

1 **Referees' Comments and Author's Responses**

2 **Comment by Anonymous Referee #1**

3 Received and published: 21 April 2015

4 It is a very nice example of correct detecting relationships between climatological time
5 series. I'd like to notice that the math used is completely rigorous and well founded.
6 The presentation is clear and thorough. I recommend the manuscript for publication
7 with a couple of minor remarks. 1. A concept of linearly regular process is not a
8 common knowledge for readers of this journal. Please provide an appropriate
9 reference on line 23, page 6. 2. Line 15, page 7: "at" should be replaces by "as". 3. A
10 caption to Fig.2 should be extended to clarify which curve is what. 4. One of the
11 conclusions (I completely agree with) is "AMO is affected by SST3.4 but does not affect
12 SST3.4 explicitly". Does it mean that SST is a cause and AMO is a consequence? It
13 would be good to see comments on causation inferences in the conclusion section.

14 **Author's response to Anonymous Referee #1**

15 Received and published: 26 April 2015

16 Thank you for your comments.

17 1. Agreed. A clarification of this concept will require additional information about time
18 series' spectral properties which is not relevant for this study. Therefore, the entire
19 paragraph (lines 10-13 on page 703) should be removed.

20 2. Page 704, line 14. It's a misunderstanding. It should be "at" not "as".

21 3. Sorry, I believe that the caption is quite clear.

22 4. According to Eq. (9), p. 707, AMO and SST interact (all coefficients are statistically
23 significant) and this fact is included into conclusions (see #3, p. 711). A physical
24 nature of this phenomena requires a fluid dynamics model and cannot be explained
25 from probabilistic positions.

26 Sincerely, V. Privalsky

27

1 **Referee's response to the Author**

2 **L. Piterbarg**

3 piter@usc.edu

4 Received and published: 6 May 2015

5 Thanks for your explanations and comments. I notice that the manuscript has been
6 somewhat improved and in my opinion now it is completely ready for publication.

7 Interactive comment on Earth Syst. Dynam. Discuss., 6, 699, 2015.

8 **Comment by Prof. A. Gluhovsky**

9 **A. Gluhovsky**

10 aglu@purdue.edu

11 Received and published: 14 May 2015

12 The manuscript should be published as it contains important information about the
13 proper approach to multivariate time series analysis little known to climatologists as
14 well as new results about a teleconnection in the climate system. It is well written,
15 contains solid science, and provides vital information for those analyzing climate data.
16 It is probably worth mentioning in the paper that spectral methods of multivariate time
17 series analysis were also pioneered in the 1960s by C. W. J. Granger (2003 Nobel
18 Prize in Economics).

19 **Author's response to Prof. Glukhovsky**

20 vprivalsky@gmail.com

21 Received and published: 19 May 2015

22 Thank you for your suggestion. This was an omission. Two references will be added
23 to the article (p. 701, line 7 and in References):

24 Granger C.W.J, Hatanaka M. Spectral Analysis of Economic Time Series. Princeton
25 University Press 1964

26 and

1 Granger, C.W.J. Investigating causal relations by econometric models and
2 crossspectral methods. 37, 424-438, 1969.

3 **Prof. Glukhovsky's response to the author**

4 **A. Gluhovsky**

5 aglu@purdue.edu

6 Received and published: 20 May 2015

7 Thanks for having added the appropriate references. In my view, as mentioned in my
8 previous comment, the manuscript should be published.

9 **Comment by Anonymous Referee #2**

10 Received and published: 18 May 2015

11 I appreciate the approach of the author, but I wonder why no comparison (or even
12 citation) is made to works by researchers in networks and global teleconnections? For
13 example, Guez et al, EPL 2013; Ludescher et al, PNAS 2013, and others.

14 In my opinion, the suggested multivariate analysis should be developed at the same
15 scale as in the mentioned references, and directly compared in the same case study.

