The Problem of Red Noise in Climate Regime Shift Detection

Sergei N. Rodionov

Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington

Abstract

Time series of observations generated by a stationary red noise process are characterized by long intervals when the observations remain above or below the overall mean value. These intervals can be easily misinterpreted as "climatic regimes" with different statistics. A "prewhitening" procedure that removes the red noise component from the time series prior to an application of a regime shift detection technique is discussed. The key elements of this procedure are subsampling and bias correction of the least squares estimate of the serial correlation. A new technique to obtain a bias-corrected estimate of the autoregressive parameter is proposed. The approach is applied to the annual Pacific Decadal Oscillation (PDO) index that exhibits a regime-like behavior. The PDO regimes appear to be more than just a manifestation of a red noise process.

1. Introduction

Currently, a popular interpretation of long-term variability in the climate and biological records is based on the concept of "regimes" and "regime shifts." Common definitions of these terms usually involve the notion of multiple stable states in a physical or ecological system and a rapid transition from one state to another. The regime concept received a strong impetus after the shift in the North Pacific climate in 1976-77 [*Miller et al.*, 1994]. This shift clearly exhibited itself in the phase change of the first principal component of sea surface temperature in the North Pacific, known as the Pacific Decadal Oscillation or PDO [*Mantua et al.*, 1997]. The biological response to this shift is well documented (e.g., *Hare and Mantua*, 2000].

Rudnick and Davis [2003] questioned the interpretation of the PDO series as a sequence of genuine "regimes" with different statistics. Using Monte Carlo simulations, they showed that an equally plausible model for the PDO would be a Gaussian red noise process with stationary statistics. Exploring the idea that true regime shifts require the underlying dynamics to be nonlinear, *Hsieh et al.* [2005] arrived at the conclusion that large-scale marine ecosystem, due to their nonlinearity, have the capacity for dramatic change in response to stochastic fluctuations in basin-scale physical factors. However, key physical variables for the North Pacific, such as the PDO, are not deterministically nonlinear, and are best described as linear stochastic.

The purpose of this paper is to suggest a procedure of red noise removal from a time series with potentially "true" regime shifts. After this "prewhitening," the time series can be processed with any regime shift detection method. This approach is applied to the PDO series to see whether or not it represents something more than just a realization of a red noise process.

2. Structural Time Series Model

A structural time series model is one which is set up in terms of components that have a direct interpretation [*Harvey and Shephard*, 1993]. For example, a time series $\{X_t, t = 1, 2, \ldots, n\}$ can be seen as the sum of trend f_t and irregular component ε_t:

$$X_t = f_t + \varepsilon_t, \tag{1}$$

where ε_t are normally distributed independent random variables with zero mean value and variance σ^2. In the case of two regimes with different mean values, μ_1 and μ_2, and known change point c

$$f_t = \begin{cases} \mu_1, & t = 1, 2, \ldots, c-1, \\ \mu_2, & t = c, c+1, \ldots, n. \end{cases}$$

A realization of process (1) for $\mu_1 = -1$, $\mu_2 = 1$, $\sigma^2 = 1$, $c = 21$, and $n = 40$ is presented in Fig. 1a. Having a time series of observations, the direct approach to regime shift detection is to formulate the null hypothesis H_0 regarding the lack of a regime shift at $t = c$ (H_0: $\mu_1 = \mu_2 = \mu$), obtain the estimates $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\sigma}^2$, and then using, for example, the Student's t-test, try to reject the null hypothesis at the required probability level p. For the sample in Fig. 1a, the estimates are: $\hat{\mu}_1 = -1.08$, $\hat{\mu}_2 = 1.09$, $\hat{\sigma}^2 = 0.95$ and the null hypothesis can be rejected at $p = 0.00000003$ (two-tail t-test).

