Optimal Average of Regional Temperature with Sampling Error Estimation

Samuel S. Shen
Center for Climate System Research
University of Tokyo
Tokyo 153, Japan
and
Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1

and
Xiaochun Wang
Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta T6G 2G1

[Original manuscript received 28 May 1995; in revised form 5 November 1996]

ABSTRACT *This paper presents an optimal averaging scheme that yields the area average of a climate field over a region. The scheme consists of two steps: fitting the observation data to a priori covariance functions and computing the weights for each of the observation stations. A surface air temperature dataset of 23 stations in Northeast China from 1961 to 1990 is used to demonstrate the use of the scheme and its advantages. The results show that, compared to the arithmetic averaging scheme, the optimal averaging scheme has a smaller sampling error and less deviation from the mean value of the random sampling error, in particular, when there are only a small number of stations.*

RÉSUMÉ *Cet article expose un schème pour obtenir une moyenne optimale donnant, au-dessus d'une région, une moyenne spatiale d'un champ climatique. Le schème comprend deux étapes: l'agencement des données d'observation aux fonctions de covariance et le calcul des coefficients de pondération à chacune des stations d'observation. Pour démontrer le fonctionnement du schème et de ses avantages, on a eu recourt à un ensemble de données de températures de l'air en surface, comprenant 23 stations du nord-est de la Chine au cours de la période de 1961 à 1990. Pour un petit nombre de stations en particulier, les résultats montrent, lorsqu'on le compare au schème de la moyenne arithmétique, que le schème pour obte-*

ATMOSPHERE-OCEAN 35 (2) 1997, 147–160 0705-5900/97/0000-0147$1.25/0

nir une moyenne optimale présente une plus petite erreur d'échantillonnage et moins d'écart que la valeur moyenne de l'erreur d'un échantillonnage pris au hasard.

1 Introduction

The purpose of this paper is to present an optimal averaging scheme and show that it is worthwhile to consider using an optimization method for spatial averages, in particular when observation stations are sparse.

Properly weighting the data from each observation station in order to obtain the best average of an observed climate field over a region of interest is an important task in climatological statistics. It can happen that the poor result of averaging is due to an improper weighting scheme rather than an insufficient number of samples, especially when the field is highly inhomogeneous. Taking the global average annual mean temperature as an example, it used to be thought that at least several hundred stations were needed to get a reasonable result. Shortly after the famous publications of Jones et al. (1986a, b), Hansen and Lebedeff (1987) and Vinnikov et al. (1990), there appeared a question regarding the accuracy of their analyses (Madden et al., 1993; Weber and Madden, 1995). There were good reasons to ask this question since each paper used a different number of stations (1783 in Jones et al., 2685 in Hansen and Lebedeff and 566 in Vinnikov et al.) and a different weighting scheme to calculate the global average. After the publication of Jones (1994) and Shen et al. (1994), it became clear that the annual mean surface air temperature field has relatively few degrees of freedom, most likely between 24 and 100, and the global average can be computed rather accurately with 60 or so well-distributed stations on the globe if a proper weight is assigned to each station. (For the January temperature, Madden et al. (1993) estimated 135 degrees of freedom using model output.) Therefore, if there are many redundant stations, one can get a good average without optimizing the weights. This explains why the results of Jones et al. (1986a, b), Hansen and Lebedeff (1987) and Vinnikov et al. (1990) are highly correlated since all of them have many redundant stations.

Despite the above statement, for the dataset considered in the present paper, we will show that the sampling error is noticeably smaller when the optimal weight scheme is used and when one has fewer stations. This makes the optimal averaging scheme useful when a very high accuracy of a spatial average is required, such as the case of ground truth validation of satellite images (North et al., 1994). We will illustrate the significant differences by Monte Carlo experiments (Section 4).

The spectral method in Shen et al. (1994) is basically a double projection: the temperature field was projected onto the Empirical Orthogonal Function (EOF) basis and the EOFs were further projected onto spherical harmonics. The present study tries to provide an optimal averaging scheme when an orthogonal basis, such as the spherical harmonics, cannot easily be found. This is often the case when one considers a regional average.

The context of this paper is still in the category of classical objective analysis

(Gandin, 1965; Daley, 1991) and does not intend to give an overall picture of all the averaging methods available and their results. Hence, we have made selective citations of the past work on averaging the surface temperature field, no matter regional or global, just to demonstrate the motivation for our current work. As far as the methodology is concerned, our method is similar to that of kriging averaging (Rendu, 1981). The minor difference is that we have included a least square data fitting step to obtain the parameters in the assumed covariance functions. This step is similar to that used in Reynolds and Smith (1994) for optimal interpolation. Our fitted covariance functions are isotropic around each point but not homogeneous over the region in question (see Section 3). Hence our method is also different from the optimization method of Vinnikov et al. (1990) in which a zonally homogeneous and isotropic covariance structure is assumed (also see Section 3).