16 Without such comparison, the paper lacks up-to-date field overview, in my opinion.

17 **Author's response to comments by Anonymous Referee #2**

18 vprivalsky@gmail.com

19 Received and published: 19 May 2015

20 Thank you for your suggestion.

21 The article by Guez et al (2013) does not deal with time series analysis methods (at
22 least, explicitly), uses cross-correlation coefficients, and its objects of study do not
23 include the Atlantic Multidecadal Oscillation. Its "... aim is to follow climate dynamics
24 and to investigate the temporal stability of their structure" (p. 68006-p1), which is not
25 a goal in my article. The data consists of daily values of air temperature anomalies
26 "(actual values minus the climatological averaged over the years for each day)" on a
27 spatial grid, not of mean annual values as in my article. The authors do not analyze

1 any time series in the frequency domain nor does the article contain time series models
2 in the form of stochastic difference equations. Thus, it has no direct bearing on either
3 the mathematical methods described in my article (parametric multivariate time series
4 analysis), upon the time scales (climate), or upon the geographical objects (no AMO
5 data).

6 The article by Ludescher et al (2013) deals with spatial gridded data in the Niño area
7 3.4 (that is, SST3.4). Similar to the Guez et al article, it is based on the climate network
8 approach, not on time series analysis approach used in my article. It uses
9 crosscorrelation between nodes of the grid (random variables analysis) within the area
10 (no teleconnections through oceans as in my article) and deals with forecasting El Niño
11 events (not a subject in my article). Again, no overlap with my article either in methods
12 of analysis or the goals.

13 With all due respect, the articles that you recommend are not relevant to my article.

14 Referring to them in my article would be inappropriate.

15 Sincerely, V. Privalsky

16

List of corrections in the article

1

2 Page ## are shown in accordance with this document

3

1. Page 8, line 5: two references added in accordance with Prof. Gluhovsky's
recommendation

4

5

2. Page 10, lines 2-4: the paragraph removed in accordance with Prof. Piterbarg's
recommendation

6

7

3. Page 18, lines 14-17: two references added to the list of references in accordance
with Prof. Gluhovsky's recommendation

8

9

4. Page 20, lines 7- 10: caption to Figure 2 is corrected in accordance with Prof.
Piterbarg's recommendation

10

11

5. Page 21, line 3: caption to Figure 3 is corrected.

1 On Studying Relations between Time Series in Climatology

2 V. Privalsky¹

3 [1]{Space Dynamics Laboratory (ret.), Logan, Utah, USA}

4 Correspondence to: V. Privalsky (vprivalsky@gmail.com)

6 Abstract

7 Relationships between time series are often studied on the basis of crosscorrelation coefficients
8 and regression equations. This approach is generally incorrect for time series irrespective of the
9 crosscorrelation coefficient value because relations between time series are frequency-
10 dependent. Multivariate time series should be analyzed in both time and frequency domains,
11 including fitting a parametric (preferably, autoregressive) stochastic difference equation to the
12 time series and then calculating functions of frequency such as spectra and coherent spectra,
13 coherences, and frequency response functions. The example with a bivariate time series
14 'Atlantic Multidecadal Oscillation (AMO) – sea surface temperature in Niño area 3.4 (SST3.4)'
15 proves that even when the crosscorrelation is low, the time series' components can be closely
16 related to each other. A full time and frequency domain description of this bivariate time series
17 is given. The AMO – SST3.4 time series is shown to form a closed feedback loop system with
18 a two-year memory. The coherence between AMO and SST3.4 is statistically significant at
19 intermediate frequencies where the coherent spectra amount up to 55% of the total spectral
20 densities. The gain factors are also described. Some recommendations are offered regarding
21 time series analysis in climatology.

