative scale. A thorough resolution of these issues requires detailed modeling of the hydrologic indicator (e.g., Nicholson and Yin, 2001) and/or statistical analysis of the relationship with precipitation (e.g., Nicholson, 1998b). In some cases, this has been done. In others, the hydrologic indicator was subjectively interpreted.

The indicators in question include several lakes and the maximum and minimum levels of the Nile flow during the course of the year. Lakes integrate precipitation over their catchment. For small lakes, it can generally be assumed that the lake represents the nearby surrounding region. This was the assumption for Lakes Rukwa, Chilwa, Stefanie, Turkana and Naivasha. For Lakes Victoria and Malawi, statistical correlations determined the region best represented by the lake. Lake Victoria's catchment lies within rainfall region 32 and the lake level primarily reflects catchment rainfall (Nicholson et al., 2000). Although Lake Albert receives outflow from Lake Victoria, its level is also influenced by rainfall in its catchment and inflow from the Semliki River. These sources are in rainfall region 31 and Lake Albert was assigned to this region. Lake Tanganyika's catchment lies within rainfall region 39. Lake Malawi's catchment straddles several uncorrelated rainfall regions, but statistical analysis shows that its level is most strongly correlated with rainfall in region 38 (Nicholson, 1998b). This is reasonable, as the lake's major tributaries lie in this region (Pilskaln and Johnson, 1991). Lake Chad is more difficult to interpret because its tributaries originate in equatorial regions (rainfall region 25) but the lake itself lies in the Sahel (straddling regions 15 and 20). Its level is influenced by rainfall in both the equatorial region and the Sahel (Nicholson, 1981b). Here the assumption is made that it is most representative of region 20, in which most of the over-lake rainfall occurs. This is an appropriate "compromise," in that region 20 correlates with both regions 15 and 25 (Nicholson, 2001).

The water-balance model of Yin and Nicholson (2002) demonstrated that rainfall in Lake Victoria's catchment is reflected by the year-to-year change in the level of the lake. This relationship may depend to some extent on basin characteristics and the dominant water-balance terms. However, it is assumed to be valid for the lakes utilized in creating this archive. This means that rainfall can be inferred from lake levels alone only when information about the level is available for two consecutive years.

Data on Nile flow were obtained from Toussoun (1925), who reported minimum and maximum levels at the Roda gauge (near Cairo) since the 7th century AD. Interpretation is complex because rainfall in several geographic regions influences its flow. We assumed that, to a first approximation, the summer minimum reflects rainfall in the source region of the White Nile (i.e., near Lake Victoria, region 32) and that its subsequent flood stage reflects rainfall over the Ethiopian highlands (region 27) (Lamb, 1966). As for the nature of the relationship to rainfall, it was assumed that the Nile flood stage directly reflects rainfall over the highlands but that its summer minimum reflects the year-to-year change in region 32 rainfall. The justification for the latter assumption is as follows. The year-to-year change in the level of Lake Victoria, the source of the Nile, reflects annual rainfall. Its discharge, which largely determines the summer Nile flow, is directly correlated with the level of the lake (Nicholson et al., 2000).

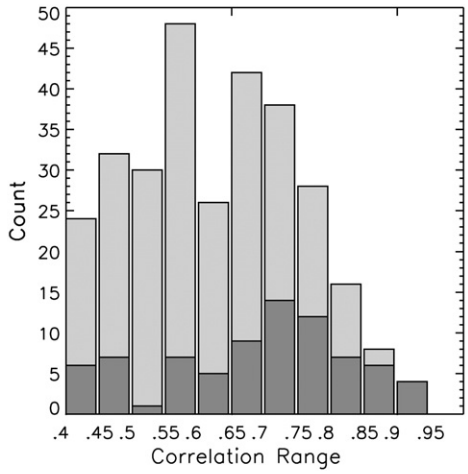

Figure 4. Histogram of correlation coefficients for all potential equivalent region substitutions. Dark shading represents the highest inter-region correlation for each region. Light shading indicates the values of all other inter-regional correlations exceeding 0.4.

The source of data for all lakes except Bosumtwi and Malawi was the lake chronologies published in Nicholson (2001). One of them, the Lake Ngami chronology, was not utilized because of ambiguities in its interpretation. The chronologies were examined to identify usable data points. It was initially assumed that levels based on reports of a continual rise or fall over a multi-year period would be usable. However, an initial test of the impact of these data showed that such generalizations masked high-frequency fluctuations, resulting in an unrealistic red spectrum of resultant rainfall time series. Thus, the only usable points from the chronologies were cases with reliable values of lake level in two consecutive years.

Lake Malawi data were obtained from Bergonzini (1998). Nearly continuous lake-level information was available from 1882 to 1900 and miscellaneous levels were available in the 1860s and 1870s.

The Lake Bosumtwi data were obtained from the Shanahan et al. (2009) contribution to the NOAA's Paleoclimate Data Center (Data Contribution Series #2009-084). Oxygen isotopes were measured in laminated sediments in a lake core. These are interpreted by those authors as a proxy for the precipitation/evaporation balance. Assuming that evaporation is less variable from year to year than precipitation, the isotopic record was assumed in this study to represent annual rainfall. The lake was assigned to the region in which it lies, rainfall region 28.