In contrast to a brief mathematical analysis described elsewhere (Shen et al., 1995), the present research is focused more on the sampling error analysis and the comparison between the optimal and arithmetic averaging schemes with respect to the sampling error. We attempt to present our method in such a way that readers can easily code their own program and perform computations. The purpose of presenting our results in such a way is to let researchers who work on regional climate change consider the optimal averaging scheme as an option for computation in the situation of sparse observation stations.

In Section 2, the optimal averaging scheme is derived. Section 3 describes the dataset, the averaging of a covariance function around a station, and the result of the average temperature in Northeast China using 23 stations. The Monte Carlo experiments designed to compare the sampling errors from the optimal averaging scheme and the arithmetic averaging scheme are presented in Section 4. Section 5 contains conclusions and discussions.

2 Theory of optimal averaging scheme

The theory of the optimal averaging scheme is presented in terms of monthly average temperature anomaly after the removal of the seasonal cycle and standardization (details in Section 3). Let $\mathbf{r}$ be the position of a point in the region Ω and $\Theta(\mathbf{r}, t)$ the monthly average temperature at point $\mathbf{r}$ and month t. The true regional average of the temperature anomaly over Ω is

$$\bar{\Theta}(t) = \frac{1}{A} \int_{\Omega} d\Omega \, \Theta(\mathbf{r}, t) \tag{1}$$

where A is the area of the region Ω, and $d\Omega$ is the integration element.

The estimate $\bar{\Theta}(t)$ using observation data from N stations is

$$\hat{\bar{\Theta}}(t) = \sum_{i=1}^{N} w_i \Theta(\mathbf{r_i}, t), \tag{2}$$

where w_i is the weight for the ith station ($i = 1, 2, \cdots, N$). If $w_1 = w_2 =$

$\cdots = w_N = 1/N$, then $\hat{\bar{\Theta}}$ is the arithmetic average. If we choose the weights $w_1, w_2, \cdots, w_N$ such that they minimize the mean square error (MSE) of the estimator $\hat{\bar{\Theta}}(t)$ for $\bar{\Theta}$ defined by

$$\varepsilon^2 = \langle (\bar{\Theta}(t) - \hat{\bar{\Theta}}(t))^2 \rangle, \tag{3}$$

then $w_1, w_2, \cdots, w_N$ are called the optimal weights and the corresponding $\hat{\bar{\Theta}}$ is called the optimal average. In the above, $\langle \cdots \rangle$ stands for the ensemble average. We regard the time-series of $\bar{\Theta}(t)$ and $\hat{\bar{\Theta}}(t)$ as approximately stationary. Then ε^2 is independent of time. Here we have a dilemma: there are some trends in the time-series $\bar{\Theta}(t)$ and $\hat{\bar{\Theta}}(t)$, (hence the time-series of $\bar{\Theta}(t)$ and $\hat{\bar{\Theta}}(t)$ are not stationary), but at this stage of research development we do not have another method which abandons the stationarity assumption. The compromise is that the amplitude of the trend is small compared with the standard deviation of the same time-series in a time interval of five or ten years. In the present study, the ratio of the former to the latter is less than 1/7 (see the bottom panel of Fig. 3) and the trend is regarded as being small. If the trend is too large to be regarded as a "small" trend, one can carry out a detrending procedure as follows: use the stationarity assumption to process the average to get the trend and subtract this trend from the original time-series. The resulting time-series should be closer to a stationary one.

However, we need to detect the trend correctly even if it may be small. We should choose our weights so as not to deform the shape of the trend. Therefore, a normalization condition on the weights is imposed:

$$\sum_{i=1}^{N} w_i = 1. \tag{4}$$

The expansion of the MSE formula leads to

$$\varepsilon^2 = \frac{1}{A^2} \int_\Omega d\Omega \int_\Omega d\Omega' \, \rho(\mathbf{r}, \mathbf{r}') - \frac{2}{A} \sum_{i=1}^{N} w_i \int_\Omega d\Omega \, \rho(\mathbf{r}, \mathbf{r_i}) + \sum_{i,j=1}^{N} w_i w_j \, \rho(\mathbf{r_j}, \mathbf{r_i}) \tag{5}$$

where

$$\rho(\mathbf{r_j}, \mathbf{r_i}) = \langle \Theta(\mathbf{r_i}, t)\Theta(\mathbf{r_j}, t) \rangle \tag{6}$$

is the covariance matrix. Here again since it is assumed that $\Theta(\mathbf{r_j}, t)$ is a stationary time-series, $\rho(\mathbf{r_j}, \mathbf{r_i})$ is independent of t.