In climate research, the number of observations n typically ranges from a few dozens to a hundred or so points (years). Working with these relatively short time series, it is hard to draw any definitive conclusion about the underlying process based just on the data alone [*Percival et al.*, 2001]. For example, the time series in Fig. 1a might be easily mistaken for a realization of a stationary red noise process. This process is usually modeled by the first order autoregressive model, AR1, although other models can also be used [*Stephenson et al.*, 2000]. The AR1 model is defined as

$$(X_t - \mu) = \rho(X_{t-1} - \mu) + \varepsilon_t \,. \tag{2}$$

For the process to be stationary and causal, it is necessary for the autoregressive parameter ρ to satisfy the condition $|\rho| < 1$. When ρ is positive, the process is red noise, because its energy monotonically decreases as the frequency increases. If $\rho = 0$, it is white noise when the same energy is found at all frequencies. For negative values of ρ, it becomes violet noise, with energy monotonically increasing as the frequency increases. If $\rho = 1$, the process is called "random walk", for which the increments $(X_t - X_{t-1})$ are purely random.

By letting $\mu' = (1 - \rho)\,\mu$, (2) can be rewritten as

$$X_t = \rho X_{t-1} + \mu' + \varepsilon_t \,, \tag{3}$$

Without loss of generality, assuming $\mu' = 0$ and recursively substituting $X_{t-1}, X_{t-2}, \ldots$ by their own AR1 expressions, (3) can be rewritten as

$$X_t = \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-k} \tag{4}$$

Equation (4) shows that X_t is a result of integration, or low-pass filtering when $\rho > 0$, of all random impulses in the past [*Box and Jenkins*, 1970]. Such low-pass filters are common in nature. The ocean, for example, integrates atmospheric impulses, which can be viewed as white noise [*Hasselmann*, 1976]. Another example of an excellent low-pass filter is water level in large terminal lakes, such as the Caspian Sea, which efficiently integrates variations in runoff and over-lake precipitation [*Rodionov*, 1994].

Due to inertia in red noise processes determined by the value of ρ, they are characterized by extended intervals or "runs," when the time series remains above or below its mean value. These runs can be misinterpreted as different "regimes." Figure 1b shows a realization of AR1 process with $\rho = 0.7$. The regime shift at $i = 28$ could be statistically significant at the 0.05 level based on the t-test, if the data points were independent. Therefore, it is necessary to either recalculate the significance level by taking into account the serial correlation or use the so-called "prewhitening" procedure, which consists of removing red noise by using the difference ($X_t - \hat{\rho} X_{t-1}$). Both cases require the estimate $\hat{\rho}$ of the AR1 coefficient, which can be obtained using the entire series of observations.

The situation becomes more complicated, if the time series contains both regime shifts and red noise, i.e., the underlying model is

$$X_t = \rho X_{t-1} + f_t' + \varepsilon_t , \tag{5}$$

where $f_t' = f_t - \rho f_{t-1}$. In this case, using all the available data to estimate ρ would be misleading. For example, the ordinary list squares (OLS) estimate of ρ for the time series in Fig. 1a is $\hat{\rho}$ = 0.36, if all 40 data points are used.

A possible solution to this problem is to use subsampling. The size of subsamples should be chosen so that the majority of them do not contain change points. Assuming that regime shifts occur at a regular interval of l years, this condition is satisfied if the subsample size m is less than or equal to $(l + 1)/3$. In this case, the estimate of ρ can be chosen as the median value among the estimates for all subsamples. In practice, however, finding the right value of m requires some experimentation as discussed below. After the red noise is removed by taking the differences ($X_t - \hat{\rho} X_{t-1}$), the filtered time series $Z'_t = f_t' + \varepsilon_t$ can be processed with one of regime shift detection methods.