22 1 Introduction

23 Studying relations between time series on the basis of observations is a common task in all
24 branches of Earth sciences. Normally, it requires getting quantitative answers to the following
25 questions:

- 26 • what is the optimal time-domain stochastic model for a given multivariate time series?
- 27 • which components of the time series could be regarded as inputs and outputs of
28 respective climatic system?
- 29 • is there any interaction between the inputs and the outputs (are there any closed-
30 feedback loops within the system)?

- 1 • what are the statistical properties of the multivariate time series in the time and
2 frequency domains?

3 In this article, we will apply the methods first developed in theory of information (Gelfand and
4 Yaglom, 1956), time series analysis (Bendat and Piersol, 1966; Box and Jenkins, 1970),
5 [econometrics \(Granger and Hatanaka, 1964; Granger, 1969\)](#), and [in geophysics \(Robinson,](#)
6 [1967; Robinson and Treitel, 1980\)](#) to study relations between multivariate time series of
7 climatic data; the goal is to describe time series in the time and frequency domains, including
8 climatic teleconnections that can hardly be found within the correlation/regression approach.
9 Although methods of multivariate time series analysis are also described, mostly in the
10 frequency domain, in climatology (von Storch and Zviers, 1999) and oceanography (Emery and
11 Wilson, 2004), they are rarely applied in both time and frequency domains (e.g. Privalsky, 1988,
12 1995) or just in the frequency domain (*e.g.*, Oladipo, 1987, Schneider and Schönwiese, 1989;
13 Ghanbari et al., 2009; Park and Dusek, 2013).

14 **2 Elements of multivariate time series analysis**

15 Note first that the linear correlation/regression approach as a means of studying relations
16 between scalar time series, including teleconnections within the climatic system, is generally
17 inapplicable to time series analysis. The simplest example given in Privalsky and Jensen (1995)
18 and repeated in Emery and Wilson (2004) is a zero crosscorrelation coefficient between two
19 strictly linearly connected white noise sequences, one of which is obtained by applying a shift
20 operator to the other. A low correlation coefficient may occur between any time series related
21 to each other through more complicated but still strictly linear transformations. In particular, it
22 can be a time series and its first difference, or any autoregressive-moving average (ARMA)
23 time series and its innovation sequence, or the time series at the input and output of a linear
24 filter. The general statement is that if a time series is obtained from another time series through
25 a strictly linear inertial transformation the correlation coefficient between them will not be equal
26 to 1 in spite of the strictly linear dependence between them.

27 Relations between two time series (say, A and B), which are not mutually independent, always
28 correspond to one of the following three situations: A affects B but not vice versa, B affects A
29 but not vice versa, A and B affect each other (interaction). Again, the correlation/regression
30 approach does not allow one to determine what the actual situation is. It can be done within the
31 framework of the time-and-frequency domain analysis of multivariate time series.

1 The linear regression equation $B = \alpha A + \varepsilon$, where α is a constant and ε is a white noise sequence
 2 (regression error) means that the spectra $s_A(f)$, $s_B(f)$ of the time series A and B are identically
 3 shaped because $s_A(f) = \alpha^2 s_B(f) + \sigma_\varepsilon^2$ (where f is frequency and σ_ε^2 the regression error variance).
 4 This result is irrelevant when A and B are time-invariant random variables but if A and B are
 5 time series, it puts an unnecessary limitation upon their properties. In the general case, the
 6 shapes of the spectra are not identical, which would mean that ε is not white noise thus making
 7 the regression equation inadequate. This is another reason why both the crosscorrelation
 8 coefficient and respective linear regression equation cannot describe relations between time
 9 series.

10 The problem is solved if one uses methods of time series analysis including simultaneous
 11 description of multivariate time series in the time and frequency domains. It means fitting a
 12 stochastic difference equation to the time series, analyzing its properties in the time domain and
 13 then calculating and analyzing functions that describe the time series in the frequency domain.
 14 For a number of considerations (see below), the approach used here will be limited to the
 15 autoregressive (AR) case. Also, we will be regarding only the bivariate case. The extension to
 16 higher dimensions is rather simple (*e.g.*, Bendat and Piersol, 1966; Robinson, 1967) and will
 17 be briefly described at the end of this section.