The final issue was converting the proxy data to the seven-class system. For Lake Bosumtwi a mean and standard deviation for the 19th-century data were calculated and the same conversion was used as for gauge data. This Nile data were similarly converted. For the remaining lakes, too few data points were available to use this approach. Instead the categorization was subjective, but it was based to some extent on the frequency distribution of these categories in modern rainfall data.

Merging gauge data and historical (documentary and hydrological) information

Regional averages were created by merging all available information within the region. A simple average of all data for the region was not used because one cannot assume that gauge and historical data

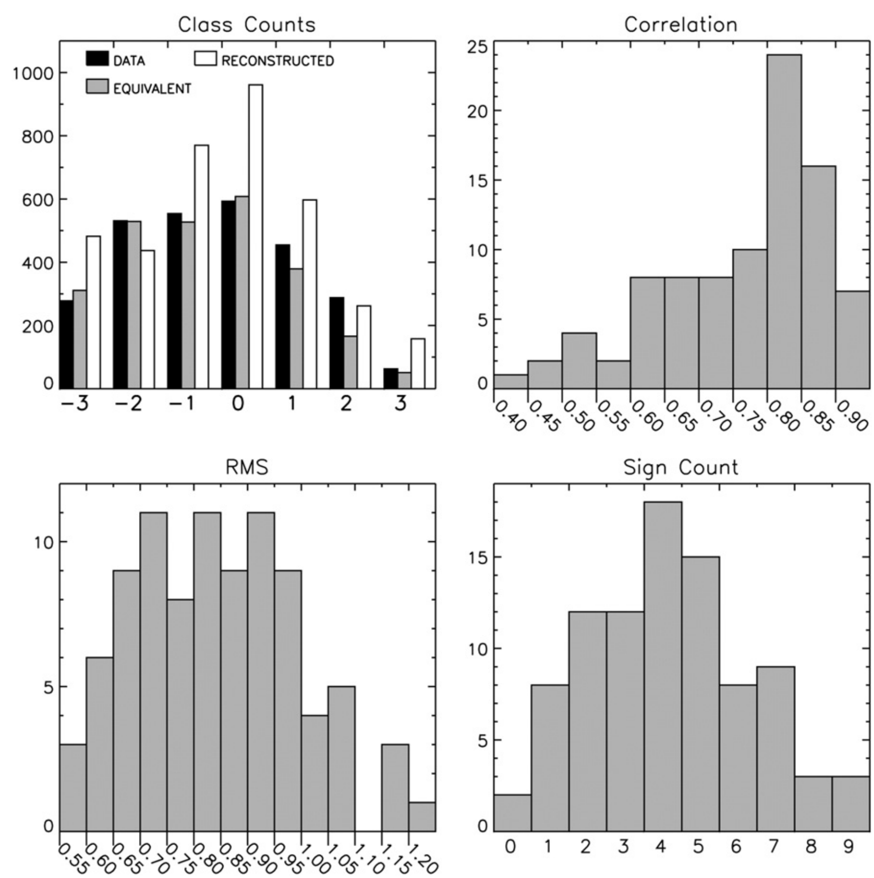

Figure 5. Histograms of data input and error statistics. Ranges of histogram classes are indicated at the bottom of each graph. Top left: For each of the seven classes of wetness indices the number of cases representing actual data input, equivalent regions, and spatially reconstructed data. Top right. Correlation between reconstructed and observed regional wetness indices. Bottom left. Root mean square error (in units of wetness classes) for reconstructed regional wetness indices. Bottom right. The histogram classes denote the number of cases in which the reconstructed and observed index values are of opposite sign. Y-axis gives the number of regions falling into each category.

https://doi.org/10.1016/j.yqres.2012.03.012 Published online by Cambridge University Press

are equally representative of large-scale conditions. By definition a drought covers a relatively large spatial area. Generally, only the larger and more intense droughts get reported. A lake integrates conditions over its catchment. On the other hand, a rain gauge provides a point measurement that, in the tropics, is probably not representative of a very large area. Lebel and Le Barbé (1997), for example, show that when the gauge density in Sahelian West Africa is two stations per 10,000 km^2, the error is roughly 5% for annual data. The error is considerably greater for a single station. Five or more stations reduce the error to 3% or less. For this reason, a credible report of drought or reliable lake levels may provide better information at the regional scale than does a single gauge measurement.

The question then arises as to how to weight historical and station data. Algorithms were derived for the weighting, based on past experience of the first author, the examination of the impact of station number on spatial averages presented in Nicholson (1986), and error estimates in the literature, such as those of Lebel and Le Barbé (1997). This reasoning suggested the algorithms shown in Table 3 for combining gauge and historical data. As an example, if only one gauge record is available, but two or more historical entries are available, the gauge data are given a zero weighting and one defers to the average of the historical entries μH. The average historical and gauge entries are equally weighted when one historical and two gauge entries are available. When four or more stations are available, the historical data are given a zero weighting. Note also that the resultant values are rounded to the nearest integer.