It is important to note that the magnitude of shifts in Z'_t is reduced by a factor of $(1 - \hat{\rho})$, which makes the shift detection more difficult. It is partly offset by a reduction of the variance in Z'_t by a factor of $(1 - \hat{\rho}^2)$. Also, what helps determining the timing of regime shifts is that the value of Z'_c tends to be amplified in the filtered time series. Indeed, within the regimes (when $f_t = f_{t-1}$), f_t' is constant and equal to the reduced mean value μ_1' for the first regime and μ_2' for the second. At the change point, which is the first point of the second regime, $f_c' = \mu_2 - \rho\mu_1$. Substituting $\mu_2 \pm \Delta\mu$, where $\Delta\mu$ is the difference between the mean values of the regimes, for μ_1, the latter can be rewritten as $f_c' = \mu_2 - \rho\mu_2 \pm \rho\Delta\mu = \mu_2' \pm \rho\Delta\mu$. It shows that Z'_c will be higher (lower) than other values for the second regime with the same white noise impulse ε_t by $\rho\Delta\mu$, when the regime shift is up (down). This amplification of change points facilitates the regime shift detection when the sequential method is used [*Rodionov*, 2004].

3. Parameter Estimation

The major problem in the outlined method is an accurate estimation of ρ for short subsamples of size m. It is well-known that the conventional estimators, such as the OLS or maximum likelihood techniques, yield biased estimates for ρ [*Shaman and Stine*, 1988]. There are two sources of the bias. First, if the true mean of the series μ is known, the serial correlations, will, in general, be biased (except when $\rho = 0$). The source of this bias is that the bivariate probability density function (X_{t-1}, ε_t) is not well defined [*Johnston*, 1984]. In practice, the mean has to be estimated from the sample, and this introduces a much larger bias, which is present even if $\rho = 0$ [*Marriott and Pope*, 1954].

Much research has been devoted to estimating the bias, although most efforts have considered the first order term of the bias, $O(m^{-1})$, and the case when the mean is known. Among those who considered a more complex situation with the unknown mean were *Marriott and Pope* [1954] and *Kendall* [1954], who gave the formula for the expected value of the OLS estimator of ρ:

$$E(\hat{\rho}) = \rho - \frac{1+3\rho}{m-1} + O\left(\frac{1}{m^2}\right). \tag{7}$$

In a practical situation, ρ is unknown. The expected value of $\hat{\rho}$ is also unknown, but following *Orcutt and Winokur* [1969], the procedure is to substitute $\hat{\rho}$, which is known, for $E(\hat{\rho})$ and then solve equation (7) for ρ. Solving for ρ and denoting this revised estimate of ρ by $\hat{\rho}^c$ yields

$$\hat{\rho}^c = \frac{(m-1)\hat{\rho}+1}{(m-4)} \tag{8}$$

Another method tested here is based on the assumption that the first approximation of the bias is approximately inversely proportional to m and is always negative [*Orcutt and Winokur,* 1969; *Stine and Shaman*, 1989]. Therefore, the first order bias-corrected estimate $\hat{\rho}^{c,1}$ is

$$\hat{\rho}^{c,1} = \hat{\rho} + 1/m\,.$$

The residual bias is also inversely proportional to m and its magnitude is a linear function of ρ. Since ρ and $\hat{\rho}$ are linearly related (which can be seen in Monte Carlo simulations), the latter can be substituted for the former. The method (called here IPN4 for short) uses three additional corrections of a smaller magnitude:

$$\hat{\rho}^{c,k} = \hat{\rho}^{c,k-1} + |\,\hat{\rho}^{c,k-1}\,|/m\,, \quad k = 2, 3, 4.$$

The IPN4 and Marriott-Pope and Kendall (MPK) methods were compared in a series of Monte Carlo experiments. The sample means and variances obtained in those experiments for the MPK estimates $\hat{\rho}^c$ were close to those computed by *Marriott and Pope* [1954, Table 5]. Table 1 shows the results of the experiment when one thousand time series of size $n = 40$ were generated for each given value of ρ. The OLS estimates were calculated for the subsample size $m = 5, 6, \ldots, 30$, and then corrected by the MPK and IPN4 methods. The results for $m = 5$, 10 and 20 are presented in Table 1. As seen, the OLS estimates are biased substantially even at $m = 20$ and the bias becomes larger as the serial correlation increases. The results of the MPK and IPN4 methods are similar to each other for $m \geq 10$. For smaller m, however, IPN4 substantially outperforms MPK in terms of both the magnitude of the bias and variability of the estimates.

