Holocene variations in the global hydrological cycle quantified by objective gridding of lake level databases

A. E. Vau and K. Gajewski
Laboratory for Paleoclimatology and Climatology, Department of Geography, University of Ottawa, Ottawa, Ontario, Canada

Abstract. Lake level fluctuations provide evidence about past variations in the global hydrological balance. The geostatistical approach is here used to more objectively identify global patterns using an ensemble of lake level databases by examining spatial autocorrelation between sites. The spatial structures of the lake level data are then modeled and grids produced for the last 12,000 years at 3000-year intervals using ordinary and indicator kriging techniques. The two gridding techniques produced almost identical estimated regional lake status patterns, thus suggesting a robust estimation. The resulting lake-status grids are in general agreement with previous paleoclimatic reconstructions using only site-by-site lake status point maps; however, the reduction of local fine-scale variability resulted in more coherent regional spatial patterns in areas of high local variability. The 6 ka lake-status grids were compared to simulations of four atmospheric general circulation models to illustrate their usefulness in validating broad-scale climate model outputs.

1. Introduction

To study past climates, atmospheric general circulation models (AGCMs) are used to simulate broad-scale patterns in atmospheric circulation, and test hypotheses of forcing mechanisms. Numerous time periods throughout the Earth’s history have been simulated using AGCMs, but the Quaternary has received the most attention because the boundary conditions are better known, and there are large amounts of data available for validation. Modeling experiments of past climates [e.g., Kutzbach, 1981; Boer et al., 1984; Kutzbach and Guetter, 1986; Rind et al., 1989; Schulz et al., 1992; Marshall et al., 1997; Hewitt and Mitchell, 1997; Teixier et al., 1997; Kutzbach et al., 1998; and Vettoretti et al., 1998, 2000a, 2000b] need to be compared to maps of paleoclimate data [COHMAP, 1988; Wright et al., 1993; Webb, 1998]. Data-model comparisons can contribute to paleoclimate research by determining which climate model simulations best represent past climates. To be useful for comparison with model output, paleoclimate data need to (1) provide broad spatial coverage, global if possible, and (2) be derived independently of the model output.

Lake level variations provide one source of information about past climates. A lake responds to the hydrological balance over its surface and catchment over many timescales [Street-Perrott and Harrison, 1985]. Local nonclimatic and climatic factors control lake depth and area extent, and any individual lake integrates these factors in its record of water level fluctuations.

Lake level fluctuations leave traces in the shores, including drowned or raised beaches. Changes in the area and depth influence the biotic and abiotic processes of the lake and its catchment, and these leave a record in lake sediments, so sediment-stratigraphic analyses can be used to reconstruct past lake level fluctuations [Harrison et al., 1991]. Various combinations of geomorphic, geological, and biostratigraphic data provide evidence of variations in lake area and depth [Street-Perrott and Harrison, 1985]. Although several factors controlling a lake’s response to climate have been proposed for several timescales, the resolution of these data sets are at the millennial scale. It has been suggested that lakes are in equilibrium with climate at this scale [Mason et al., 1994]. Street and Grove [1976, 1979], Street-Perrott and Harrison [1985], Harrison and Metcalfe [1985a, 1985b], Harrison [1989], Guitot et al. [1993], Yu and Harrison [1995a, 1995b], and Cheddadi et al. [1997] illustrate the use of lake level data in paleoclimate research. Typically, site-by-site point maps are presented, illustrating lake level status (i.e., the variations in lake depth) for an array of lakes. Although the regional-scale hydrological status can be visually interpreted, any one region may contain considerable intersite variability, and this leads to ambiguity in data-model comparison. This sub-regional-scale variability has a number of climatic or nonclimatic sources which may be significant for geomorphic and ecological interpretation, but it obscures the regional-scale signal that is of interest in comparisons with global climate models. When the majority of lakes in a region fluctuate in a similar fashion, climate is probably the dominant controlling factor [Harrison et al., 1991], and it is this pattern that is interpreted in the above studies.

