NOTES AND CORRESPONDENCE

The Interpolation of Data Series Using a Constrained Iterating Technique

ALI HARZALLAH

Laboratoire de Météorologie Dynamique du CNRS, Ecole Normale Supérieure, Paris, France

13 May 1994 and 19 January 1995

ABSTRACT

A consistent interpolation technique, applicable to data series, is presented. Demonstrative examples are given where consistency is defined as conservation of mean or of second-order moment; the interpolants are linear or spline functions. The choice of interpolants can be extended to other desired definitions. The method is based on an iterating technique that forces the interpolant to satisfy the consistency constraint. The method works for an evenly spaced series and does not suppose periodicity. The method is applied to interpolate monthly series of sea surface temperature from yearly sampled values; it is shown that the method provides satisfactory results.

1. Introduction

Over the past few years, atmospheric simulations using general circulation models (GCMs) have shown an important improvement; the length of simulations has increased from days or months to decades and centuries. This improvement has, in particular, raised the problem of defining a consistent set of surface conditions to be used as bounds for those models. For example, daily values of sea surface temperature (SST) are generally used to force atmospheric GCMs. However, only monthly fields are usually available and one has to interpolate the data. The consistency can be defined to mean no loss of initial information nor any creation of contradictory information. Thus, it can be defined as the conservation of mean; in such a case, data values can be reconstructed exactly from the interpolant by simple averaging. It might also be appropriate to define consistency as the conservation of second- or higher-order moments, depending on the nature of the problem. For example, conserving T^4 where T refers to SST is equivalent to conserving a quantity proportional to the longwave radiation emission (this last quantity being $\epsilon\sigma T^4$, where ϵ is the emissivity and σ is the Stefan–Boltzman constant).

A commonly used method that conserves the mean is Fourier-based interpolation, illustrated by Epstein (1991). This method is simple in concept and has the important property that no additional frequencies are added to the data. The interpolation is carried out by adding zeros to spectral coefficients of data and by performing an inverse Fourier transform to the resulting extended coefficients. The method is equivalent to applying a rectangular filter to Fourier coefficients. In this method, data are considered periodic; it is therefore only applicable to climatological data as shown in Epstein (1991).

Another simpler method widely used in meteorology is linear interpolation between successive data values. This method, as discussed in Epstein, has the disadvantage of underestimating the total variation of data and not conserving the mean.

We present here an interpolation method that can be applied to periodic as well as to nonperiodic data. We show that for periodic data this method provides similar results to those obtained by the Fourier-based method. The method is presented in the next section and illustrated by some examples in section 3. A brief conclusion is presented in section 4.

2. The method

a. Derivation

Let us consider M samples S_m, $m = 1, \cdots, M$ of time series corresponding, say, to monthly mean values of SST. We wish to obtain daily values s_j, where j refers to the jth day in the series, by interpolation of the given series S_m. The series s_j must satisfy the consistency property, defined in the present example as conservation of the mean

$$\overline{s_j} = S_m,$$

where the overbar refers to the average over all days of month m. The first step is to choose an interpolation

Corresponding author address: Dr. Ali Harzallah, Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, Cedex 05, France.

© 1995 American Meteorological Society

Unauthenticated | Downloaded 02/08/23 09:04 PM UTC

function that coincides with S at the known values S_m. Linear or cubic splines are examples of such functions. Defining $S_m^{(0)} \equiv S_m$ and $s_j^{(0)}$ as the interpolant of S_m, the second step consists in recalculating, for each month m, the monthly average $\overline{s_{jm}^{(0)}}$. The monthly average $\overline{s_{jm}^{(0)}}$ will not exactly coincide with $S_m^{(0)}$ since most interpolation functions are not mean-conservative interpolants. The next step is to force mean conservation by calculating the difference

$$S_m^{(1)} = S_m^{(0)} - \overline{s_{jm}^{(0)}}.$$

Then $S_m^{(1)}$ is interpolated using the same interpolation function to obtain $s_j^{(1)}$, and the monthly average $\overline{s_{jm}^{(1)}}$ is recalculated. The procedure is continued in the same way until $\|S_m^{(n)}\|$ is less than a given level of precision. The iterating expression is thus given by

$$S_m^{(i)} = S_m^{(i-1)} - \overline{s_{jm}^{(i-1)}}.$$

The mean-conservative interpolant s_j is obtained by the sum

$$s_j = \sum_{i=1,\cdots,n} s_j^{(i-1)}.$$

One can see that $\overline{s_{jm}} = S_m^{(0)} - O_m^{(n)}$, where $O_m^{(n)}$ $[=S_m^{(n)}]$ is the iterating error depending on the desired precision. Thus, the conservation property is obtained for all interpolation functions that satisfy

