

From: Edward Cook <drdendro@ldeo.columbia.edu>  
To: Tim Osborn <t.osborn@uea.ac.uk>  
Subject: Re: N(eff) and practicality  
Date: Tue, 24 Jul 2001 08:49:14 -0400  
Cc: Phil Jones <p.jones@uea.ac.uk>, Keith Briffa <k.briffa@uea.ac.uk>

Hi Tim,

Thanks for the remarks. We can certainly spend some time talking through some of the points raised. I guess I am still finding it difficult to believe that an rbar of 0.05 has any operational significance in estimating Neff. It is kind of like doing correlations between tree rings and climate: a correlation of 0.10 may be statistically significant, but have no practical value at all for reconstruction. The same goes for an rbar of 0.05 in my mind. I agree that what I suggested (i.e. testing the individual correlations for significance and only using those above the some significance level for estimating rbar) is somewhat ad hoc and not theoretically pleasing. However, it is also true that correlations below the chosen significance threshold are "not significantly different from zero" and could be ignored in principle, just as we would do in testing variables for entry into a regression model. This would clearly muddy (a nice choice of words!) the rbar waters, I admit.

In terms of the problem I am working on (computing bootstrap confidence limits on annual values of 1205 RCS-detrended tree-ring series from 14 sites), it is hard to know what to do. Certainly, using Neff will result in almost none of the annual means being statistically significant over the past 1200 years. I don't believe that this is "true". Other highly conservative methods of testing significance result in a very high frequency of similarly negative results, i.e. the test of significance in spectral analysis that takes into account the multiplicity effect of testing all frequencies in an a posteriori way (see Mitchell et al. 1966, Climatic Change, pg. 41). If you use this correction, virtually no "significant" band-limited signals will ever be identified in paleoclimatological spectra. So, this test has very low statistical power. I think that this is the crux issue: Type-1 vs. Type-2 error in statistical hypothesis testing. The Neff correction greatly increases the probability of Type-2 error, while virtually eliminating Type-1 error. So, truth or dare.

Consider one last "thought experiment". Suppose you came to Earth from another planet to study its climate. You put out 1,000 randomly distributed recording thermometers and measure daily temperatures for 1 Earth year. You then pick up the thermometers and return to your planet where you estimate the mean annual temperature of the Earth for that one year. How many degrees of freedom do you have? Presumably, 999. Now, suppose that you leave those same recording thermometers in place for 20 years and calculate 20 annual means. From these 20-year records, you also calculate an rbar of 0.10. How many degrees of freedom per year do you have now? 999 or 9.9? What has changed? Certainly not the observation network. Does this mean that we can just as accurately measure the Earth's mean annual temperature with only 10 randomly placed thermometers if they provide temperature records with an rbar of 0.00 over a 20 year period? I wouldn't bet on it, but your theory implies it to be so. Surely, one would have more confidence (i.e. smaller confidence intervals) in mean annual temperatures estimated from a 1000-station network.

Cheers,

Ed

>Ed,

>

>re. your recent questions about Neff and rbar etc...

>

>I've thought a bit about these kind of questions over the past few years,

>but have never completely got my head around it all in a satisfactory way.

>I agree with what Phil said in his reply to you. Also, your idea of

>subsampling 40% of the cores at a time sounds reasonable, though I don't

>think it would be possible to write a very elegant statistical  
>justification! Anyway, I just wanted to add a couple of points to what  
>Phil said:  
>  
>(1) Even for very low rbar, the formula certainly works for  
>idealised/synthetic cases (i.e. with similar standard deviations and  
>inter-series correlations etc.). For example, I just generated 1000 random  
>time series (each 500 elements long) with a very weak common signal,  
>resulting in rbar=0.047. n=1000 was the closest I could get to n=infinity  
>without waiting for ages for the correlation matrix to be computed! The  
>formula:  
>  
>neff = n / ( 1 + [n-1]rbar )  
>  
>which reduces to neff = 1 / rbar for n=infinity gives neff = 20.83. For  
>such a low rbar, neff seems rather few? The mean of the variances of the  
>1000 series was 1.04677. If I took the "global-mean" timeseries (i.e. the  
>mean of the 1000 series, then it's variance was 0.05041. The ratio of  
>these variances is 20.77 - almost the same as neff! If our expectation  
>that neff should be higher than 20.83 was true, then the variance of the  
>mean series should have been much lower than it was. It should be easy to  
>try out similar synthetic tests with various options (e.g. shorter time  
>series, sets of series with differing variances, subsets with higher common  
>signal (within-site) combined with subsets with weaker common signal  
>(distant sites) etc.) to test the formula further.  
>  
>(2) I agree that rbar is computed from sample correlations rather than true  
>(population) correlations.  
>(a) For short overlaps, the individual correlations will rarely be  
>significant. But the true correlations could be higher as well as lower,  
>so rbar could be an underestimate and neff could be an overestimate! Maybe  
>you have even fewer than 20 degrees of freedom!  
>(b) I did wonder whether the sample rbar might be a biased estimate of the  
>population rbar, given that the uncertainty ranges surrounding individual  
>correlations are asymmetric (with a wider range on the lower side than the  
>higher side). But I've checked this out with synthetic data and the rbar  
>computed from short samples is uncertain but not biased.  
>(c) Just because rbar is only 0.05 does not mean that you need series 1500  
>elements long to be significant - that would be the case for testing a  
>single correlation coefficient. But rbar is the mean of many coefficients  
>(not all independent though!) so it is much easier to obtain significance.  
>Not sure how you'd test for this theoretically, but a Monte Carlo test  
>would work, given some assumptions about the core data. For 100 cores,  
>each just 20 years long, a quick Monte Carlo test indicates that an rbar of  
>0.05 is indeed significant - therefore rbar=0.05 in your case with > 100  
>cores, many of which will be > 20 years long, should certainly be significant.  
>  
>Looking forward to your visit! We can discuss this some more.  
>  
>Tim  
>  
>  
>Dr Timothy J Osborn | phone: +44 1603 592089  
>Senior Research Associate | fax: +44 1603 507784  
>Climatic Research Unit | e-mail: t.osborn@uea.ac.uk  
>School of Environmental Sciences | web-site:  
>University of East Anglia | http://www.cru.uea.ac.uk/~timo/  
>Norwich NR4 7TJ | sunclock:  
>UK | http://www.cru.uea.ac.uk/~timo/sunclock.htm

=====  
Dr. Edward R. Cook  
Doherty Senior Scholar  
Tree-Ring Laboratory  
Lamont-Doherty Earth Observatory  
Palisades, New York 10964 USA  
Phone: 1-845-365-8618  
Fax: 1-845-365-8152

Email: [drdendro@ldeo.columbia.edu](mailto:drdendro@ldeo.columbia.edu)

=====