Quality control for the combined gauge-historical entries was provided by comparing the spread of estimates from the various sources. If more than a two-category spread existed among the entries for an individual region and year, each of those entries was re-evaluated. The choice of two categories was governed by the fact that such a spread could still occur when the entries are consistently indicating either dry conditions or wet conditions. Moreover, it would not be unusual to find such a spread in rain-gauge data. Only a small number of such cases occurred. In most of these, it was found that an error was made in determining the location or year of a documentary entry. Only eight "conflicts" could not be resolved in this way.

Statistical inference for missing regions: equivalent regions

Over much of Africa the spatial coherence of the interannual variability of rainfall goes well beyond the space scales of the regions shown in Figure 3 (Nicholson, 1993). In many cases the inter-regional correlations are very high, exceeding 0.8 over nearly a century (Nicholson, 2001). These local relationships are used to fill in gaps with data from nearby regions, when the relationship is sufficiently strong.

Nicholson (2001) determined that when the correlation between two regions exceeds 0.5 the regions are appropriate substitutes for each other. For the records utilized in that study, a correlation of 0.5 well exceeded the 1% significance level. With the longer data set utilized in the current study, a correlation of 0.4 has roughly the same significance level and is therefore adopted as a threshold for acceptable regional equivalence.

For each of the 90 regions, the "equivalent regions" are listed in Table 4. Some areas of extremely high coherence, such as the Sahel, have numerous possible substitute regions. In other locations, fluctuations have little spatial coherence and no regional substitutes are acceptable. That is the case for 12 of the 90 regions. This is generally the

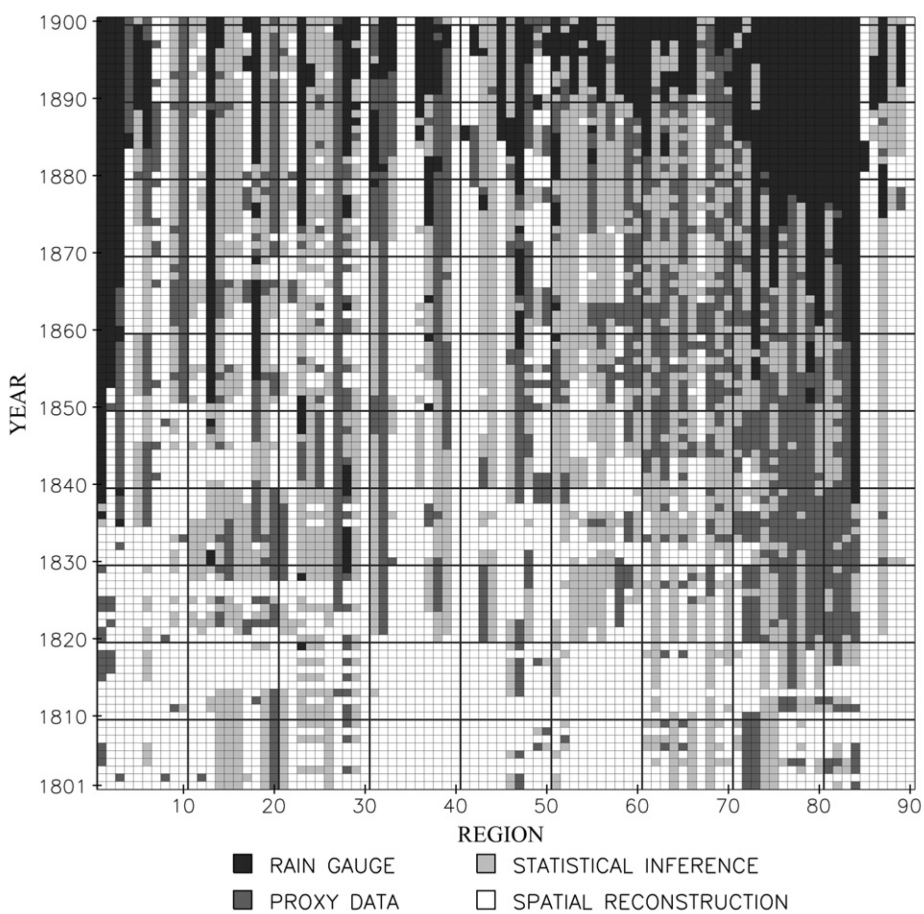

Figure 6. Matrix of sources for the final data matrix (Fig. 7). Shadings distinguish between gauge data, other direct information (documentary and hydrologic), regional substitutions, and spatial reconstruction.

https://doi.org/10.1016/j.yqres.2012.03.012 Published online by Cambridge University Press