Lake level data have been used to verify climate model simulations by making qualitative visual comparisons of the observations in lake-status change [Street and Grove, 1979; Harrison and Metcalfe 1985a, 1985b; Harrison et al., 1991; Yu and Harrison, 1996] or by semiquantitative regional averages of lake data [Kutzbach and Street-Perrott, 1985; Street-Perrott, 1986; Harrison, 1989; Webb et al., 1993]. More recently, Qin et al. [1998] quantitatively compared the lake data that had continuous records for 0 and 6 ka with model results. These approaches can be improved by using a method...
to determine the regional signal in areas that contain fine-scale variability. The large-scale structure of environmental data can be modeled using spatial statistics if the statistical effects of any significant small-scale variability can be reduced. In this paper, lake level data are analyzed to identify broad-scale regional patterns in a 3-ka interval for the past 12,000 years. A geostatistical approach is used to measure the spatial autocorrelation between sites over different spatial scales at specific time periods in order to extract the regional patterns at the global scale. Once patterns are identified, the gridding of these spatial patterns requires a model. Geostatistical models enable estimation at unknown locations between sites where the lake status is measured. Estimation enables mapping of the spatially coherent patterns in an objective way. The mapping is done here under two assumptions. First, it is assumed that climate is spatially continuous within regions. It then follows that the prevailing climate conditions, favoring high lake level in a given region, will change along a climatic gradient toward climatic conditions favoring low lake levels through a transitional region over which intermediate lake status will prevail. Next, again assuming that climate is spatially continuous within regions, we do not assume a spatial climate gradient. Therefore the lake level data are assumed to be spatial categorical data with mutually exclusive classes. It then follows that prevailing climate conditions favoring high lake level in a given region does not necessarily change along a climatic gradient toward a low lake-level region. All three lake statuses are independently analyzed. Therefore regions of high lake levels can be adjacent to low or intermediate regions of lake level status. As an example of the application of these grids, we then compare them to climate model simulations at 6 ka.

2. Data

2.1. Lake Level Data

A global lake-level database was assembled using three available lake-status data sets in order to maximize geographical coverage. The Oxford lake-level database [Street-Perrott et al., 1989; COHMAP Members, 1994], which emphasizes closed-basin lakes, is global in extent and covers the last 30 ka estimated at 1000-year intervals. However, geographical coverage is uneven. The European lake-status database [Yu and Harrison, 1995a] contains records from 118 sites across Europe spanning 30 ka estimated at 500-year intervals. In this database, temperate or overflowing lakes are also included. The Former Soviet Union (FSU) and Mongolia lake-status database [Tarasov et al., 1994, 1996] consists of 98 basins for the FSU and five basins from Mongolia. It covers the last 15 ka at 500-year intervals, where western Russia is particularly well represented. Lakes where water level change appears to have been influenced by nonclimatic factors, such as tectonism, glacier fluctuations, or human impact have been excluded from these databases [Street-Perrott and Harrison, 1985; Yu and Harrison, 1995a; Tarasov et al., 1994, 1996].

The database used in this study thus consists of records of paleolake level reconstruction at 1000-year intervals for 579 lakes spanning the last 30 ka. However, site density decreases greatly earlier than 12 ka (Figure 1); therefore only the past 12,000 years will be analyzed in this paper. Each of the

![Figure 1](Image.png)

**Figure 1.** Distribution of available sites, and lake status for 0, 3, 6, 9, and 12 ka.
Table 1. Data-Model Criterion Index Scheme [modified from Qin et al., 1998]

<table>
<thead>
<tr>
<th>Lake Status (6-0k)</th>
<th>P-E (6-0k)</th>
<th>Agreement</th>
<th>Disagreement</th>
<th>Kriging Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF ΔS &gt; 0.5 and ΔP-E &lt; 0</td>
<td>1</td>
<td>0</td>
<td>ordinary</td>
<td></td>
</tr>
<tr>
<td>IF ΔS &gt; 0 and ΔP-E &lt; 0</td>
<td>1</td>
<td>0</td>
<td>indicator</td>
<td></td>
</tr>
<tr>
<td>IF ΔS &lt; -0.5 and ΔP-E &gt; 0</td>
<td>1</td>
<td>0</td>
<td>ordinary</td>
<td></td>
</tr>
<tr>
<td>IF ΔS &lt; 0 and ΔP-E &gt; 0</td>
<td>1</td>
<td>0</td>
<td>indicator</td>
<td></td>
</tr>
<tr>
<td>IF ΔS &lt; 0.5 and &gt; -0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>ordinary</td>
<td></td>
</tr>
<tr>
<td>IF ΔS = 0</td>
<td>0.5</td>
<td>0.5</td>
<td>indicator</td>
<td></td>
</tr>
</tbody>
</table>

databases was verified using the available documentation. Inconsistencies such as duplicate data and coding was made consistent with the Oxford lake level database. The global lake level database contains continuous and discontinuous chronological sequences; sites that have no record for a particular time interval are given a “no-status” code. Chronologies are provided either by 14C dating or correlation with a well-established regional stratigraphy, and dating control follows on the COHMAP scheme [Webb, 1985].

Coding of the lake status (or relative water depth) for each site has been standardized into three categories in order to render all basins comparable regardless of size [Street-Perrott and Harrison, 1985]. Categorization of the Oxford lake level database is based on the absolute altitudinal range of fluctuation, where low lake status is assigned to 0-15% of the total range of fluctuation, including dry lakes; intermediate status to 15-70% of the total range and high status to 70-100% of the total range of fluctuation, including overflowing lakes [Street-Perrott and Harrison, 1985]. The boundaries between lake status classes in the European and FSU and Mongolia databases are defined so as to obtain a broadly similar frequency of occurrence of each lake status category [Yu and Harrison, 1995a; Tarasov et al., 1994, 1996].

