The Value and Use of Averaged Homogenized Data Derived from Coads

Peter B. Wright
14 Shrewsbury Lane
Edgmond Newport, Shropshire TF10 8HT U.K.

Introduction

COADS is used for 2 main purposes:

1) Studying climate changes, trends and fluctuations;
2) Verifying models.

In this paper, we concentrate on the first of these uses.

The vast majority of climate changes involve space scales of at least 5 degrees latitude and 10 degrees longitude. Therefore, data are needed only on this or larger scales.

If the data were "perfect", the question is not too important. For example, with stations over a continent, one may simply average a number of stations over the required spatial area, or use one station as a sample for the area.

But COADS data are not perfect. They exhibit many sampling limitations, big gaps, and some periods and areas where data are very sparse. It is thus necessary to make best use of all observations.

If we simply average observations over an area, we are confronted by severe sampling problems. A month may be colder than the previous year simply because there happened to be no data for the warmer part of its area this particular year. The 2-degree resolution of the basic COADS products is generally considered to be the largest area for which such averaging is desirable. We need a method of averaging which does not introduce additional noise.

The best approach is to obtain anomalies and average those. This is because anomaly fields are in general more spatially coherent than actual fields.

Procedure

A method of generating anomalies in 4 degree latitude by 10 degree longitude boxes was developed some years ago, and has been outlined in several papers (Wright, 1986; Wright et al., 1985, 1988). In brief, it goes as follows. The following analysis is performed for each calendar month.

First, it is necessary to produce a mean field. For this purpose, we average the data over a specified reference period for each 2-degree box. If data are sparse, there will be sampling problems, because certain areas may be biased towards years that happened to have plentiful observations. Since the mean field is smooth, it is appropriate to calculate the mean in areas

of plenty and then interpolate and/or smooth into areas where there were few observations or few years represented. It is important that gaps are filled, because it enables us not to waste the observations there when it comes to calculating anomalies.

Then, anomalies in 2-degree boxes are obtained. This is straightforward: for each 2-degree box, the mean is subtracted. If there is a gradient across the 2-degree box, then the mean may not be fully representative of all parts of the box, but in practice it is not practical to take account of the position of each observation within the box.

Finally, anomalies are averaged over all the 2x2 boxes that comprise the 4xl0 box. This is satisfactory even if some boxes are missing, because the anomaly field is usually spatially coherent.

The result comprises the best available coverage, it does not waste any observations, and it is on a resolution appropriate for most studies. It is much more compact than the 2x2 data set, and is therefore more amenable to use on personal computers.

One limitation of the 4xl0 data set is that a 5x5 degree spacing is more common in gridded fields. If a future COADS release is on a 1x1 resolution instead of 2x2, it would be easy to construct a 5x5 instead of a 4xl0 set.

Uses of these data

For climate change studies. Virtually all climatically-significant features involve areas larger than 4xlO degrees. Indeed many features, such as the area of the equatorial Pacific used for Southern Oscillation (SO) studies, are much larger (e.g. Wright, 1984). Indices of the SO and for other key regions may readily be calculated from the 4x10 data.

For environmental workers. People studying other environmental subjects, such as pollution or fisheries, want to view their results in a climate context, and they want climate data sets that they can easily use. Those looking at global- or ocean-scale phenomena want to obtain the data from meteorologists and oceanographers with the fine detail averaged or smoothed out. They also want us to iron out all the inhomogeneities we can find. The 4xl0 set, being less noisy than the 2x2, is readily amenable to homogeneity investigations (Wright, 1986; Jones et al., 1986).

For comparison with models. Model data is often on a finer scale than 4xl0 degrees. But when verifying simulations against reality, the grosser scales are often all that is relevant. For example, Luksch and von Storch (1991) performed EOF analysis using COADS 4x10, which is certainly fine enough to resolve the first 2 EOFs over an ocean.

General. Users with vast computer power and an interest in detail have the choice of whether to use raw or processed data. But those with PCs do not have the choice. Those who are not

Meteorologists or Oceanographers do not want the choice. Therefore it is essential that averaged and homogenized data sets are made available.

Future

Climate change is becoming of importance to everyone. Therefore the need for global studies will increase. People in other environmental fields will need readily usable climate data. They will not want to plough through data on a fine space scale, or to have to repeat the research into inhomogeneities.

It is proposed that we undertake an appropriate averaging of the latest COADS for all relevant elements. We should then correct the series for inhomogeneities. The resulting sets would be easy to work with on PCs, and could be made readily available to the scientific community.

It is suggested further that we might set up global data sets by extending COADS, with suitable land station data, for all individual months this century. These also could be combined with all homogenization information available. This would aim to be the best available data set for climate change and environmental studies.

References

Jones, P.D., Wigley, T.M.L. and Wright, P.B., 1986: Global temperature variations between 1861 and 1984. *Nature*, **322**, 430-434.

Luksch, U. and von Storch, H., 1991: Modelling the low-frequency sea surface temperature variability in the North Pacific. Max-Planck Institut für Meteorologie, Report No 62.

Wright, P.B., 1984: Relationships between indices of the Southern Oscillation. *Monthly Weather Review*, **112**, 1913-1919.

Wright, P.B., 1986: Problems in the use of ship observations for the study of interdecadal climate changes. *Monthly Weather Review*, **114**, 1028-1034.

Wright, P.B., Mitchell, T.P. and Wallace, J.M., 1985: *Relationships between surface observations over the global oceans and the Southern Oscillation*. NOAA Data Report ERL PMEL-12.

Wright, P.B., Wallace, J.M., Mitchell, T.P. and Deser, C., 1988: Correlation structure of the El Niño/Southern Oscillation phenomenon. *Journal of Climate*, **1**, 609-625.