18 Let the bivariate time series $\mathbf{x}_n = [x_{1,n}, x_{2,n}]'$ be a (zero mean) sample record of an ergodic
 19 discrete-time random process; here $n = 1, \dots, N$ is the dimensionless argument, $N\Delta t$ time series
 20 length in time units Δt , and the strike means matrix transposition. The sampling interval Δt is
 21 supposed here to be equal to 1.

22 In the time domain, the time series is described with a stochastic difference equation

$$23 \quad \mathbf{x}_n = \sum_{j=1}^p \Phi_j \mathbf{x}_{n-j} + \mathbf{a}_n, \quad (1)$$

24 where $p < \infty$ is the order of autoregression,

$$25 \quad \Phi_j = \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} \end{bmatrix} \quad (2)$$

26 are matrix AR coefficients, and $\mathbf{a}_n = [a_{1,n}, a_{2,n}]'$ is a bivariate innovation sequence (white noise)
 27 with a covariance matrix

$$\mathbf{R}_a = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}. \quad (3)$$

~~If the white noise \mathbf{a}_n is stationary, the AR coefficients are not time dependent, and the roots of the characteristic equation corresponding to Eq. (1) lie outside the unit circle, the bivariate time series (1) presents a sample of a stationary linearly regular random process.~~

The order p of an optimal AR model that agrees with the observed time series \mathbf{x}_n must be chosen on the basis of quantitative statistical considerations. Probably, the most efficient way to select an optimal order is to use order-selection criteria such as Akaike's AIC, Schwarz-Rissanen's BIC, Parzen's CAT, and Hannan-Quinn's ϕ (e.g., Box et al., 2008; Parzen, 1977; Hannan and Quinn, 1979).

Properties of the time series \mathbf{x}_n in the frequency domain are defined with the spectral matrix

$$\mathbf{s}(f) = \begin{bmatrix} s_{11}(f) & s_{12}(f) \\ s_{21}(f) & s_{22}(f) \end{bmatrix}, \quad (4)$$

which is obtained through a Fourier transform of stochastic difference equation (1) fitted to the time series \mathbf{x}_n . Here f is frequency in cycles per sampling interval (in our case, year⁻¹), $s_{11}(f)$, $s_{22}(f)$ are spectra and $s_{12}(f)$, $s_{21}(f)$ complex-conjugated cross-spectra of the time series components $x_{1,n}$ and $x_{2,n}$. In particular, the coherence function

$$Co_{12}(f) = \frac{|s_{12}(f)|}{[s_{11}(f)s_{22}(f)]^{1/2}} \quad (5)$$

describes the linear dependence between the time series components $x_{1,n}$ and $x_{2,n}$ in the frequency domain. It can be thought of as a frequency-dependent set of correlation coefficients between components of a bivariate time series. It is the coherence function $Co_{12}(f)$ (and not the crosscorrelation coefficient) that describes the degree of linear dependence between two scalar time series. Values of $Co_{12}(f)$ satisfy the condition $0 \leq Co_{12}(f) \leq 1$.

The importance of the coherence function in time series analysis and modelling is illustrated with the following property. If the components of an ergodic bivariate time series present processes at the input and output of any linear time-invariant system, the coherence between them will be equal to 1 at all frequencies where the spectral density is not too close to zero.