2.2. Model Simulations

The maps derived in this study were compared to four climate simulations of 6 ka and modern, which were available to us in order to illustrate the use of these lake-status grids. These model outputs are two versions of the National Center for Atmospheric Research (NCAR) community climate model (CCM0 and CCM1) [Kutzbach and Guetter, 1986; Kutzbach et al., 1998; COHMAP Members, 1988] and two versions of the Canadian Center for Climate Modeling and Analysis (CCCMA) general atmospheric circulation model [Vettoretti et al., 1998].

The NCAR CCM0 and CCM1 models have a resolution grid size of ~4.5° latitude by 7.5° longitude corresponding to an R15 spectral truncation run. The CCM0 simulation [Kutzbach and Guetter, 1986; COHMAP Members, 1988] was

Figure 2. Variogram analysis using all available points at (a) 0 ka and (b) 6 ka and after filtering at (c) 0 ka and (d) 6 ka.
run in “perpetual” January and July mode with prescribed effective soil moisture. Sea surface temperatures (SSTs) and sea ice distribution at 6 ka were prescribed to be the same as in the control run. CO$_2$ was set to 330 ppmv in both 6 ka and control experiments. For this paper the computed annual P-E averages (mm/year) were calculated following Qin et al. [1998], using the January and July averages weighted accordingly to the length of the summer and winter half year. The CCM1 simulation [Kutzbach et al., 1998] was run with a full seasonal cycle. Interactive soil moisture was calculated using a one-layer soil with a field capacity of 15 cm. There was a 50m mixed-layer ocean allowing SSTs and sea ice distribution to be calculated. CO$_2$ was set to a “preindustrial” level of 267 ppmv in both 6 ka and control simulations. In this case, the computed annual P-E (mm/year) averages were calculated using the monthly averages and weighted according to the length of each month [Qin et al., 1998].

The two sets of experiments from the CCCMA represent a second generation AGCM (GCMII). Both model simulation outputs used in this paper have an ~3.75° latitude by 3.75° longitude grid size corresponding to a T32 spectral triangular truncation [Petit et al., 1998]. The first model (hereinafter referred to as CCCfex) was run for 12 annual cycles where the last 10 annual cycles were used to compute monthly, seasonal, and annual climatology averages. The CCCfex simulation uses prescribed SSTs and sea ice for both 6 ka and control experiments. CO$_2$ concentrations were set to “modern” levels of 345 ppmv for the control simulation while a “preindustrial” CO$_2$ level of 280 ppmv was prescribed for the 6 ka simulation. In the second set of simulations (hereinafter referred to as CCCcex) the model calculates SSTs and sea ice using a mixed-layer ocean and sea-ice module [McFarlane et al., 1992]. CO$_2$ is set to a “preindustrial” level of 280 ppmv for both the control and the 6 ka simulations. The CCCcex control simulation was run for 50 annual cycles, where the last 10 annual cycles were averaged to produce monthly, seasonal, and annual modern climatologies. The CCCcex simulation for 6 ka was run for 40 annual cycles, where the last 10 annual cycles were averaged to produce the 6 ka climatologies. In both sets of simulations, annual P-E averages (mm/year) were computed using the annual P-E (mm/d) weighted according to the length of a year.

### 3. Methods

In this paper the geostatistical approach is used to describe, measure, and grid the dependence between lake sites at different spatial and temporal scales. Results include the (1) identification of large-scale patterns in the lake-status data, (2) identification of small-scale variability or influential outliers, and (3) modeling of the underlying spatial structure. These will permit robust estimates of regional lake status through time.

#### 3.1 Variogram Analysis

To model these data and permit the gridding of the lake status maps, variograms were estimated. The variogram or indicator