We try to find the weight for every station by minimizing ε^2 with (4) as a constraint. The weights obtained by this procedure are optimal. The method of Lagrange multiplier is used for the minimization. The Lagrange functional is

$$J[w_1, \cdots, w_N] = \varepsilon^2[w_1, \cdots, w_N] - 2\Lambda\left(\sum_{i=1}^{N} w_i - 1\right) \tag{7}$$

where -2Λ is the Lagrange multiplier. At the critical point of the Lagrange functional, we have

$$\frac{\partial J}{\partial w_i} = 0, \quad i = 1, \cdots, N \quad \text{and} \quad \frac{\partial J}{\partial \Lambda} = 0. \tag{8}$$

The above yields

$$\sum_{j=1}^{N} w_j \rho(\mathbf{r_i}, \mathbf{r_j}) - \Lambda = \bar{\rho}(\mathbf{r_i}), \qquad i = 1, \cdots, N, \tag{9}$$

$$\sum_{i=1}^{N} w_i = 1, \tag{10}$$

where

$$\bar{\rho}(\mathbf{r_i}) = \frac{1}{A} \int_{\Omega} d\Omega\, \rho(\mathbf{r}, \mathbf{r_i}). \tag{11}$$

The solution of (9)–(10) yields the optimal weights $w_1, \cdots, w_N$. The above equations look similar to a kriging system (Rendu, 1981). But we will compute $\bar{\rho}(\mathbf{r_i})$ in a rather different way from the conventional kriging method.

Let $[b_{ij}]$ denote the inverse of the matrix $[\rho_{ij}] = [\rho(\mathbf{r_i}, \mathbf{r_j})]$. Then the solution of (9)–(10) can be expressed by

$$w_i = \sum_{j=1}^{N} b_{ij}[\Lambda + \bar{\rho}(\mathbf{r_j})], \qquad i = 1, 2, \cdots, N, \tag{12}$$

$$\Lambda = \frac{1 - \sum_{i,j=1}^{N} b_{ij} \bar{\rho}(\mathbf{r_j})}{\sum_{i,j=1}^{N} b_{ij}}. \tag{13}$$

Now $\bar{\rho}(\mathbf{r_i})$ is a new regional average. It signifies the importance of station i when its observation data are used to compute the regional average, since $\bar{\rho}(\mathbf{r_i}) = \langle \bar{\Theta}(t)\Theta(\mathbf{r_i}, t) \rangle$ is the correlation between the data of station i and the regional average temperature. What we originally wanted to compute is the regional average temperature over Ω. Now we have to compute $\bar{\rho}(\mathbf{r_i})$ before we can solve the equations (9)–(10) for the optimal weights. Therefore, the original averaging problem has been converted into a new averaging problem. Fortunately, the latter average can be estimated more accurately since the covariance often has a pattern.

3 Computational procedures and results

The different datasets support different covariance patterns. Vinnikov et al. (1990) presumed an empirical correlation coefficient formula for every latitude band of 30° degrees. For 30°–60°N, the presumed formula is

$$\rho(s) = \exp(-0.21s^{0.893})J_0(0.852s) \tag{14}$$

where $\rho(s)$ is the correlation coefficient, s is the distance between stations with 10^3 km as the unit, and J_0 is the zeroth order Bessel function. However, one may not always want to start with a fancy pattern of this sort and often a simple normal distribution, with a couple of tuneable free parameters, can be a good candidate for the first try. Our assumed covariance structure is isotropic around a station but different from one station to the other (and hence nonhomogeneous). Around the ith station it is assumed that

$$\hat{\rho}(\mathbf{r}, \mathbf{r_i}) = a_i \exp\left(-\frac{|\mathbf{r} - \mathbf{r_i}|^2}{d_i^2}\right). \tag{15}$$

Then we use the observation data to fit this pattern to determine a_i and d_i. Hence we have a covariance pattern for each station i and this pattern is fitted with N covariance data $\rho(\mathbf{r_i}, \mathbf{r_j}), j = 1,2,\cdots,N$. This is different from the conventional objective analysis that assumes a single covariance function which will be fitted by $N \times (N-1)/2 + N$ covariance data $\rho(\mathbf{r_i}, \mathbf{r_j})$, $i,j = 1,2,\cdots,N$ and $j \geqslant i$.

We wish to point out that since the data usually cannot fit the pattern (15) when $\mathbf{r} - \mathbf{r_i} = 0$, a_i is thus not equal to unity even for standardized data. In this sense, a_i is not the point variance of Θ at $\mathbf{r_i}$ and d_i is not the e-folding length scale of the Θ field. For the best fitting, we usually have $a_i < 1$ and d_i is larger than the spatial length scale which is about 800 km for an inland area such as Northeast China.

a *Data*

The data used in our computation experiment consist of monthly average surface air temperatures from 23 stations in Northeast China (Fig. 1). The time period is from 1961 to 1990. Part of the data (16 stations) were compiled in Tao et al. (1991), in which the station history, observation instruments and quality control procedures are described. In order to have as many station observations as possible in this region, we have included 7 stations from another dataset (Wang and Xiao, 1991). According to Tao et al. (1991), before the 1950s there were different regulations among the stations with regard to the observation time, method and the way to calculate monthly average (since the People's Republic of China was established in October 1949). After that time, the regulations for all stations became consistent. For example, the daily temperature was calculated as the average of 4-time observations. The monthly temperature is calculated as the average of daily temperatures. Figure 1 depicts the location of the 23 stations and the 104 grid points on the $1° \times 1°$ grid in the region.