$$\|S_m^{(i)}\| < \|S_m^{(i-1)}\|.$$

Let us now consider the case where we wish the consistency property to be conservation of second-order moment (no conservation of the mean is required)

$$\overline{s_{jm}^2} = V_m,$$

where V_m are known values. Simple calculations show that the iterating expression is given by

$$S_m^{(i)} = -\sum_{k=1,\cdots,i} \overline{s_{jm}^{(k-1)}} + \beta\{V_m - \overline{[\sum_{k=1,\cdots,i} s_j^{(k-1)}]_m^2}\}^{1/2},$$

with

$$\beta = \mathrm{sgn}[\sum_{k=1,\cdots,i} \overline{s_{jm}^{(k-1)}}].$$

b. Linear and cubic-spline interpolations

To illustrate the convergence of the procedure, let us now consider the cubic-spline function defined for the interval $[m, m+1]$ by

$$s_j = AS_m + BS_{m+1} + CS_m'' + DS_{m+1}'',$$

where

$$A = \frac{x_{m+1} - x}{\Delta x_{m^+}}, \quad B = 1 - A,$$

$$C = \frac{(A^3 - A)\Delta x_{m^+}^2}{6}, \quad D = \frac{(B^3 - B)\Delta x_{m^+}^2}{6},$$

with x referring to the abscissa axis (time), $\Delta x_{m^+} = x_{m+1} - x_m$ ($\Delta x_{m^-} = x_m - x_{m-1}$), and x_m and S_m'' are the abscissa and the second derivative at month m. If we choose S_1'' and S_M'' equal to zero, we obtain the so-called *natural cubic spline.*

The average of the function s over month m is given by

$$\overline{s}_m = \frac{1}{8} S_{m-1} + \frac{3}{4} S_m + \frac{1}{8} S_{m+1} - \frac{1}{384} [7\Delta x_{m^-}^2 S_{m-1}'' + 9(\Delta x_{m^-}^2 + \Delta x_{m^+}^2)S_m'' + 7\Delta x_{m^+}^2 S_{m+1}'']. \quad (1)$$

In the linear interpolant case ($C = D = 0.$),

$$\sum_m (\overline{s}_m - S_m)^2 = \sum_m \left(\frac{1}{8} S_{m-1} - \frac{1}{4} S_m + \frac{1}{8} S_{m+1}\right)^2.$$

Using the fact that

$$\sum_m S_m S_n \leqslant \sum_m S_m^2,$$

we obtain

$$\sum_m (\overline{s}_m - S_m)^2 \leqslant \frac{1}{4} \sum_m S_m^2.$$

This shows that the iterating technique converges in the case of linear interpolation. Note that the convergence does not depend on the length of the intervals.

In the case of cubic splines, similar calculations show that

$$\sum_m (\overline{s}_m - S_m)^2 \leqslant \frac{1}{4} \sum_m (S_m)^2 + \frac{431}{192} (\Delta x_{\max})^4 \sum_m (S_m'')^2,$$

where $\Delta x_{\max}$ is the largest interval in the series. A sufficient condition of convergence is therefore given by

$$(\Delta x_{\max})^4 \sum_m (S_m'')^2 \leqslant \frac{144}{431} \sum_m S_m^2.$$

This condition shows that the convergence depends on the choice of the second derivative of the spline. It also depends on the length of the largest interval but not on how intervals are given.

In the case where we wish to conserve the second-order moment, we specifically chose β in Eq. (1) so that $\|S_m^{(i)}\|$ decreases. In linear interpolation, it is easily seen that the procedure always converges, and $\|S_m^{(i)}\| < \|S_m^{(i-1)}\|$ is always verified. This results from the fact that a linear interpolation leads to series with variance always less than the variance of the initial series. For other functions, the procedure converges if the interpolant satisfies this last condition.