4. The Effect of Prewhitening

Since prewhitening reduces the magnitude of regime shifts, the Monte Carlo technique was used to evaluate this effect. Using a Gaussian random number generator, 1000 time series of length 100 were generated for each pair of ρ and regime shift magnitude tested. A regime shift was imposed at point 51. These time series were processed using the sequential method [*Rodionov*, 2004] with two options, with and without prewhitening. In both cases the method was run with the following parameters*: target significance level $p = 0.01$, cutoff length $l = 20$ and Huber weight parameter $h = 1$. The serial correlation was estimated using the IPN4 method, with $m = 12$. After the prewhitening the values of the time series were considered independent, and the final significance levels for the shifts were calculated using the standard 2-tailed test with unequal variances. When prewhitening was not applied, the degrees of freedom for the regime shift index (which is used in the method to determine the timing of shifts) and the final significance levels were corrected for serial correlation. The correction formula, based on the so-called "equivalent sample size," is described in *Von Storch and Zwiers* [1999, p. 115].

Figure 2 illustrates the results of this Monte Carlo experiment for two magnitudes of the shift set at σ and 2σ. The number of hits (that is, correctly identified regime shifts) was calculated as the sum of regime shifts detected within plus/minus two points from point 51. The number of false alarms was calculated as the sum of regime shifts detected outside this area and excluding the last 20 points. Both numbers are shown as per 100 time series. As expected, the number of hits declines with the increase in serial correlation, and more so if the prewhitening is used (Fig. 2a). On the other hand, the number of false alarms substantially increases with ρ when prewhitening is not used (Fig. 2b). At $\rho = 0.9$, it reaches 100, which means that on average there is one false alarm in each time series. When prewhitening is used, the number of false alarms remains relatively low for the entire range of ρ. Therefore, prewhitening is a more conservative way of regime shift detection, with less chances of false alarm.

5. Is the PDO Simply Red Noise?

The red noise reduction technique, discussed above, was applied to the annual PDO values for the period 1900-2005. The MPK and IPN4 estimates are practically the same for subsample size $m > 11$ (Fig. 3). The estimates remain relatively stable at $\hat{\rho} \approx 0.45$, as m increases to 27. For greater m, $\hat{\rho}$ jumps to a higher level of about 0.60. This behavior of $\hat{\rho}$ is typical for the time series that represent a mixture of red noise with shifts in the mean. It shows that a characteristic time scale of the PDO regimes is about 25-30 years.

Figure 4a shows the PDO series and a stepwise trend with regime shifts in 1948, 1976 and 1999. These shifts were detected by applying the sequential method with the following parameters: $p = 0.05$, $l = 20$ and $h = 1$. The significance levels for the first two regime shifts are $2.1 \cdot 10^{-5}$ and $1.1 \cdot 10^{-5}$, respectively. These levels were calculated using the 2-tailed t-test and then

---

* More details about the method and parameters it takes can be found at www.BeringClimate.noaa.gov. The website also contains a downloadable macro with the code written in VBA for Excel.

adjusted for serial correlation using the equivalent sample size technique [*Von Storch and Zwiers*, 1999]. The autoregressive parameter $\hat{\rho}$ = 0.46 was estimated using the IPN4 method with the subsample size m = 12. As for the potential regime shift in 1999, the test is still in progress, since the number of years after that shift is less than the cutoff interval of 20 years used here. As of 2005, the significance level for this shift is 0.14.

Figure 4b shows the same PDO series after prewhitening, when the red noise component was removed by taking the difference ($X_t - 0.46X_{t-1}$). The sequential method was applied to this filtered time series with the same values of p, l and h parameters as above. Although the magnitude of the shifts in this case is reduced in comparison to those in Fig. 4a, they still remain statistically significant at the levels of $4.6 \cdot 10^{-4}$ for 1948 and $2.1 \cdot 10^{-4}$ for 1976. No shift was detected in 1999. To make the shifts of 1948 and 1976 to be statistically insignificant at the 0.05 level, the AR1 coefficient should be 0.8 or greater. Based on Monte Carlo simulations, the 95% confidence interval for ρ = 0.46 is 0.22-0.66. Hence, it is unlikely that these regime shifts are just manifestations of a red noise process. The red noise component (Fig. 4c), which is represented here by the difference in the PDO series before and after the prewhitening, accounts for about 25% of the total variance in PDO.