| Table 2. Summary of Local Indicators of Spatial Autocorrelation (LISA) |
|--------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 ka | a | b | c | 3 ka | a | b | c | 6 ka | a | b | c |
| Distance | + | - | % | Ratio | Distance | + | - | % | Ratio | Distance | + | - | % |
| 5 | 3.9 | 0.4 | 95.7 | 0.1 | 5 | 3.3 | 1.1 | 95.6 | 0.3 | 5 | 2.7 | 1.6 | 95.7 | 0.6 |
| 10 | 8.6 | 1.5 | 89.9 | 0.2 | 10 | 5.7 | 3.6 | 90.7 | 0.6 | 10 | 6.4 | 3.5 | 90.1 | 0.5 |
| 25 | 18.3 | 3.7 | 77.9 | 0.2 | 29 | 16.4 | 8.7 | 74.9 | 0.5 | 27 | 17.9 | 7.5 | 74.7 | 0.4 |
| 33 | 21.1 | 4.1 | 74.8 | 0.2 | 90 | 27.6 | 22.4 | 50.0 | 0.8 | 81 | 33.6 | 16.5 | 49.9 | 0.5 |
| 76 | 42.1 | 8.0 | 49.9 | 0.2 | 179 | 39.6 | 29.5 | 30.9 | 0.7 | 176 | 44.5 | 25.1 | 30.4 | 0.6 |
| 150 | 57.9 | 11.0 | 31.0 | 0.2 | 213 | 44.3 | 31.1 | 24.6 | 0.7 | 213 | 46.4 | 28.0 | 25.6 | 0.6 |
| 200 | 65.8 | 12.7 | 21.5 | 0.2 | 300 | 49.2 | 35.0 | 15.8 | 0.7 | 300 | 52.8 | 30.7 | 16.5 | 0.6 |
| 300 | 73.5 | 14.0 | 12.5 | 0.2 | 400 | 51.6 | 38.5 | 9.8 | 0.7 | 400 | 57.9 | 32.3 | 9.9 | 0.6 |
| 400 | 79.1 | 14.8 | 6.2 | 0.2 | 500 | 53.3 | 39.9 | 6.8 | 0.7 | 500 | 60.0 | 33.1 | 6.9 | 0.6 |
| 500 | 80.6 | 15.5 | 3.9 | 0.2 | | | | | | | | | |
| 9 ka | a | b | c | 12 ka | a | b | c |
| Distance | + | - | % | Ratio | Distance | + | - | % |
| 5 | 3.0 | 0.6 | 96.3 | 0.2 | 5 | 2.9 | 0.0 | 97.1 | 0.0 |
| 10 | 5.2 | 3.0 | 91.8 | 0.6 | 10 | 5.3 | 0.0 | 94.7 | 0.0 |
| 34 | 16.2 | 9.5 | 74.4 | 0.6 | 67 | 17.1 | 8.2 | 74.7 | 0.5 |
| 93 | 28.4 | 22.0 | 49.7 | 0.8 | 162 | 31.8 | 18.2 | 50.0 | 0.6 |
| 200 | 37.5 | 35.4 | 27.1 | 0.9 | 200 | 34.1 | 22.4 | 43.5 | 0.7 |
| 300 | 44.2 | 38.4 | 17.4 | 0.9 | 300 | 47.1 | 20.6 | 32.4 | 0.4 |
| 400 | 50.0 | 38.4 | 11.6 | 0.8 | 400 | 50.0 | 29.4 | 20.6 | 0.6 |
| 500 | 50.3 | 41.8 | 7.9 | 0.8 | 500 | 50.0 | 35.9 | 14.1 | 0.7 |

Distance expressed in kilometers and represent small distances (5 and 10 km), first quartile, median, mean, third quartile, and higher interpoint distances until maximum geographical coverage at 500 km.

a Local spatial autocorrelation (+/-) expressed in percentages of sites available (i.e., 0 ka - 535, 3 ka - 366, 6 ka - 375, 9 ka -328, and 12 ka - 170).

b Percentage of sites with no neighbors.

c Ratio of negative to positive correlations.
Figure 3. Variograms and fitted models under continuous assumption at (a) 0 ka and (b) 6 ka, and using indicators at 0 ka for (c) high, (e) intermediate, (g) low and at 6 ka for (d) high, (f) intermediate, and (h) low lake-status categories.
variogram was given as

$$\gamma(h) = \frac{1}{2} \left[ N(h) \right] \sum [z_i - z_j]^2,$$

where $\gamma(h)$ is semivariance for interval distance class $h$; $z_i$ is measured sample value at point $i$, $z_j + h$ is measured sample value at point $i + h$; and $N(h)$ is total number of sample couples for the lag interval $h$, are spatial statistics used to describe the extent or range of spatial autocorrelation at different spatial scales. These statistics describe the spatial continuity or autocorrelation as a function of distance between pairs of sample sites [Isaaks and Srivastava, 1989; Deutsch and Journel, 1992]. As with many statistics, these are sensitive to outliers, and it may not be possible to model the underlying spatial structure if these outliers are too influential.

3.2 Local Spatial Analysis or Low-Pass Spatial Filtering

We next need to identify objectively only the large-scale signal and remove the fine-scale variability in the data, which will enable the estimation of the variogram. One approach is to arbitrarily remove outliers from the data using variogram tools [Isaaks and Srivastava, 1989; Getis and Ord [1992] suggested that a more appropriate approach to deal with influential points is to focus on local patterns of spatial autocorrelation. In this paper we use the local indicators of spatial association (LISA) statistics [Anselin, 1995] given as

$$I = \frac{1}{\sum w_{ij}(z_i - z_j)^2 / n} \sum w_{ij}(z_i - z_j),$$

where $z_i$ is some variable measured at different points $(i,j)$ in space, $w_{ij} = 1$ for those points, $j$, within a prespecified distance (neighborhood) of locality $i$ and zero otherwise, which decompose global statistics, such as the Moran’s I, into the contribution of each site within a specific distance radius.