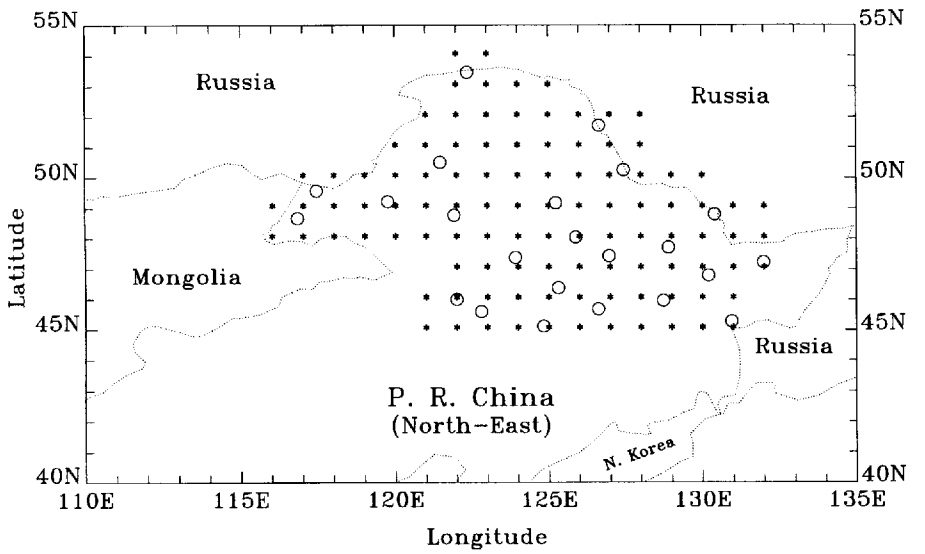

Fig. 1 The locations of 23 stations (circles) from Northeast China and the 104 grid points (stars) regularly distributed (1° × 1°) in the region.

The seasonal cycle of the data of a station is the 30-year average of this station data for each month and it thus has 12 values for each station. The temperature anomaly data of a station is obtained by subtracting, from each monthly value at the station, the seasonal cycle of the station for that month. Then the temperature anomaly time-series of 360 months is standardized by being divided by its standard deviation. This is done for every station. To get an idea of the magnitude of the anomalies in terms of °C, we need to include the average standard deviation of the 23 stations' monthly temperature for every month: 2.368°C for January, 2.363°C for February, 2.450°C for March, 1.770°C for April, 1.331°C for May, 1.337°C for June, 0.987°C for July, 1.091°C for August, 1.014°C for September, 1.367°C for October, 2.674°C for November, and 2.570°C for December. The average standard deviation of the 23 stations' annual mean temperature is 0.830°C.

b *Fitting data to the covariance pattern*

The minimization objective function for data fitting is

$$E = \sum_{j=1}^{N} \left[\ln \left(\frac{\hat{\rho}(\mathbf{r_j}, \mathbf{r_i})}{\rho(\mathbf{r_j}, \mathbf{r_i})} \right) \right]^2 . \tag{16}$$

The ideal case is that $\rho(\mathbf{r_j}, \mathbf{r_i}) = \hat{\rho}(\mathbf{r_j}, \mathbf{r_i})$ for every j. Then $E = 0$ is the minimum, otherwise $E > 0$.

The minimization condition is

$$\frac{\partial E}{\partial a_i} = 0 \quad \text{and} \quad \frac{\partial E}{\partial d_i^2} = 0. \tag{17}$$

Let $\beta_{ij} = \ln\rho(\mathbf{r_i}, \mathbf{r_j})$ and $s_{ij} = |\mathbf{r_i} - \mathbf{r_j}|$. From the above minimization condition, we can have

$$d_i^2 = \frac{N\sum_{j=1}^{N} s_{ij}^4 - (\sum_{j=1}^{N} s_{ij}^2)^2}{\sum_{j=1}^{N}\beta_{ij}\sum_{j=1}^{N} s_{ij}^2 - N\sum_{j=1}^{N}\beta_{ij}s_{ij}^2}, \tag{18}$$

$$a_i = \exp\left[\left(d_i^2\sum_{j=1}^{N}\beta_{ij} + \sum_{j=1}^{N} s_{ij}^2\right)/(Nd_i^2)\right]. \tag{19}$$

For the Northeast China dataset, the averages of a_i and d_i are

$$\frac{1}{23}\sum_{i=1}^{23} a_i = 0.790 \text{ (dimensionless)}, \quad \frac{1}{23}\sum_{i=1}^{23} d_i = 1.648 \times 10^3 \text{ km}$$

The standard deviations of a_i and d_i are 0.0847 and 507 km respectively. Hence the estimations of a_i and d_i are reasonably robust. Figure 2 shows the robustness and goodness of the fit of the covariance pattern (15) with the above averaged parameters a and d. The vertical coordinates of the $23 \times 22/2 + 23$ points are the entries of the covariance matrix. The horizontal coordinates are the distance between each pair of points.