The spectral matrix (4) describes a linear stochastic system with the time series $x_{2,n}$ and $x_{1,n}$ as the system's input and output, respectively. The coherent spectral density, or coherent spectrum

$Cs_{12}(f) = Co_{12}^2(f)s_{11}(f)$ defines the part of the output spectrum $s_{11}(f)$ that can be explained by

1 the linear dependence between $x_{1,n}$ and $x_{2,n}$. The coherent spectrum $Cs_{21}(f) = Co_{12}^2(f)s_{22}(f)$
2 describes the share of the spectrum $s_{22}(f)$ defined by the contribution of $x_{1,n}$ to $x_{2,n}$. Finally, the
3 complex-valued frequency response function $G_{12}(f) = s_{12}(f)/s_{22}(f)$ shows in what manner
4 the spectral energy $s_{22}(f)$ of the input $x_{2,t}$ is transformed into the spectral energy $s_{11}(f)$ of the
5 output $x_{1,t}$ as well as the phase difference between them (the gain factor $|G_{12}(f)|$ and phase factor
6 $\varphi_{12}(f)$, respectively). The spectral characteristics calculated on the basis of Gaussian AR models
7 with properly selected autoregressive orders satisfy the requirements of the maximum entropy
8 method in spectral analysis. This is one of the reasons for selecting the AR modelling.
9 In the general case of an M -variate time series $\mathbf{x}_n = [x_{1,n}, \dots, x_{M,n}]'$, the time domain model is still
10 given by Eq. (1) with the matrix AR coefficients

$$11 \quad \Phi_j = \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} & \dots & \phi_{1M}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} & \dots & \phi_{2M}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1}^{(j)} & \phi_{M2}^{(j)} & \dots & \phi_{MM}^{(j)} \end{bmatrix} \quad (6)$$

12 and with the innovation sequence $\mathbf{a}_n = [a_{1,n}, \dots, a_{M,n}]'$. Its covariance matrix

$$13 \quad \mathbf{R}_a = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1M} \\ R_{21} & R_{22} & \dots & R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M1} & R_{M2} & \dots & R_{MM} \end{bmatrix}. \quad (7)$$

14 Consequently, the spectral matrix (4) changes to

$$15 \quad \mathbf{s}(f) = \begin{bmatrix} s_{11}(f) & s_{12}(f) & \dots & s_{1M}(f) \\ s_{21}(f) & s_{22}(f) & \dots & s_{2M}(f) \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1}(f) & s_{M2}(f) & \dots & s_{MM}(f) \end{bmatrix}, \quad (8)$$

16 where $s_{ii}(f)$ and $s_{ij}(f)$ are spectral (if $i = j$) and cross-spectral (if $i \neq j$) densities, respectively, of
17 the time series components $x_{i,n}$, $i = 1, \dots, M$. The elements of the spectral matrix (8) are used to
18 calculate spectra, multiple and partial coherence functions, multiple and partial coherent
19 spectra, and $M-1$ frequency response functions (see Bendat and Piersol, 2010). The spectral
20 matrix (8) is Hermitian.

1 It should be noted that if the multivariate time series is long (by orders of magnitude longer
2 than the largest time scale of interest) and if the spectra of its components are intricate, the
3 above-described approach may not be the best, especially in the time domain – because of the
4 high order of the optimal stochastic difference equation. In such cases the analysis may have to
5 be limited to a frequency domain description of the time series by using the nonparametric (e.g.,
6 Percival and Walden, 1993, Bendat and Piersol, 2010) approach. However, this can hardly
7 happen if one is interested in properties of the time series at climatic time scales.

8 **3 An example of bivariate time series analysis**

9 The example with actual climatic data given below proves that the components of a bivariate
10 time series can be connected to each other in spite of the fact that the crosscorrelation coefficient
11 between them is low. It also provides a simultaneous description of a climatic system in both
12 time and frequency domains. The case with a high crosscorrelation coefficient between the
13 components of the ENSO system (Southern Oscillation Index and SST variations in the Niño
14 area 3.4) has been treated in detail in Privalsky and Muzylev, 2013 where it was shown, in
15 particular, that both time series are close to white noise, interact with each other mostly through
16 the innovation sequence, and that the coherence function, coherent spectra, and the frequency
17 response functions between SOI and SST are frequency dependent.