6. Concluding Remarks

The "prewhitening" procedure proposed here is designed to remove the red noise component from time series prior to applying a regime shift detection method. It does not intend to reveal the nature of climate regimes; it is simply a way to see whether or not these regimes can be more than just a red noise process. Two major elements of the procedure are: 1) subsampling and 2) bias correction of the OLS estimates of ρ. The size of subsamples should be as small as possible to minimize the effect of regime shifts on $\hat{\rho}$. On the other hand, it should be as large as possible to minimize the sampling variability of $\hat{\rho}$. These opposite requirements to the size of subsamples make it difficult to apply this procedure to short time series with decadal or shorter regimes. It works well, however, for relatively long time series with the regimes on the multidecadal time scale, such as the PDO.

Acknowledgements

To be added (N. Bond, D. Percival).

References

Box, G. E. P., and G. M. Jenkins (1970), *Time Series Analysis, Forecasting and Control*, Second printing, Holden-Day, San Francisco, Calif.

Hare, S. R., and N. J. Mantua (2000), Empirical evidence for North Pacific regime shifts in 1977 and 1989, *Progr. Oceanog., 47*, 103-146.

Harvey, A. C. and N. Shephard (1993), Structural time series models, in *Handbook of Statistics*, vol. 11, edited by G. S. Maddala, C. R. Rao and H. D. Vinod, Elsevier Science Publishers B.V.

Hasselmann, K. (1976), Stochastic climate models I: Theory, *Tellus, 28*, 473-485.

Hsieh, C. H., S. M. Glaser, A. J. Lucas, and G. Sugihara (2005), Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean, *Nature, 435*, 336-340.

Johnston, J. (1984), *Econometric Methods*, McGraw-Hill, New York.

Kendall, M. G. (1954), Note on bias in the estimation of autocorrelation, *Biometrika, 41*, 403-404.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997), A Pacific interdecadal climate oscillation with impacts on salmon production, *Bull. Amer. Meteorol. Soc., 78*, 1069-1079.

Marriott, F. H. C., and J. A. Pope (1954), Bias in the estimation of autocorrelations, *Biometrika, 41*, 390-402.

Miller, A. J., D. R. Cayan, and J. M. Oberhuber (1994), The 1976-77 climate shift of the Pacific Ocean, *Oceanography, 7*, 21-26.

Orcutt, G. H., and H. S. Winokur, Jr. (1969), Autoregression: Inference, estimation, and prediction, *Econometrica, 37*, 1-14.

Percival, D. B., J. E. Overland, and H. O. Mofjeld (2001), Interpretation of North Pacific variability as a short- and long-memory process, *J. Climate, 14*, 4545-4559.

Rodionov, S. (2004), A sequential algorithm for testing climate regime shifts, *Geophys. Res. Lett., 31*, L09204, doi:10.1029/2004GL019448.

Rodionov, S. N. (1994), *Global and Regional Climate Interactions: The Caspian Sea Experience*, Kluwer Academic Pub., Dordrecht, The Netherlands.

Rudnick, D. I., and R. E. Davis (2003), Red noise and regime shifts, *Deep-Sea Research, 50*, 691-699.

Shaman, P., and R. Stine (1988), The bias of autoregressive coefficient estimators, *J. Am. Statist. Assoc., 83*, 842-848.

Stephenson, D. B., V. Pavan, and R. Bojariu (2000), Is the North Atlantic Oscillation a random walk? *Int. J. Climatol., 20*, 1-18.

Stine, R., and P. Shaman (1989), A fixed point characterization for bias of autoregressive estimators, *Ann. Statist., 17*, 1275-1284.

Von Storch, H., and F. W. Zwiers (1999), *Statistical analysis in climate research*, Cambridge University Press.