We point out again that a_i are not the point variances and d_i are not e-folding spatial correlation length scales. (The average e-folding spatial length scale for the same region is 739 km.)

c *Computation of* $\bar{\rho}(\mathbf{r_i})$

To compute $\bar{\rho}(\mathbf{r_i})$ we put a dense uniform longitude-latitude grid network on the region Ω. Then $\bar{\rho}(\mathbf{r_i})$ can be computed by

$$\bar{\rho}(\mathbf{r_i}) = \frac{1}{N_{grid}}\sum_{g=1}^{N_{grid}} a_i \exp\left(-\frac{|\mathbf{r_g} - \mathbf{r_i}|^2}{d_i^2}\right) \tag{20}$$

in which N_{grid} is the total number of grid points in region Ω and $\mathbf{r_g}$ is the position for gth grid point in Ω. Here we are designing a scheme to compute the regional average and the area in each grid box is considered the same. Of course if the north-south span of the region is very large, one has to include the area factor (i.e. the cosine of the latitude) into the above numerical integration formula. In our computation, the

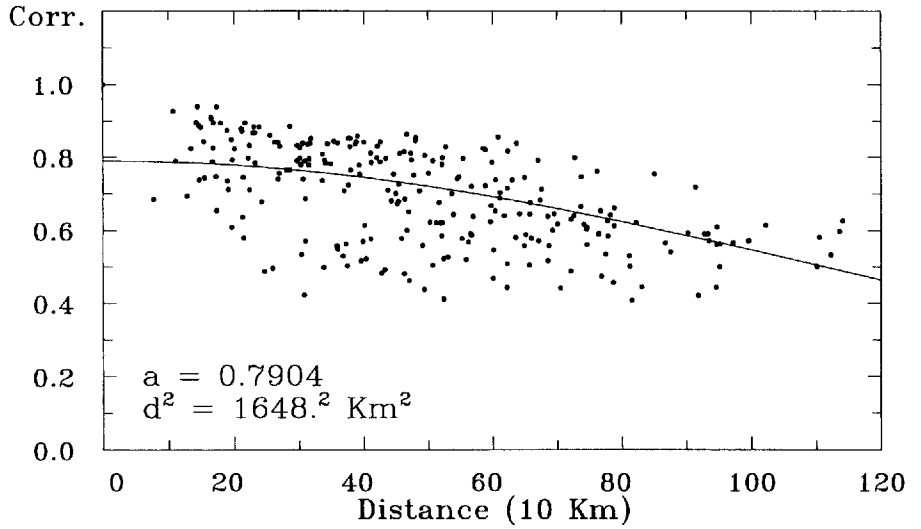

Fig. 2 The isotropic covariance for monthly temperature is shown as a function of distance. The solid curve shows the results of formula (15), with the parameters a_j and d_j averaged over 23 stations: $d = \frac{1}{23}\sum_{j=1}^{23} d_j = 1648$ km and $a = \frac{1}{23}\sum_{j=1}^{23} a_j = 0.7904$.

dense grid is a $1° \times 1°$ network and there are 104 grid points which are in or near the region Ω.

For comparison, the top panel of Fig. 3 shows the difference of the regional average monthly temperature computed by the optimal and arithmetic averaging schemes. The difference between the results yielded from the two methods is of noticeable size. Please note that the units of the vertical scale are °C which resulted from multiplying the dimensionless $\hat{\Theta}$ by the standard deviation for each month. The standard deviations for monthly data are roughly in the range of 1.0–2.7°C as described in Section 3a.

The bottom panel of Fig. 3 shows the times-series of the optimally averaged monthly temperature (solid line). The dashed line is the 13-point moving average. The dotted line is the linear trend obtained by the least square fitting of the solid line. The formula of the trend is

$$-0.51900 + 0.00287 \times t$$

where the unit of t is month. The net increase of the temperature from 1960 to 1991 according to this trend is: $0.00287 \times 360 = 1.0$°C. This increase is larger than that of the global or Northern Hemispheric average surface air temperature since our dataset is for a middle latitude land region.