The mean-conservative interpolation can be obtained directly in the linear case (see also de Boor 1978) and in the cubic spline case by solving the matrix formed

Unauthenticated | Downloaded 02/08/23 09:04 PM UTC

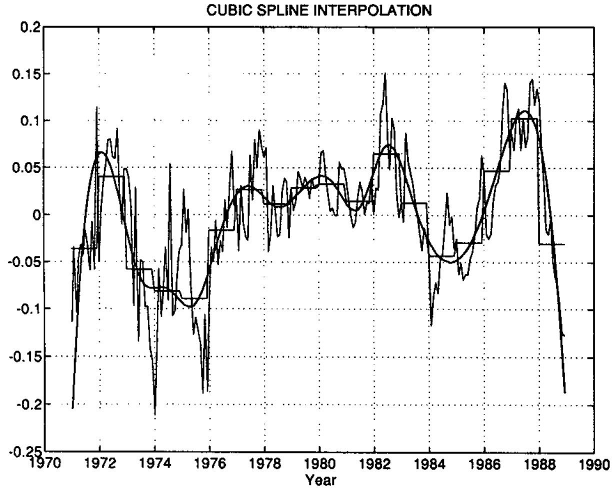

FIG. 1. Cubic-spline interpolation of yearly averaged sea surface temperature anomalies at Niño-3. The original monthly mean values are shown together with the annual average and monthly values obtained using the iterating technique with mean-conservation constraint. Abscissas are years (1970–88) and ordinates are degrees Celsius.

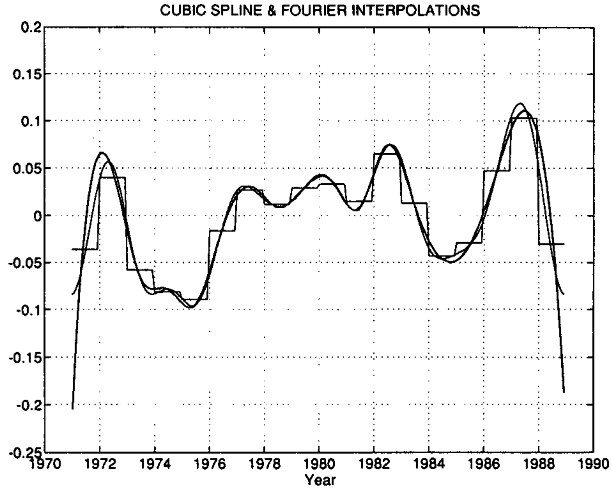

FIG. 2. Cubic-spline and Fourier-based interpolations of yearly averaged sea surface temperature anomalies at Niño-3. The yearly averaged values are shown together with monthly values obtained by the iterating technique (thick line) and by the Fourier-based method (thin line), both with the constraint of mean-conservation. Abscissas are years (1970–88) and ordinates are degrees Celsius.

by $M - 2$ Eq. (1) for $m = 2, \cdots, M - 1$ and two additional conditions for the left and right bounds.

In the case where the second-order moment is conserved, a linear interpolation leads to

$$\overline{s_m^2} = \frac{1}{24} S_{m-1}^2 + \frac{7}{12} S_m^2 + \frac{1}{24} S_{m+1}^2 + \frac{1}{6} S_m S_{m-1} + \frac{1}{6} S_m S_{m+1} .$$

This expression is nonlinear and therefore the set of resulting equations cannot be solved directly by matrix inversion.

3. Examples

The examples shown here are based on a series of SST averaged over the region extending from 5°S to 5°N and from 150° to 90°W, the so-called Niño-3 region. The original data are monthly means and cover the period 1970–88. The mean annual cycle was first removed to retain only anomalies. In the examples, we have averaged the data over each year. Thus, we obtain 19 samples corresponding to the 19-yr period. We use the method proposed here to obtain monthly mean values from the yearly averaged ones. We can therefore compare the original and interpolated series. Figure 1 shows the original monthly values of SST together with the yearly averaged ones and the results of a cubic-spline interpolation. The cubic spline used is the CSINT interpolation procedure of the IMSL library. Endpoint conditions require that the third derivative of the spline be continuous at the second and next-to-last points (see de Boor 1978). Figure 1 shows that the method interpolates efficiently the yearly averaged series. The method converged after six iterations with an absolute error of 10^{-3}. One should notice that abscissas of maxima of the interpolated series do not always correspond to those of the yearly averaged series. This is inherent to the constraint of mean conservation; the interpolant should sometimes be above the mean and at other times below, so that its average can equal the mean.