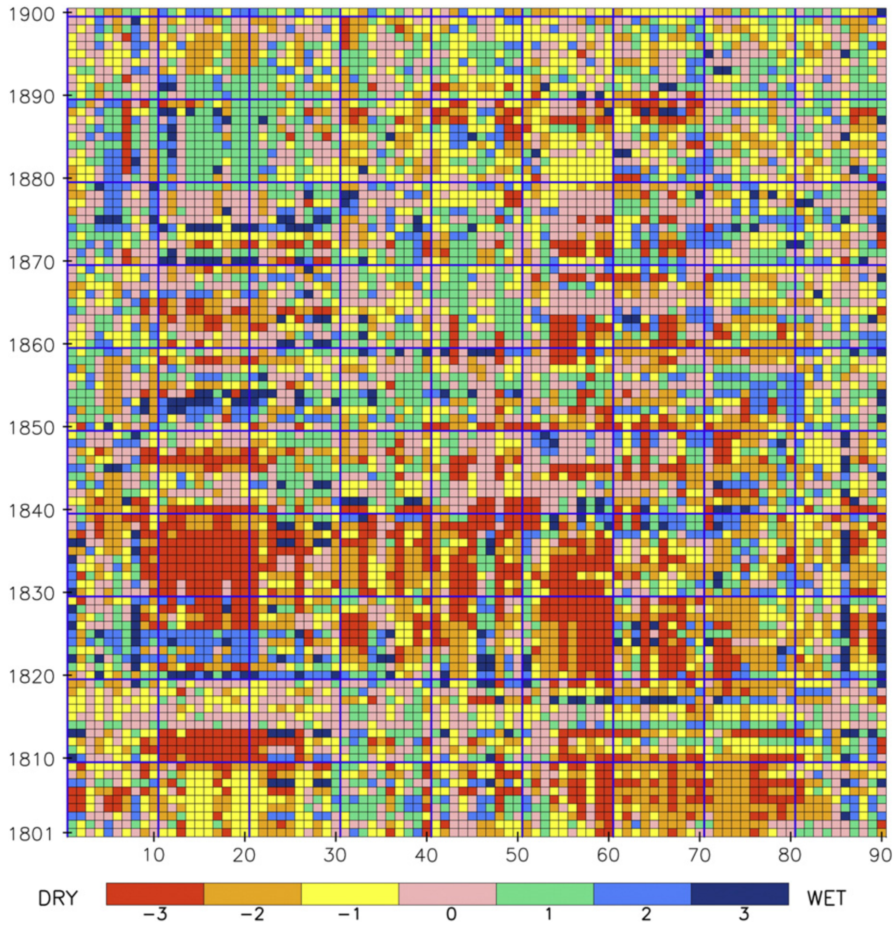

Figure 7. The final data matrix of "wetness" indices for 1801 to 1900 and the 90 regions indicated on the x-axis. From left to right, the regions approximately extend by latitude from the northern to southern extremes of Africa.

case for desert regions and for areas of complex topography, such as western equatorial Africa and Ethiopia.

Figure 4 shows the histogram of inter-regional correlations for the acceptable regional substitutions. It includes two types of information. The dark shading represents the values of the correlation coefficients between each region and the region with which it is best correlated. The light shading represents the correlation coefficients between all other pairs of regions that can be substituted for each other. Figure 4 indicates, for example, that 52 of the 78 regions have at least one other region with which the correlation exceeds 0.65.

Two approaches were tested for filling in regional gaps by substituting "equivalent" regions. In one case, the value of the most highly correlated region was adopted. In the second case, the average of the values for all possible substitute regions was used. The results showed little difference and the former approach was adopted.

Spatial reconstruction

Several methods are available for proxy-based climate field reconstruction using nonlocal relationships. In these methods, large-scale covariance information is used to calibrate instrumental data against proxy indicators (Zhang et al., 2004). The most common are multivariate principal component regression, composite plus scaling, and regularized expectation maximization ('RegEM') (Neukom et al., 2010). The appropriate method depends on the variable being estimated and the nature of the proxy variable or variables (Smerdon et al., 2011).

In this study, principal component analysis (PCA) was used in conjunction with multivariate regression to reconstruct the missing regional values for the nineteenth century. To assign a value for a specific region in a year with no data, the first step was to apply PCA on the 20th century time series of all regions with data available for the 19th-century year in question. The leading principal components (PCs) that cumulatively explain a large fraction of the total variance were retained. This serves to minimize random noise. The retained PCs were then used as the independent variables in a multivariate regression model whose dependent variable was the rainfall of the region of interest. The resultant linear regression equation was utilized to calculate the missing value. This process was repeated for all regions/years with missing data in the 19th century. Various numbers of PCs were initially retained to determine the sensitivity of the analysis to the inclusion/retention of PCs. Results were relatively stable for numbers ranging from 10 to 20 and the decision was made to retain 20 PCs.

A tally of the composition of the final data set is shown in Figure 5. In each anomaly class, regional data, equivalent regions, and spatial reconstructions each provide roughly the same number of regional index values. The exception is the zero class, in which the largest amount of data comes from spatial reconstruction. This is not surprising because in many cases the absence of data reflects normal conditions (i.e., no need to comment on the quality of the season).