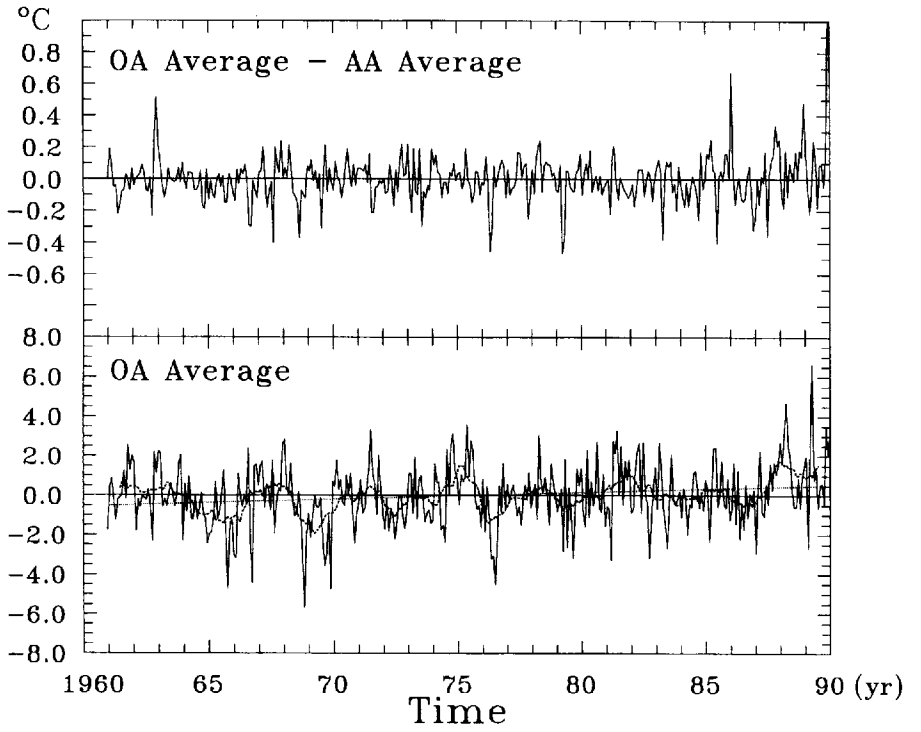

Fig. 3 The top panel shows the difference between the optimal average (OA) and the arithmetic average (AA) of the monthly temperature. In the bottom panel, the solid line is the optimally averaged monthly temperature. The dashed line is the 13-point moving average. The dotted line is the linear trend obtained by the least square fitting of the solid line.

4 Monte Carlo experiments

A Monte Carlo method is used to compare the sampling errors resulting from the optimal and arithmetic averaging schemes. We performed a sequence of experiments with the number of stations included to compute the regional average decreasing gradually, from 19, 17, ⋯, until 3. The sample size is 1000. For example, when we experiment with 15 stations, we randomly draw 15 stations out of the 23 stations and compute the regional average using the optimal averaging scheme and the arithmetic scheme respectively. We define a percentage sampling error (PSE) to show the accuracy of a regional average method.

We first compute the regional average of the temperature anomaly using all the 23 stations by the optimal averaging scheme and the arithmetic averaging scheme respectively. The average of the results from these two schemes is used as the standard value, denoted by STD_m for the mth month. Namely,

STD_m = [optimal average of 23 stations + arithmetic average of 23 stations]/2. (21)

We will check how far the averages, using the data from fewer stations, deviate from this standard value. This deviation is represented by a root mean square error:

$$RMSE = \left[\frac{1}{M}\sum_{m=1}^{M}(X_m(p) - STD_m)^2\right]^{\frac{1}{2}}. \tag{22}$$

Here M is the length of the data in months which is $12 \times 30 = 360$, $X_m(p)$ is the regional average temperature for the mth month computed using p randomly drawn stations (from the 23 stations) by either the optimal averaging scheme or the arithmetic averaging scheme.

Now we can define the percentage sampling error

$$PSE = RMSE/DEV \times 100\% \tag{23}$$

in which DEV is the standard deviation of STD_m.

We would like to check the reliability of the sampling error formula (5). The PSE computed by formula (5) is denoted by PSE_f and is given by

$$PSE_f = \varepsilon \times 100\%. \tag{24}$$

The three terms in (5) are computed in the following way

$$\frac{1}{A^2}\int_\Omega d\Omega \int_\Omega d\Omega' \, \rho(\mathbf{r}, \mathbf{r}') \approx \frac{1}{N \times N_g}\sum_{i=1}^{N}\sum_{g=1}^{N_g} a_i \exp\left(-\frac{|\mathbf{r_g} - \mathbf{r_i}|^2}{d_i^2}\right), \tag{25}$$

$$\frac{2}{A}\sum_{i=1}^{N} w_i \int_\Omega d\Omega \, \rho(\mathbf{r}, \mathbf{r_i}) \approx 2 \times \sum_{i=1}^{p} w_i \frac{1}{N_g}\sum_{g} a_i \exp\left(-\frac{|\mathbf{r_g} - \mathbf{r_i}|^2}{d_i^2}\right), \tag{26}$$

$$\sum_{i,j=1}^{N} w_i w_j \, \rho(\mathbf{r_j}, \mathbf{r_i}) \approx \sum_{i,j=1}^{p} w_i w_j \rho(\mathbf{r_i}, \mathbf{r_j}). \tag{27}$$