18 The El Niño–Southern Oscillation (ENSO) system is believed to affect many phenomena in the
19 Earth climate (e.g., Philander, 1990). We will construct an autoregressive model of the bivariate
20 time series $\mathbf{x}_n = [x_{1,n}, x_{2,n}]'$, which consists, respectively, of the annual Atlantic Multidecadal
21 Oscillation (AMO) time series $x_{1,n}$ (Kaplan SST V2 data provided by the NOAA/ OAR/ESRL
22 PSD, Boulder, Colorado, USA, see Enfield et al., 2001) and sea surface temperature SST3.4 –
23 an oceanic component of the ENSO system ($x_{2,n}$). The SST3.4 data is computed from the
24 HadSST3 file at the University of East Anglia site (see Morice et al., 2011), for the same time
25 interval from 1876 through 2014 (Fig. 1a).

26 As seen from the figure, the two components behave in a different manner: AMO contains
27 much stronger low-frequency variations than SST3.4. The correlation between AMO and
28 SST3.4 (Fig. 1b) is very low, with the crosscorrelation coefficient 0.06 and the maximum
29 absolute values of the crosscorrelation function below 0.26. With the crosscorrelation-based
30 approach that prevails in climatology, the conclusion would have to be that the two scalar time
31 series are either not related to each other at all or that the connection between them is very
32 weak. And it would not be correct.

1 Consider first the time domain properties. All four above-mentioned order selection criteria
2 selected the same order $p = 2$. The respective AR(2) model is

$$\begin{aligned} 3 \quad x_{1,n} &\approx 0.59x_{1,n-1} + 0.09x_{2,n-1} + 0.21x_{1,n-2} - 0.07x_{2,n-2} + a_{1,n} \\ 4 \quad x_{2,n} &\approx -1.28x_{1,n-1} + 0.29x_{2,n-1} + 1.00x_{1,n-2} - 0.22x_{2,n-2} + a_{2,n} \end{aligned} \quad (9)$$

4 with the innovation covariance matrix

$$5 \quad \mathbf{R}_a = \begin{bmatrix} 0.014 & 0.007 \\ 0.007 & 0.290 \end{bmatrix}. \quad (10)$$

6 All coefficients in Eq. (9) are statistically significant at a confidence level of 0.95.

7 Obviously, Eq. (9) describes a closed feedback loop system: AMO ($x_{1,n}$) depends upon two of
8 its previous values and upon two previous values of SST3.4 ($x_{2,n}$) and SST3.4 depends upon
9 two previous values of both SST3.4 and AMO. The eigen-frequency of this system was found
10 to be 0.24 year^{-1} so that respective period of about 4 years is the time required to close the
11 system's feedback loop.

12 The stochastic difference equations (9) and the innovation sequence covariance matrix (10)
13 allow one to understand how much of the variances of AMO and SST3.4 can be explained by
14 the "deterministic" components of the model (9) that describe the dependence of $x_{1,n}$ (AMO)
15 and $x_{2,n}$ (SST3.4) upon their own past values and upon past values of the other time series. The
16 variance of AMO $\sigma_1^2 \approx 0.035(\text{°C})^2$ while, according to Eq. (10), the variance R_{11} of the
17 innovation sequence $a_{1,n}$ is $0.014(\text{°C})^2$. Therefore, the part of the AMO variance σ_1^2 which
18 cannot be explained by the dependence of the time series upon their past behavior is $0.014/0.035$
19 ≈ 0.4 . Thus, Eq. (9) allows one to explain about 60% of the AMO variance by its dependence
20 upon its own past values and upon the past values of SST3.4.

21 The SST3.4 variance $\sigma_2^2 \approx 0.37(\text{°C})^2$ while the variance R_{22} of the innovation sequence $a_{2,n}$ is
22 $0.29 (\text{°C})^2$. Thus, the unexplainable part of the SST3.4 variance amounts to almost 80% of the
23 SST3.4 variance. The results for AMO and SST3.4 differ because AMO and SST3.4 have very
24 dissimilar spectra (Fig. 2a).