Figure 6 shows the source of the "data" for each region and year. Shadings distinguish between gauge data, other direct information (documentary and hydrologic), regional substitutions, and spatial reconstruction. Overall, there is much more direct information (including gauges) available for southern Africa than for other regions. The availability of the first two classes becomes relatively high in the 1840s AD for southern Africa, 1850s for West and North

https://doi.org/10.1016/j.yqres.2012.03.012 Published online by Cambridge University Press

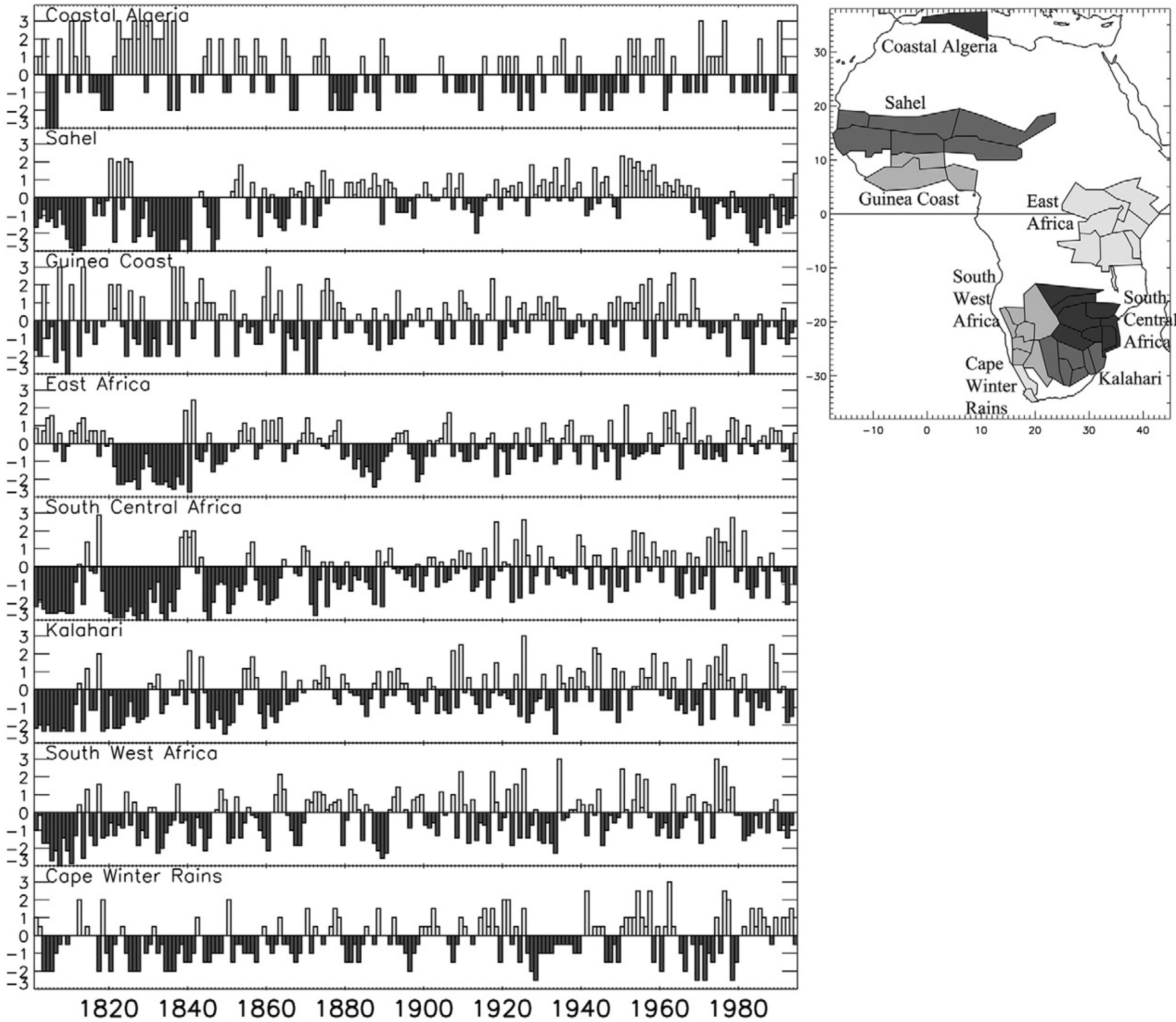

Figure 8. Select multi-region time series of "wetness" from 1801 to 1998 for eight geographical sectors of Africa (location shown in the inset map). The term "Kalahari" is used here in a broad sense as an identifying label. The Kalahari region sensu stricto is a smaller region.

Africa, and 1870s AD for equatorial Africa. On the other hand, there is little direct information or regional substitution for region 8 (eastern Sahara) or for regions 34, 35, 85 and 90 (central equatorial Africa). Prior to 1820, most of the data set is based on spatial reconstruction and the regional values become questionable.

Figure 7 shows the final data matrix. The most striking feature is the tendency for increased aridity during the first few decades of the nineteenth century. Anomalies tended to be smaller during the last half of the century, but overall rainfall was generally higher than in the first half of the century. A notable period of above normal rainfall prevailed in the Sahel (roughly regions 9 to 23) in the 1880s and early 1890s AD, but elsewhere a tendency toward more arid conditions commenced around 1880 AD.

Error estimates

An approximate estimate of error in the 19th-century data set can be made by applying the same method to reconstructing 20th-century data. Thus, for each of the 90 regions a PC analysis was performed using all data except values for that region and the PCs were utilized to reconstruct values for the region in question. The period 1929 to 1990 AD was chosen for this analysis because the regional data set is complete for these years.