Table 1. Sample means and standard deviations (in parentheses) of three estimates of ρ for different subsample size m.

m	ρ	OLS	MPK	IPN4
5	0.0	-0.28 (0.14)	-0.12 (0.59)	-0.02 (0.12)
5	0.4	-0.07 (0.15)	0.69 (0.60)	0.24 (0.22)
5	0.8	0.15 (0.18)	1.55 (0.71)	0.60 (0.30)
5	1.0	0.27 (0.19)	2.08 (0.81)	0.80 (0.34)
10	0.0	-0.11 (0.17)	0.00 (0.25)	0.02 (0.17)
10	0.4	0.18 (0.17)	0.44 (0.26)	0.39 (0.21)
10	0.8	0.48 (0.15)	0.86 (0.23)	0.76 (0.20)
10	1.0	0.59 (0.14)	1.06 (0.22)	0.92 (0.20)
20	0.0	-0.04 (0.18)	0.00 (0.21)	0.01 (0.18)
20	0.4	0.30 (0.18)	0.41 (0.22)	0.40 (0.21)
20	0.8	0.63 (0.15)	0.81 (0.18)	0.79 (0.18)
20	1.0	0.78 (0.13)	0.98 (0.16)	0.96 (0.15)

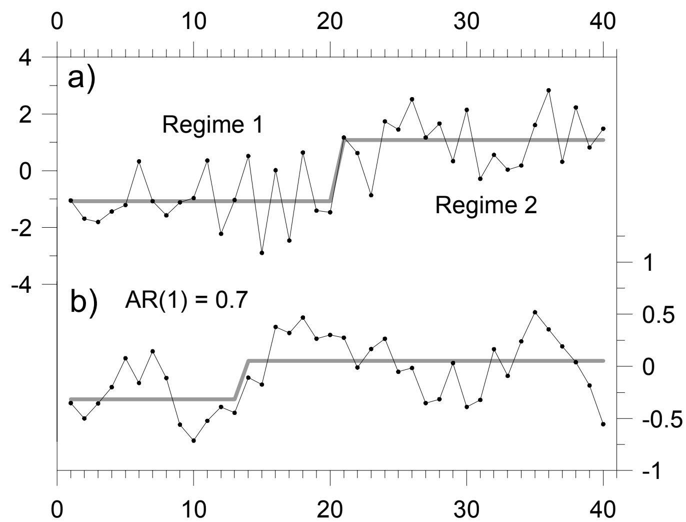

Fig. 1. Realizations of a) white noise process ($\sigma^2 = 1$) with a shift in the mean at t =21 from μ_1 = -1 to μ_2 = 1, and b) red noise process with ρ = 0.7. The shift at t = 14 in the latter case would be statistically significant at the 0.05 level, if the data points were independent.