A local indicator of spatial association (LISA) statistic, such as the local Moran’s I, in its standardized form, within specific distance radii, identifies sites that are different from (negative spatial autocorrelation) or similar to their neighbors (positive spatial autocorrelation).

The local Moran I statistic is used in this study to filter the data to extract the large-scale signal at a consistent scale. Alternative methods previously used include using only sites with continuous records [Harrison, 1989; Harrison et al., 1993], quality dating control schemes [Guiot et al., 1993; Cheddadi et al., 1997; Qin et al., 1998], and the use of maps using all points with a qualitative interpretation based on the majority lake status of an area [Street-Perrrot and Harrison, 1984; Harrison and Metcalfe, 1985a, 1985b]. This study differs in that the exclusion is done objectively by deciding on the scale of study and maintaining this scale through the study.

3.3 Gridding and Mapping

Gridding of the lake level data requires a model for estimating values in areas that have not been sampled. Geostatistical models (spherical, exponential, Gaussian, linear) were fitted to the least variable variograms. The choice of model was decided using best fit statistics such as $r^2$ (regression) and RSS (reduced sum of squares) coefficients. In cases where two or more models gave similar values, preference was given to the model that yielded the largest range or distance of spatial autocorrelation. Gridding was performed using ordinary kriging under the climate-continuum assumption and indicator kriging under the independent-region assumption. Two different estimation methods were used to test the robustness of the estimated lake level patterns. Both methods involve weighted linear combinations where weights assigned by local estimation should account for both the

![Figure 4. Lake status grids under ordinary kriging or continuous assumption at 0, 3, 6, 9, and 12 ka. Scale goes from 1 (wet) to 3 (dry).](image)
distance and the possible redundancy between sample sites [Isaaks and Srivastava, 1989], as in the case of clustering. In addition, this study uses local estimation rather than global to take into account the irregular distribution of sites, clustering, and scale effect present in the data which may not be representative of the globe and to limit extrapolation in areas with sparse data representation.

Ordinary kriging estimates unknown locations using a common model for all three lake statuses so that continuous estimates of lake status are possible. Indicator kriging provides a value of uncertainty or probability of occurrence for the binary transformed data of each independent lake status, rather than an estimate of the value [Deutsch and Journel, 1992; Goovaerts, 1997]. Indicator kriging involves three distinct steps: (1) the binary transformation of each observation into a vector corresponding to the three lake status categories (e.g., high = 1,0,0), (2) indicator variogram analysis and modeling of each independent category, and (3) reclassification of each estimated grid node probability of occurrence into a corresponding lake status (i.e., high, intermediate, low) based on the maximum probability vector [Goovaerts, 1997].

3.4 Data-Model Comparison

The simulated precipitation minus evaporation (ΔP-E, 6 ka minus modern) value from each model output is first interpolated to the estimated lake-status grid. The difference in lake status (ΔS, 6 ka minus modern) at 6 ka of each estimated grid node is then compared with the interpolated ΔP-E at 6 ka using a modified version of the Qin et al. [1998] agreement index, where 1 means an agreement and 0 a disagreement between lake status (ΔS) and model-simulated ΔP-E at 6 ka. Where no evident lake status change occurs (ΔS = 0), a potential agreement value of 0.5 is assigned to the data model comparison. However, a potential agreement value is not attributed to the model simulation that yields the smallest ΔP-E value as in the work of Qin et al. [1998] (Table 1). This is done in order to emphasize the agreement between data and different spatial resolution models. We then compute the number of grid nodes that are in agreement, potential agreement, and disagreement expressed in percentages between data and model output.

4. Results

4.1 Exploratory Spatial Data Analysis

Patterns in spatial data may be difficult to identify because of fine-scale variability [Isaaks and Srivastava, 1989]. A detailed visual inspection was conducted using symbol maps to analyze the spatial distribution of lake status sites, and aid in

![Figure 5. Probabilities of occurrence for each lake status category using indicator kriging method at 0 and 6 ka.](image-url)
the identification of distinct spatial patterns for all five time periods under investigation.

In general, some regions of the world, such as Europe and northern Africa, are well represented, while others are underrepresented, for example, most of eastern Asia and South America (Figure 1). Site density increases from 12 ka to present (Figure 1). A more detailed visual inspection reveals the presence of fine-scale spatial variability in most regions, especially at 3 and 9 ka (Figure 1). In spite of this spatial variability, distinct patterns in lake status can be visually identified at 0 ka, especially over North Africa and North America (Figure 1). At 3 ka the patterns in most regions of the world are not so obvious as at 0 ka. In addition, the number of available sites in several regions has declined (Figure 1). Lake status at 6 ka visually exhibits more coherent patterns, with less fine-scale variability than at 3 ka. However, in many areas both high and low lake statuses are found in close proximity. Fine-scale spatial variability is evident in most regions of the world at 9 ka and is comparable to that seen at 3 ka, which suggests a possible transition in the global hydrological cycle. Africa is the only region that exhibits low spatial variability (Figure 1). Lake status at 12 ka shows more homogenous spatial patterns in lake level status in most regions of the world, due in part to the decreased geographical coverage (Figure 1).