The differences between the regional index value based on gauge data and that in the reconstruction were expressed in three ways. These included a correlation coefficient, a root mean square error estimate, and a "count" of the number of cases in which the series had indices of the opposite sign. The results are too voluminous to include here, but are entered as part of our contribution to the Paleoclimatology Data Center. The observed versus reconstructed values for the 90 regions mostly correlate at 0.8 or higher. The correlation is less than 0.6 for only 9 of the 90 regions. In general the root mean square error is less than one anomaly class for nearly all regions. The third statistic is a count of cases in which the anomalies have different signs. Cases of -1 and $+1$ were not in the tally, as both indicate near-normal rainfall. The error in anomaly signs is less than 10% (i.e., less than 6 cases in 62 years) for 23 of the 90 regions. It exceeds 13% for only 6 of the 90 regions.

Select examples of time series and annual anomaly patterns

Figure 8 shows time series for eight multi-region sectors, arranged roughly from north to south. These highlight some of the patterns already evident in the data matrix. With the exception of northern Algeria, all show the extensive aridity in the first three decades of the nineteenth century. Most of the sectors also show a general tendency

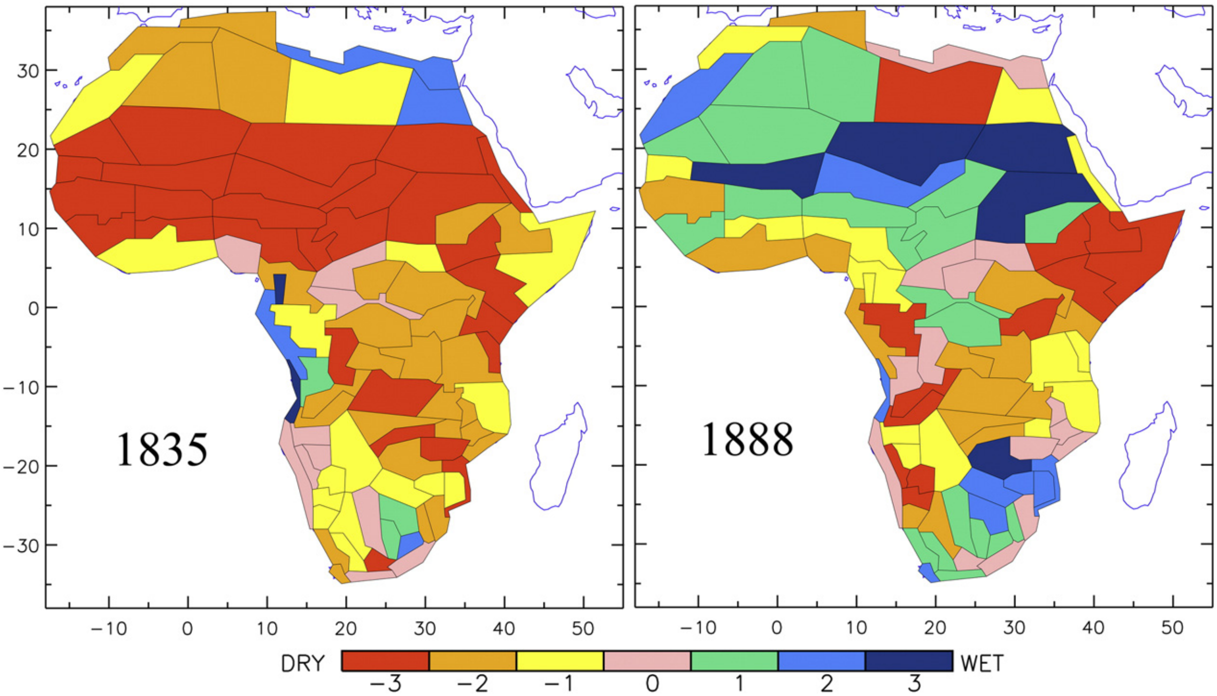

Figure 9. Rainfall anomaly patterns for 1835 and 1888. The anomalies relate to 90 regions shown in Fig. 3.

for higher rainfall throughout the early and mid-20th century, or later.

Figure 9 shows the spatial anomaly pattern for two years, 1835 AD and 1888 AD. The pattern for 1835 AD suggests that the early nineteenth-century period of aridity is nearly ubiquitous. Positive anomalies are restricted to the region near the Benguela Coast. The anomaly pattern for the year 1888 illustrates an opposition between the equatorial and subtropical regions, with dry conditions throughout most of the area between roughly 10° N and 20° S. This is one of the most common anomaly patterns during the 20th century, but did not appear frequently in the 19th century.

Archived data

The African historical data set has been contributed to the Paleoclimate Data Center in Boulder, Colorado. The working title is "Africa 2001 Precipitation" and it is filed under the first author's name. The contribution number is 2011-130. It consists of four separate entries, together with metadata and documentation. Each has been subjected to extensive quality control. Part 1 is the gauge data set (title: 19th century African gauge precipitation data set). It includes 48 stations in Algeria, 93 stations in South Africa and 283 stations scattered over the rest of Africa. Included are long-term means and standard deviations for each month and for annual totals, as well as information on location. Part 2 (title: African historical regional data) is the semi-quantitative data matrix with an annual "wetness" value for each of the 90 regions and each year of the nineteenth century (shown graphically in Fig. 7). Part 3 (title: African historical documentary information) is the documentary record based on descriptive historical sources. Its format is that shown in Table 2.