In the above, N is 23, N_g is 104 (the total number of regularly distributed grid points), p is the number of stations used, and $\rho(\mathbf{r_i}, \mathbf{r_j})$ is computed by

$$\rho(\mathbf{r_i}, \mathbf{r_j}) \approx \frac{1}{360}\sum_{t=1}^{360} \Theta(\mathbf{r_i}, t)\Theta(\mathbf{r_j}, t). \tag{28}$$

We also carried out Monte Carlo experiments for PSE_f by randomly selecting p stations from the 23 stations. The PSE and PSE_f results are listed in Table 1. This table also lists the PSE results for the uniform weight. "Mean" in the table is the average of the 1000 values (because 1000 random drawings are performed) of PSE,

TABLE 1. The percentage sampling error (PSE) for the optimal averaging scheme (OA), and the arithmetic averaging scheme (AA). The PSE_f is the PSE estimated by using the formula (5).

Number of Stations (p)	Scheme OA		PSE_f		Scheme AA	
	Mean	Dev	Mean	Dev	Mean	Dev
3	30.16	5.31	29.12	4.92	36.80	8.39
5	21.79	3.59	21.97	3.22	27.64	5.28
7	16.93	2.57	17.94	2.27	22.18	3.86
9	13.95	2.09	15.46	1.81	18.46	3.00
11	11.60	1.74	13.70	1.45	15.51	2.46
13	9.82	1.47	12.37	1.19	13.13	2.00
15	8.44	1.20	11.37	0.98	11.18	1.66
17	7.26	0.97	10.54	0.77	9.34	1.38
19	6.23	0.65	9.85	0.54	7.54	0.99

and "Dev" is the deviation of these 1000 values from their "Mean". When the number of stations used in the computation increases, the accuracy of the regional average temperature also increases and hence *PSE* decreases. When the number of stations used is more than 15, the *PSE* is less than 12%. Thus, the smaller the "Mean" *PSE* is, the better the scheme. From Table 1, we can see that the optimal averaging scheme is consistently better than the arithmetic averaging scheme. The deviation of *PSE* for the optimal averaging scheme is consistently smaller than that for the arithmetic averaging scheme and hence the optimal averaging scheme can provide more robust results. We also see that *PSE* and PSE_f are about the same when we have fewer stations (less than 7). When we have more stations, the formula (5) tends to overestimate the sampling error (i.e., $PSE_f > PSE$).

We stress that the parameters a_i and d_i (needed for the optimal averaging scheme) are still those computed from the 23 stations even though we draw only p stations in the Monte Carlo experiments. One needs reasonable covariance patterns in order to use the optimal averaging scheme correctly, and these covariance patterns can only be reasonably estimated by using sufficiently dense data. Please also see the discussion in the next section.

5 Conclusion and discussion

The percentage sampling error in Table 1 shows the advantage of the optimal averaging scheme over the arithmetic averaging scheme. It follows that the optimal averaging scheme can achieve a more accurate regional average temperature. The deviation of the sampling error of the optimal averaging scheme is also less than that of the arithmetic averaging scheme. If the covariance pattern can be simulated more accurately, there is still room for the enhancement of the accuracy of the regional average by using the optimal averaging scheme. This is the rationale for using the EOF expression of a covariance function:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{n=1}^{\infty} \lambda_n \psi_n(\mathbf{r}) \psi_n(\mathbf{r}'),$$

where λ_n and ψ_n are respectively the eigenvalues and eigenfunctions (EOFs) of the covariance function $\rho(\mathbf{r}, \mathbf{r}')$. One can make use of the more recent observation data or GCM output to generate λ_n and ψ_n which are supposed to be robust over a period of decades to a century. This philosophy was proposed in Shen et al. (1994) where EOFs were computed from 40 recent years (1951–1990) of data (when data are dense and more reliable) and the average was computed for 100 years (1890–1990) (when data were sparse and not very reliable). It has attracted some attention for designing detection strategy. The combination of this philosophy and the idea of optimal regional averaging is an interesting research topic and will be deferred to subsequent studies.

Non-stationarity is always a problem in the studies of climate time-series. So far, there are no systematic statistical tools to deal with general non-stationary time-series. The common practice is to regard the anomaly time-series as being approximately stationary after: (a) the removal of the seasonal cycle and the trend, and (b) standardization. Since the trend in our data is small compared with the standard deviation in a time interval of five or ten years, the presence of the trend does not affect our computations based upon the stationarity property of the data and hence we choose not to go through the detrending procedure described earlier in Section 2.