25 Both AMO and SST3.4 time series are Gaussian so that their autoregressive spectral estimates
26 satisfy the requirements of the maximum entropy spectral analysis. The AMO spectrum $s_{11}(f)$
27 quickly decreases with frequency, which is characteristic of spatially-averaged climatic
28 processes. Such behavior of the spectrum means that AMO is strongly dependent on its past
29 values. The SST3.4 spectrum $s_{22}(f)$ is not monotonic, has a wide maximum at intermediate

1 frequencies and, in general, does not differ much from a white noise spectrum. The dependence
2 on its past values is weak.

3 Though the crosscorrelation coefficient between AMO and SST3.4 is very low, the maximum
4 entropy estimate of coherence $Co_{12}(f)$ shown in Fig. 2b is statistically significant everywhere
5 except at the frequencies below 0.13 year^{-1} and above 0.44 year^{-1} . It is bell-shaped and exceeds
6 0.6 within the frequency band from 0.20 year^{-1} to 0.36 year^{-1} , that is, roughly, at time scales
7 between 3 and 5 years. Its maximum value is 0.74 at $f \approx 0.27 \text{ year}^{-1}$. This behavior of the
8 coherence function means that the components of this bivariate time series contribute to each
9 other up to about 55% of the spectral density at intermediate frequencies.

10 The coherence between AMO and SST3.4 is weak at the low-frequency end, where AMO's
11 spectral energy is much higher than elsewhere. The high coherence occurs at intermediate
12 frequencies where the spectral density of AMO is much lower. The strong dependence of AMO
13 on its past values and the relative closeness of the SST3.4 spectrum to a constant seem to be the
14 reasons why the stochastic model (9) can explain so much of the total AMO variance and less
15 of the total SST3.4 variance.

16 The contribution of SST3.4 to the AMO spectrum is $Cs_{12}(f) = Co_{12}^2(f)s_{11}(f)$ where $s_{11}(f)$ is
17 the AMO spectrum (Fig. 3a). Respective contribution of AMO to SST3.4 shown in Fig. 3b is
18 $Cs_{21}(f) = Co_{12}^2(f)s_{22}(f)$. These coherent spectrum estimates are statistically significant within the
19 frequency band from 0.18 year^{-1} to 0.38 year^{-1} where they amount to 25% – 55% of the spectral
20 densities $s_{11}(f)$ and $s_{22}(f)$. This is a substantial contribution but it occurs within the frequency
21 band where the spectral density of AMO is at least an order of magnitude smaller than at lower
22 frequencies where the coherence between AMO and SST3.4 is not significant. The
23 crosscorrelation coefficient that “integrates” the coherence function is small in spite of the
24 relatively close connection between the two processes at moderate frequencies for at least two
25 reasons: the complex structure of the interdependence between the time series components
26 expressed by Eqs. (9) and (10) which cannot be described with a linear regression equation and
27 the low absolute contribution of SST3.4 to the AMO variance.

28 This example also shows that using proper methods of analysis allows one to avoid filtering of
29 time series in order to suppress ‘noise’. Indeed, though the low-frequency variations dominate
30 the spectrum of AMO, the coherence function has revealed the ‘signal’ – a teleconnection
31 between AMO on SST3.4 at intermediate frequencies where the AMO spectrum is low. This is
32 another useful property of autoregressive time and frequency domain models.