Part 2 includes extensive information about the regional data set. For each region and year the final index value is given. Also indicated is a count of the number of entries (gauge plus historical) available for the reconstruction and the source of the reconstruction for that region and year, when no actual data are available. The "source" is then indicated as spatial reconstruction or equivalence substitution. When data are available that column gives the values of each available indicator.

Part 4 (title: validation of spatial reconstruction) gives the error estimates for each region, as described above. The value of each error parameter is given for each of the 90 regions. This allows the user of the data set to have an idea of how reliable the values are for the individual regions during the 19th century.

Acknowledgments

The authors acknowledge the support of two historical grants from NOAA: NAO80AR4310731 and A46GP0285. They also thank Dave Nash and Georgina Endfield for contributing to the archive their invaluable missionary records of Southern African climate. We also thank Tom Johnson and Tim Shanahan and coauthors for lake-level data.

References

Barth, H., 1859. Travels and Discoveries in North and Central Africa (1849–1855). Cass, 3 volumes, New York.

Bergonzini, L., 1998. Bilans hydriques du lacs (Kivu, Tanganyika, Rukwa, et Nyassa) du Rift l'est africain. Musée Royal de l'Afrique-Central, Tervuren, Belgium. Annales Sciences Géologiques 183 pp.

Compo, G.P., et al., 2011. The Twentieth Century Reanalysis Project. Quarterly Journal of the Royal Meteorological Society 137, 1–28.

Fichtler, E., Trouet, V., Beeckman, H., Coppin, P., Worbes, M., 2004. Climatic signals in tree rings of *Burkea africana* and *Pterocarpus angolensis* from semiarid forests in Namibia. Trees Structure and Function 18, 442–451.

Grab, S.W., Nash, D.J., 2010. Documentary evidence of climate4 variability during cold seasons in Lesotho, southern Africa, 1833–1900. Climate Dynamics 34, 473–499.

Halfman, J.D., Johnson, T.C., 1988. High-resolution record of cyclic climatic change during the past 4 ka from Lake Turkana. Geology 16, 496–500.

Johnson, T.C., Brown, E.T., McManus, J., Barry, S., Barker, P., Gasse, F., 2002. A high-resolution paleoclimate record spanning the past 25,000 years in southern East Africa. Science 296 (113–114), 131–132.

Kelso, C., Vogel, C., 2007. The climate of Namaqualand in the nineteenth century. Climatic Change 83, 357–380.

Lamb, H.H., 1966. Climate in the 1960s — changes in the world's wind circulation reflected in prevailing temperatures, rainfall patterns and levels of the African lakes. The Geographical Journal 132, 183–212.

Lebel, T., Le Barbé, L., 1997. Rainfall monitoring during HAPEX-Sahel. 2. Point and areal estimation at the event and seasonal scales. Journal of Hydrology 188, 97–122.