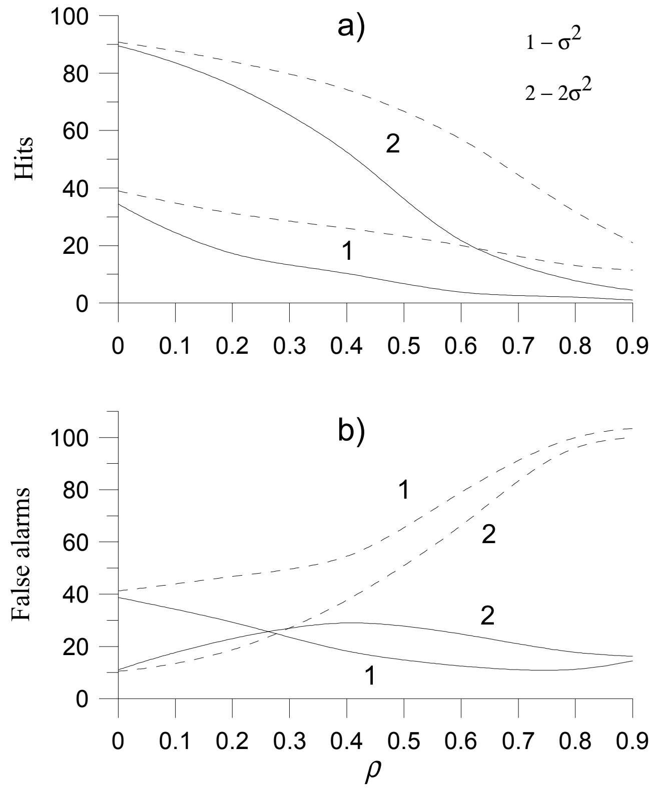

Fig. 2. Percentage of a) hits and b) false alarms in one thousand time series of size 100 with (solid line) and without (broken line) prewhitening. The regime shifts were imposed at point 51 and had the magnitudes of one and two standard deviations.

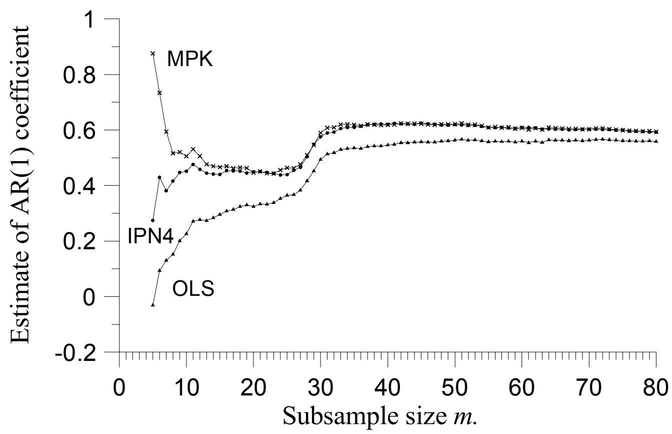

Fig. 3. OLS estimates of the annual PDO index with no bias correction and using the MPK and IPN4 techniques.

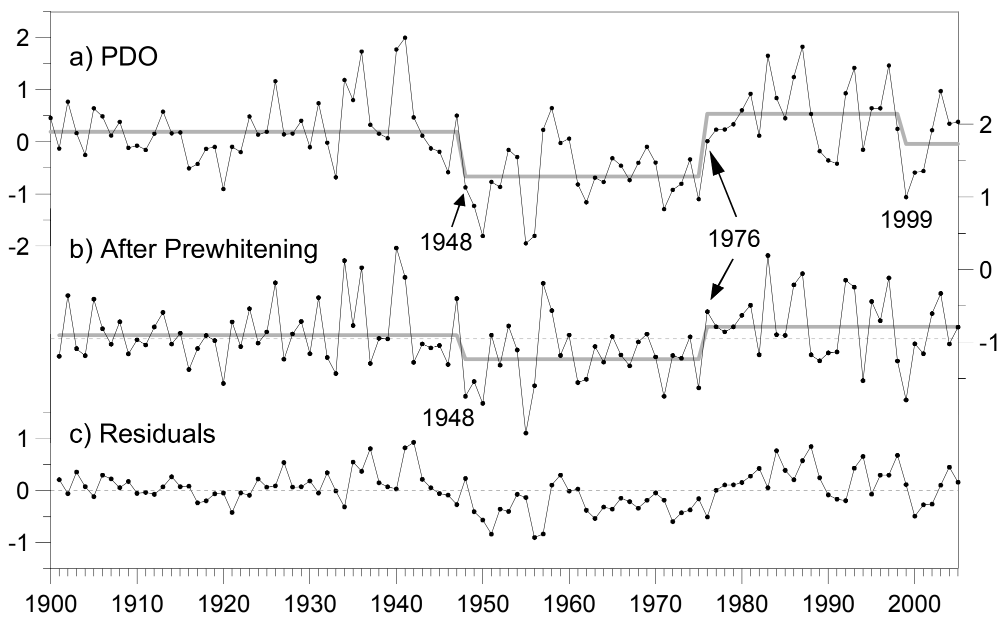

Fig. 4. a) Annual PDO index, 1900-2005, with a stepwise trend, b) the same time series after prewhitening, and c) the difference between the time series in a and b.