4.2 Variogram Analysis

Variograms were fitted to all data points for all five time periods [Viau, 1999], but only 0 and 6 ka are shown here. The particular values of the separation distance, on the x axis of the variograms, illustrate one of the multiple trials conducted, all of which resulted in similar spatial structures. In general, spatial continuity or autocorrelation is present at 0 ka, because the semivariance increases as separation distances increase (Figure 2a). The nugget (γ intercept), or fine-scale variability at lower lag distances, is relatively low, suggesting that a large number of sites in close proximity are similar in lake status. However, the variogram at 6 ka shows a lack of spatial autocorrelation. Many sites in close proximity exhibit negative spatial autocorrelation and therefore local variability is high (Figure 2b).

The results for 6 ka suggest that modeling the underlying spatial structure using all data points is not possible. Indeed, the analysis for all five time periods under both assumptions (i.e., continuous interval versus categorical indicators) showed that spatial autocorrelation was present at 0 ka and to a lesser extent at 12 ka. However, the periods of 3, 6, and 9 ka showed a lack of spatial autocorrelation at all distances investigated [Viau, 1999]. These findings reveal that there are a large number of sites that exhibit opposing lake status (i.e., high-low) within relatively small distances (lower lag distances) [Viau, 1999].

4.3 Identification of Broad-Scale Regional Lake-Status Patterns

As stated above, we used the Local Moran I as a LISA statistic [Anselin, 1995], to determine local positively correlated sites and identify sites which display local negative spatial autocorrelation within specific radii, using Rookcase software [Sawada, 1999]. The Local Moran I statistic is used

Figure 6. Lake status grids after reclassification using indicator kriging for 0, 3, 6, 9, and 12 ka.
here to optimize both the spatial scale of study and the geographical coverage and also to offer a critical regional evaluation on the extent of local influential outliers. Maximum geographical coverage was obtained at 500 km distances radii, leaving less than 10% of sites with no neighbors for all five time periods (14% at 12 ka, Table 2). The radius of 500 km is chosen as the optimal scale of study because the ratio between negative and positive autocorrelation increases with distance. To maximize geographical coverage, the spatial scale must be increased to include as many data points as possible, and this increases the ratio of negative autocorrelation.

The ratio between positive and negative spatial autocorrelation, especially at small distances, is low for 0 and 12 ka but high for 3, 6, and 9 ka (Table 2), which confirms the variogram analysis as shown for 6 ka (Figure 2b). Sample sites exhibiting negative spatial autocorrelation within the 500 km scale were at this point excluded from this study. This filtering exercise permitted the extraction of the large-scale climate signal present in the lake data under a consistent scale of study. The resulting filtered data sets revealed that no major area of the globe suffered major losses in representation, rather, the distribution of excluded sites was more or less evenly distributed, suggesting that the application of the LISA statistic was successful [Vila, 1999].

4.4 Gridding and Mapping

The spatial filtering of the lake data resulted in a more clearly interpretable spatial structure, shown here for 0 and 6 ka (Figures 2c and 2d). Moreover, the initial spatial structures were retained from the original data (Figures 2a and 2b), with the only difference being a decrease in fine-scale variability at lower lag distances. Variogram models were successfully fitted, after filtering, for 0 and 6 ka under both climate-continuum and independent-region assumptions (Figures 3a-3h).

The choice of a search strategy that restricts the samples included in the local estimation process is important because it allows local rescaling [Isaaks and Srivastava, 1989]. The search radius was prescribed at 500 km, as determined above, in order to ensure the presence of neighboring data on which to estimate with confidence and also to limit extrapolation in areas where data are absent. This avoids extrapolation, which can make data-model comparison less reliable [Broccoli and Marciniak, 1996]. For all time periods, 3° latitude by 6° longitude grids was produced with oceanic nodes masked.

Gridded surfaces based on ordinary kriging maps show significant changes in the hydrological cycle during the Holocene (Figure 4). For example, Africa has dried from 9 ka to present, as has the southwest United States, while the northern North American plains have become moister. These patterns have been shown in many previous studies [Street-Perrott and Harrison, 1985; Street-Perrott et al., 1989]. More interestingly, the kriged surfaces identify shifting bands of low lake status in Europe and Russia, which are difficult to see in the point maps due to the variability in this region (Figure 4).