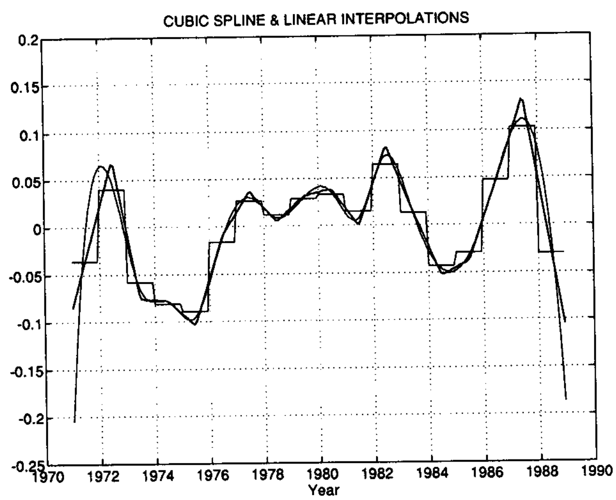

FIG. 3. Cubic-spline and linear interpolations of yearly averaged sea surface temperature anomalies at Niño-3. The yearly averaged values are shown together with monthly values obtained by the iterating technique using cubic (thin line) and linear (thick line) interpolants with the constraint of mean conservation. Abscissas are years (1970–88) and ordinates are degrees Celsius.

Unauthenticated | Downloaded 02/08/23 09:04 PM UTC

We have plotted in Fig. 2 the yearly averaged values with the interpolation results using the cubic spline as in Fig. 1 and a Fourier-based procedure analog to that used by Epstein (1991). One can see that the two methods give sensibly the same results. The largest differences lie at the left and right bounds due to the fact that the Fourier-based method supposes periodicity, whereas the present cubic spline imposes a continuous third derivative at endpoints. If periodicity is needed, one can simply apply the procedure to data that are extended on the left and the right by repeating data values. One can also directly use spline functions that satisfy periodicity (see the IMSL library).

Results of the iterating technique using a linear interpolation are shown in Fig. 3, together with results using the cubic-spline function. One can see that the results of the linear interpolation series are very close to those of the cubic-spline interpolation. It should be noted that a linear interpolation using mean-conservation underestimates the total variation less than simple linear interpolation.

In the last example, we compare results of the cubic-spline interpolation where the mean and the second-order moment are conserved. This last quantity is calculated as the sum over each year of squares of the original monthly mean values. Results, presented in Fig. 4, show that the total variation is increased when using conservation of the second-order moment. One can see, for example, that extrema during 1974–76 are better reproduced using the second-order moment (on the other hand, the values themselves are overestimated). This is of importance in some applications. For example, the variability of SST used to force atmospheric GCMs greatly influences the variability of the simulated fields. The use of a second-order moment interpolation will therefore provide a better estimation of the variability.

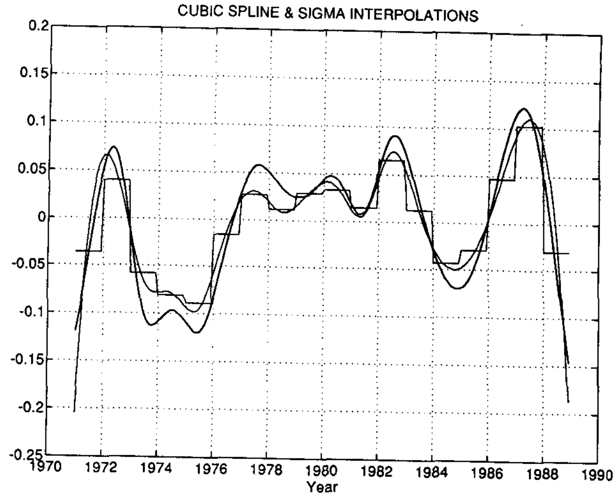

FIG. 4. Cubic-spline interpolations of yearly averaged sea surface temperature anomalies at Niño-3. The yearly averaged values are shown together with monthly values obtained by the iterating technique using mean (thin line) and second-order moment (thick line) conservation. Abscissas are years (1970–88) and ordinates are degrees Celsius.

4. Conclusions

The examples shown here have demonstrated that spline interpolants used with the mean or second-order moment conservative procedures constitute a powerful method to interpolate data series in a consistent way. It was shown that the method produces similar results to the Fourier-based technique. The interpolation conserving the second-order moment is shown to reduce the underestimation of the total variation. The main advantages of the method is that it can be used with a large number of interpolant functions with different consistency properties depending on the physics behind the interpolation.

Acknowledgments. The author is grateful to Dr. Laurent Fairhead for his help in reading the manuscript.

REFERENCES

de Boor, C., 1978: *A Practical Guide to Splines.* Springer Verlag.
Epstein, E. S., 1991: On obtaining daily climatological values from monthly means. *J. Climate,* **4,** 365–368.

Unauthenticated | Downloaded 02/08/23 09:04 PM UTC