Another question arises regarding the difference between the present method and the conventional objective analysis. As pointed out earlier in Section 3, the conventional objective analysis assumes a single covariance function like that of Vinnikov et al. (1990). It seems that the restriction of a single covariance function for the entire region makes it impossible to include any inhomogeneity property and hence forces the homogeneity assumption (Daley, 1991, pp. 109). If, assuming a single covariance function for the entire region, it is possible to use the fitted covariance function to interpolate the data onto some dense regular grid points and finally to compute the average of the interpolated field (Daley, 1991, Section 4.2). It might be an interesting project to develop a similar scheme when assuming N covariance functions. While it is natural to believe that there should be more applications of the present scheme of N covariance functions than that of a single covariance function because the former accounts for certain inhomogeneity properties, it is hard to say that the average obtained by the current method is necessarily better since for each analysis one can assume different covariance patterns and come out with different optimal averages even using the same scheme.

As for the computation, the algorithm for the optimal averaging scheme is more complicated than that of the arithmetic averaging scheme. But considering the gain from the optimal averaging scheme, this complication in computation, executed by computers anyway, is certainly worthwhile.

After we completed the revision, Gandin pointed out that the MSE formula (5) can be written in a simpler form (Gandin, 1993). Using Eqs (9) and (10), Eq. (5) can be written as

$$\varepsilon^2 = \frac{1}{A^2} \int_\Omega d\Omega \int_\Omega d\Omega' \, \rho(\mathbf{r}, \mathbf{r}') - \sum_{i=1}^{N} w_i \bar{\rho}(\mathbf{r}_i) + \Lambda. \tag{29}$$

One can then compute the MSE ε^2 following the above formula instead of the procedure given by Eqs (25)–(27).

Acknowledgment

This research was supported in part by the Atmospheric Environment Service, Canada. The authors wish to thank Rongfeng Li and Youlin Liang at the Institute of Atmospheric Physics, Chinese Academy of Sciences, for their helpful discussions regarding the Northeast China dataset and Lev Gandin for his guidance in the process of revising the paper.

References

DALEY, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

GANDIN, L.S. 1965. *Objective analysis of meteorological fields*. Gidrometeoizdat, Leningrad, Translated by Israel Program Scientific Translations, Jerusalem, 242 pp.

———. 1993. Optimal averaging of meteorological fields. NMC Office Note 397, 67 pp. [Available from NOAA/NCEP, 5200 Auth Road., Washington, DC 20233.]

HANSEN, J. and S. LEBEDEFF. 1987. Global trends of measured surface air temperature. *J. Geophys. Res.* **92**: 13, 345–13, 372.

JONES, P.D. 1994. Hemispheric surface air temperature variation: a reanalysis and an update to 1993. *J. Clim.* **7**: 1794–1802.

———; S.C.B. RAPER, R.S. BRADLEY, H.F. DIAZ, P.M. KELLY and T.M.L. WIGLEY. 1986a. Northern hemisphere surface air temperature variations: 1851–1984. *J. Clim. Appl. Meteorol.* **25**: 161–179.

———; ———, ———, ———, ——— and ———. 1986b. Southern hemisphere surface air temperature variations: 1851–1984. *J. Clim. Appl. Meteorol.* **25**: 1213–1230.

MADDEN, R.A.; D.J. SHEA, G.W. BRANSTATOR, J.J. TRIBBIA and R.O. WEBER. 1993. The effects of imperfect spatial and temporal sampling on estimates of the global mean temperature: experiment with model data. *J. Clim.* **6**: 1057–1066.

NORTH, G.R.; J.B. VALDES, E. HA and S.S.P. SHEN. 1994. The ground truth problem for satellite estimates of rain rate. *J. Atmos. Ocean. Tech.* **11**: 1035–1041.

RENDU, J.M. 1981. *An introduction to geostatistical methods of mineral evaluation*. 2nd Edition, Printpak(Cape)Ltd., Cape, 84 pp.

REYNOLDS, R.W. and T.M. SMITH. 1994. Improved global sea surface temperature analyses using optimal interpolation. *J. Clim.* **7**: 929–948.

SHEN, S.S.P.; G.R. NORTH and K.-Y. KIM. 1994. Spectral approach to optimal estimation of the global average temperature. *J. Clim.* **7**: 1999–2007.

———; X.C. WANG, Y.L. LIANG and R.F. LI. 1995. Optimal weighting scheme for averaging regional temperature I: theoretical analysis. *Chinese Sciences Bull.* **40**: 1005–1009.

TAO, S.; C. FU, Z. ZENG and Q. ZHANG. 1991. Two Long-term Instrumental Climatic Data Bases of the People's Republic of China. Oak Ridge National Laboratory/Carbon Dioxide Information Analysis Center, ORNL/CDIAC-47, NDP-039.

VINNIKOV, K.YA.; P. YA. GROISMAN and K.M. LUGINA. 1990. Empirical data on contemporary global climate changes (temperature and precipitation). *J. Clim.* **3**: 662–677.

WANG, X.C. and H.H. XIAO. 1991. 336-station 10-day surface temperature data over China. Institute of Atmospheric Physics, Chinese Academy of Sciences, Newsletter, No. 12.

WEBER, R.O. and R.A. MADDEN. 1995. Optimal averaging for the determination of global mean temperature: experiments with model data. *J. Clim.* **8**: 418–430.