Indicator kriging illustrates a high probability of low lake status located over North Africa today but not at 6 ka, as expected (Figure 5). Eastern Europe and Russia show an increased probability of high lake status at 6 ka. After reclassification of the indicator transform the resultant grids illustrate the most probable lake status for the grid points (Figure 6). Main features of the original data (Figure 1) are

![Figure 7. Difference maps between indicator and ordinary kriging estimates, computed as indicator minus ordinary kriging at 0, 3, 6, 9, and 12 ka. The ordinary kriging estimates are plotted in the background and discrepancies are plotted in the foreground.](image-url)
Figure 8. Anomaly maps of lake status at 3, 6, 9, and 12 ka under (a) ordinary kriging estimates and (b) indicator kriging estimates.

...retained, but this gridding enables the identification of lake status in areas with high spatial variability, such as Europe. For example, an area of low lake status is expanded in eastern Europe and Russia at 3 and 9 ka, compared to 0 and 6 ka (Figure 6).

The grids produced under both assumptions were compared in order to evaluate the agreement between both kriging methods (Figure 7). A leeway of ±0.5 units was allocated in order to take into account the continuous characteristic of the ordinary kriging estimates. Discrepancies occur in the proximity of boundaries between regions. This is caused in part by the estimation technique; indicator kriging gives rise to distinct breaks between regions, whereas ordinary kriging yields more continuous estimates. However, the major spatial patterns in lake status are similar.

4.5 Holocene Lake Level Variations and the Global Hydrological Cycle

The Holocene lake status at 3 ka interval is presented as anomaly maps to illustrate the spatial patterns of change in regional water balance (Figure 8). The lake-status anomalies result in five classes, where in the case of indicator kriging +2 means an increase of 2 status classes, +1 an increase of 1 status...
Plate 1. Comparison of gridded lake status maps to general circulation model outputs at 6 ka for (a) ordinary kriging and (b) indicator kriging.
class, etc. In the case of ordinary kriging, the lake status anomalies were rounded to the nearest integer.

At 12 ka the southwestern and eastern United States was wetter than today, while the southern Canadian prairies were drier (Figure 8). Although difficult to evaluate, western Europe and western Russia were also wetter. Some regions of Africa had wetter conditions, as did the Tibetan Plateau. By 9 ka, geographical coverage has increased substantially, and the spatial patterns are much more clearly interpretable (Figure 8). The conditions in southwestern United States were similar to today, while most other regions in North America, except for the Great Lakes area, were drier. These drier conditions extended over much of Europe, western Russia, and Siberia. Another striking feature was the much wetter conditions, which prevailed in North Africa and its eastern coastal region. Although not well represented, South America showed similar conditions to today. Most of North America was drier at 6 ka, while much of Europe and western Russia was similar to today except perhaps in the high northern regions, which is consistent with previous studies [Cheddadi et al., 1997; Qin et al., 1998]. Wetter conditions persisted over much of North Africa, its eastern coast and the Tibetan Plateau. Wetter conditions were also predominant in Australia and western South America. The conditions in North America at 3 ka were similar to those of today while Europe shifted to drier conditions in the midcentral and Mediterranean regions. These drier conditions extended well into western Russia. Although some regions of Africa persisted with wetter conditions, many regions indicated similar conditions to today.

4.6 Data-Model Comparison Results

The CCCcal does not simulate the persistence of low lake-levels over much of North America at 6 ka, except for an area over the Cordilleran region in western Canada and northwestern United States (Plate 1). The CCCfix also does not capture this trend in drier conditions. CCM0 captures the broad-scale patterns in lake status over much of North America, while CCM1 is in agreement with a deficit in moisture over much of North America. Lake status over much of Europe reveals little or no change in status and therefore remains difficult to assess the data-model agreements in all four model outputs used in this study [e.g., Prentice et al., 1998]. The widespread high lake levels over much of North Africa, Arabia, and northern India are not captured by CCCcal, which yields a weak annual P-E (precipitation minus evaporation) signal [Vettoretti et al., 1998]. The CCCfix simulation is in better agreement over North Africa. CCM0 fails to capture the northward extension of the Afro-Asian monsoon belt over North Africa, while CCM1 reveals better agreement over this region, except for the northwestern region as expected from the CCM0 and CCM1 model output [Jolly et al., 1998].

Three models (CCM1, CCM0, and CCCfix) yield similar results under both assumptions with 68%, 66%, and 67% of estimated grid cells in agreement respectively (Figure 9a). Only slightly lower agreement is obtained using the CCCcal simulation with 59% agreement for the continuous assumption and 61% under the indicator approach (Figures 9b and 9c).