Marty, P., 1921. Poème historique d'Abou Bakr Ibn Hejab, le Dimani. Bull. Comm. Et. Hist. et Scient. AOF 4, 252–263.
Monteil, V., 1939. Chroniques du Tichitt. Bull. Inst. Fond. Afr. Noire 1, 282–313.
Nash, D.J., 1996. On the dry valleys of the Kalahari. Documentary evidence of environmental change in central southern Africa. The Geographical Journal 162, 154–168.
Nash, D.J., Endfield, G.H., 2002a. A 19th century climate chronology for the Kalahari region of central southern Africa derived from missionary correspondence. International Journal of Climatology 22, 821–841.
Nash, D.J., Endfield, G.H., 2002b. Historical flows in the dry valleys of the Kalahari identified from missionary correspondence. South African Journal of Science 98, 244–248.
Nash, D.J., Endfield, G.H., 2008. 'Splendid rains have fallen'. Links between El Niño and rainfall variability in the Kalahari, 1840–1900. Climatic Change 86, 257–290.
Nash, D.J., Grab, S.W., 2010. "A sky of brass and burning winds." Documentary evidence of rainfall variability in the Kingdom of Lesotho, Southern Africa, 1824–1900. Climatic Change 101, 617–653.
Neukom, R., et al., 2010. Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. Climate Dynamics. doi:10.1007/s00382-010-0793-3.
Nicholson, S.E., 1978. Climatic variations in the Sahel and other African regions during the past five centuries. Journal of Arid Environments 1, 3–24.
Nicholson, S.E., 1979. The methodology of historical climate reconstruction and its application to Africa. The Journal of African History 20 (No. 1), 31–49.
Nicholson, S.E., 1980. Saharan climates in historic times. In: Williams, M.A.J., Faure, H. (Eds.), The Sahara and the Nile. A.A. Balkema, Rotterdam, pp. 73–200.
Nicholson, S.E., 1981a. The historical climatology of Africa. In: Wigley, T.M.L., Ingram, M.J., Farmer, G. (Eds.), Climate and History. Cambridge Press, pp. 249–270.
Nicholson, S.E., 1981b. Lake Chad and its relation to Sahelian climate history. The Sahara: Ecological Change and Early Economic History, pp. 35–60.
Nicholson, S.E., 1986. The spatial coherence of African rainfall anomalies. Interhemispheric teleconnections. Journal of Climate and Applied Meteorology 25, 1365–1381.
Nicholson, S.E., 1993. An overview of African rainfall fluctuations of the last decade. Journal of Climate 6, 1463–1466.
Nicholson, S.E., 1996. Environmental change within the historical period. In: Goudie, A.S., Adams, W.M., Orme, A. (Eds.), The Physical Geography of Africa. Oxford University Press, Oxford, pp. 60–75.
Nicholson, S.E., 1998a. Historical fluctuations of Lake Victoria and other lakes in the northern Rift Valley of East Africa. In: Lehman, J.T. (Ed.), Environmental Change and Response in East African Lakes. Kluwer, Dordrecht, pp. 7–35.
Nicholson, S.E., 1998b. Fluctuations of Rift Valley Lakes Malawi and Chilwa during historical times. A synthesis of geological, archaeological and historical information. In: Lehman, J.T. (Ed.), Environmental Change and Response in East African Lakes. Kluwer, Dordrecht, pp. 207–231.
Nicholson, S.E., 1999. Historical and modern fluctuations of Lakes Tanganyika and Rukwa and their relationship to rainfall variability. Climatic Change 41, 53–71.
Nicholson, S.E., 2001. A semi-quantitative, regional precipitation data set for studying African climates of the nineteenth century, part I. Overview of the data set. Climatic Change 50, 317–353.
Nicholson, S.E., Yin, X., 2001. Rainfall conditions in equatorial East Africa during the nineteenth century as inferred from the record of Lake Victoria. Climatic Change 48, 387–398.
Nicholson, S.E., Yin, X., Ba, M.B., 2000. On the feasibility of using a lake water balance model to infer rainfall. Hydrological Sciences Journal 45, 75–95.
Nicolini, G., Tarchiani, V., Saurer, M., Cherubini, P., 2010. Wood-growth zones in *Acacia seyal Delile* in the Keita Valley, Niger: is there any climatic signal? Journal of Arid Environments 74, 355–359.
Pilskaln, C.H., Johnson, T.C., 1991. Seasonal signals in Lake Malawi sediments. Limnology and Oceanography 36, 544–557.
Plote, G., 1974. L'Afrique sahelienne se desseche-t-elle? B. R. G. M., Orleans, France 61 pp.
Russell, J.M., Verschuren, D., Eggermont, H., 2007. Spatial complexity of 'Little Ice Age' climate in East Africa: sedimentary records from two crater lake basins in western Uganda. The Holocene 17, 183–193.
Shanahan, T.M., et al., 2008. The formation of biogeochemical laminations in Lake Bosumtwi, Ghana, and their usefulness as indicators of past environmental changes. Journal of Paleolimnology 40, 339–355.
Shanahan, T.M., Coauthors, 2009. Atlantic forcing of persistent drought in West Africa. Science 324, 377–380.
Smerdon, J.E., Kaplan, A., Chang, D., Evans, M.N., 2011. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. Journal of Climate 24, 1285–1309.
Supan, A., 1898. Die Verteiliung des Niederschlags auf der festen Erdoerfläche. Petermanns Mitteilungen, Ergänzungsheft No. 124, 103 pp.
Therrell, M.D., Stahle, D.W., Ries, L.P., Shugart, H.H., 2006. Tree-ring reconstructed rainfall variability in Zimbabwe. Climate Dynamics 26, 677–685.
Touchan, R., Anchukaitis, K.J., Meko, D.M., Attalah, S., Baisan, C., Aloui, A., 2008. Long term context for recent drought in northwestern Africa. Geophysical Research Letters 35, L13705.
Touchan, R., Anchukaitis, K.J., Meko, D.M., Sabir, M., Attalah, S., Aloui, A., 2011. Spatiotemporal drought variability in northwestern Africa over the last nine centuries. Climate Dynamics 37, 237–252.
Toussoun, O., 1925. Memoire sur l'histoire du Nil. Memoires a l'Institut d'Egypt, vol. 18, pp. 366–404.
Trouet, V., Haneca, K., Coppin, P., Beeckman, H., 2001. Tree ring analysis of *Brachystegia spiciformis* and *Isoberlinia tomentosa*. Evaluation of the enso-signal in the Miombo woodland of eastern Africa. IAWA Journal 22, 385–399.
Vogel, C., 1989. A documentary-derived climatic chronology for South Africa. Climatic Change 14, 291–308.
Woodruff, S.D., 2011. ICOADS Release 2.5. Extensions and enhancements to the surface marine meteorological archive. International Journal of Climatology 31, 951–967.
Yin, X., Nicholson, S.E., 2002. Interpreting annual rainfall from the levels of Lake Victoria. Journal of Hydrometeorology 3, 406–416.
Zhang, Z., Mann, M., Cook, E., 2004. Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. The Holocene 14, 502–516.