5. Discussion

The estimated lake-status anomaly grids under both continuous climate gradient and mutually exclusive lake status category assumptions are generally in agreement with climatic interpretations of previous studies using these datasets [Street-Perrott et al., 1989; Cheddadi et al., 1997; Qin et al., 1998]. At 12 ka the ice sheets were rapidly retreating. Most of North America and Europe were distinctly wetter, except for the southern Canadian prairies, possibly due to a more northerly jet stream (Figure 8). High lake levels were also predominant in many regions of Africa associated with an enhanced Afro-Asian summer monsoon and land-sea temperature contrasts [Kutzbach and Street-Perrott, 1985] (Figure 8). At 9 ka the ice sheet of Europe and Asia had nearly disappeared and sea surface temperatures were close to modern values. In North America the seasonal increase in insolation and the northward migration of the storm tracks were the main factors contributing to the decrease in lake levels over most of the regions [COHMAP Members, 1988; Harrison, 1989]. This decrease was also evident in southern Eurasia resulting from
higher evaporative rates in areas beyond the reach of the enhanced tropical monsoons [Street-Perrott, 1986]. High lake levels were evident over much of northern Africa, northern India, and southern Tibet in agreement with a strengthening of the Afro-Asian monsoons (Figure 8). By 6 ka the strong positive summer insolation anomaly over northern midlatitudes influenced the circulation and caused a widespread prevalence of low lake-levels over much of North America [Street-Perrott, 1986; Webb et al., 1998] and regions of northern Eurasia [Cheddadi et al., 1997]. The high lake levels over much of North Africa, Arabia, and northern India are in good agreement with an enhanced Afro-Asian monsoon circulation [Qin et al., 1998, Street-Perrott et al., 1989]. By 3 ka the seasonal insolation had decreased almost to modern levels as lake status in North America was rising to present levels. A decrease in the westerly transport of moisture caused a fall in lake levels in the Mediterranean regions and interior Eurasia [Street-Perrott et al., 1989]. The weakening Afro-Asian monsoons were responsible for a fall in lake levels over North Africa, Arabia, and southern Asia.

Inspection of the data and the variogram function illustrate high amounts of fine-scale variability at small distances, especially at 3, 6, and 9 ka [Vauv, 1999]. This fine-scale variability was consistent over almost all separation distances covered in this study. The data contain this variability for several reasons including, site sensitivity to climate change, local hydrological variability and problems with site stratigraphy or dating. The use of local indicators of spatial association (LISA) permitted the differentiation of local highly correlated patterns and aided in the identification of local variable sites at specific distances. The LISA results enabled the extraction of locally correlated patterns at a consistent scale of study in order to maximize global coverage. The filtering resulted in the exclusion of sites exhibiting negative spatial autocorrelation at a consistent local scale, yielding more clearly interpretable variograms upon which to estimate the underlying spatial structure.

Although both kriging approaches resulted in almost identical grids, there remain methodological questions concerning the estimation, such as anisotropy, which is the tendency for correlation to favor a particular geographic direction. This was not explored since different anisotropy can coexist at the global scale. Nonparametric validation of the models could also be attempted, such as by stochastic simulations. Estimation of unknown locations using irregularly spaced data remains problematic where some regions have more data than others do and estimations of a more variable region will always be less reliable than regions that are more homogenous.

The methodology illustrated here demonstrates that the lake level data contains large-scale patterns that can be objectively determined at a consistent scale of study. Gridding has the advantage of eliminating finer-scale variability in a data set so that the large-scale pattern can be extracted from local noise. Objectively determining homogenous spatial patterns, whether at the global or at the subcontinental scales, results in better resolution in grid-to-grid comparisons and paleoclimatic reconstruction.

Finally, the methods illustrated in this study can be applied to other paleoclimatic data, such as pollen analysis. The ability to differentiate the climatically controlled spatial patterns from local variability results in better spatiotemporal reconstructions and interpretations using proxy climate data. In the case of lake level variations, the analysis of the spatiotemporal dependence between sites can lead to a better understanding of climate change mechanisms, and hence global hydrological change, as geographical coverage increases in the future.

Acknowledgments. This paper is a contribution to the Climate System History and Dynamics Programme that is jointly sponsored by the Natural Sciences and Engineering Research Council of Canada and the Atmospheric Environment Service of Canada. We thank F. A. Street-Perrott, S. P. Harrison, and others for making the lake level data set available for study. Special thanks to M. C. Sawada for his helpful discussions and computer aid. Results of the lake-level study have been standardized into regular Cartesian grid and can be accessed through the Laboratory for Paleoclimatology and Climatology (LPC) web page.

(http://aix1.uottawa.ca/academic/arts/geographie/lpclweb/)

References

Kutzbach, J. E., Monsoon climate of the early Holocene: Climate experiment with the Earth’s orbital parameters for 9,000 years ago, Science, 214, 59-61, 1981.

K. Gajewski and A. E. Vittou, Laboratory for Paleoclimatology and climatology, Department of Geography, University of Ottawa, 60 University St., Ottawa, Ontario KIN 6N5, Canada. (avisau@uottawa.ca)

(Received December 7, 2000; revised June 22, 2001; accepted July 9, 2001.)