1 Our frequency-domain results generally agree with the earlier results by Park and Dusek (2013)
2 regarding the connection between AMO and the Multivariate ENSO Index (MEI) at
3 intermediate frequencies. The authors used a nonparametric spectral estimation – singular
4 spectrum analysis keeping 10 first empirical orthogonal functions that cover slightly over 75%
5 of the time series total variances. At frequencies above 0.15 year^{-1} , our estimate of coherence
6 is quite similar to the coherence estimate in Park and Dusek (2013). However, the authors did
7 not estimate the frequency response function (FRF) because, according to them, “a physically-
8 based transfer function is likely [to be] nonlinear”. Actually, a nonlinear theoretical model of
9 the FRF between AMO and SST3.4 time series would have been in disagreement with
10 observations because both time series are Gaussian. At frequencies from 0.1 year^{-1} to 0.4 year^{-1} ,
11 the gain factors of the empirical FRFs AMO – SST3.4 and SST3.4 – AMO (Fig. 4) behave
12 in the same manner as the coherent spectra shown in Fig. 3 and their values at intermediate
13 frequencies amount to approximately 0.1 and 4, respectively. The coherent spectra and gain
14 factors are shaped similarly because the coherence function is bell-shaped and the AMO and
15 SST3.4 spectra are rather flat at intermediate frequencies.

16 The phase factors in this case cannot give explicit information about the AMO – SST3.4 system
17 because its feedback loop is closed (interaction between AMO and SST3.4). We cannot
18 compare our spectra with those shown in Park and Dusek (2012) because their spectral
19 estimates are given without confidence bounds but generally the shapes of the spectra at
20 frequencies below 0.5 year^{-1} seem to be rather similar.

21 **4 Conclusions and comments**

22 1. Relations between time series should not be studied on the basis of crosscorrelation
23 coefficients and regression equations. An efficient approach within the framework of
24 time series analysis includes two stages both involving parametric (preferably,
25 autoregressive) modelling:

- 26 • fitting a stochastic difference equation to the time series (time domain), analyzing the
27 selected model, and
- 28 • using the fitted equation to calculate and analyze frequency domain characteristics
29 (spectra and coherent spectra, coherence function(s), gain and phase factor(s)).

30 This two-pronged approach is little known in climatology and related sciences.

- 1 2. Methods of multivariate time series analysis should be used in all cases, irrespective of
2 the value of the crosscorrelation coefficient. The crosscorrelation and regression
3 coefficients do not generally describe relations between time series. In particular, a low
4 crosscorrelation coefficient does not necessarily mean the lack of even a strictly linear
5 dependence between the time series.
- 6 3. The stochastic difference equation (9) with the innovation sequence covariance matrix
7 (10) shows quantitatively that AMO and SST3.4 interact with each other so that AMO
8 and SST3.4 can be regarded as either inputs or outputs to the AMO – SST3.4 system.
9 It also reveals that the system's memory extends for two years. The dependence of AMO
10 and SST3.4 upon their own past and upon the past of the other time series explains about
11 60% and 25% of the AMO and SST3.4 variances, respectively.
- 12 4. The frequency domain analysis of the system shows that the spectra of AMO and
13 SST3.4 behave in a different manner, with the AMO spectrum decreasing fast with
14 frequency and with a relatively flat SST3.4 spectrum.
- 15 5. In spite of the very low crosscorrelation coefficient between the time series of AMO
16 and SST3.4, a close linear dependence between them was shown to exist at intermediate
17 frequencies corresponding to time scales from 3 to 5 years. The coherence between
18 AMO and SST3.4 is statistically significant in a wide frequency band centered at 0.26
19 year⁻¹ where the coherence peaks at 0.74. This result has been obtained earlier by Park
20 and Dusek (2013) for a similar climatic system.
- 21 6. The coherent spectra AMO – SST3.4 and SST3.4 – AMO are statistically significant at
22 frequencies from 0.18 to 0.38 year⁻¹ contributing 25%-55% of the spectral densities of
23 SST3.4 and AMO to each other. The gain factors in the band behave in a manner similar
24 to the behavior of the coherent spectra and peak at about 0.1 and 4.0, respectively.

25 These conclusions provide answers to the questions formulated at the introduction to this work.
26 Time series can often be treated as Gaussian. The ability to use a Gaussian approximation is
27 important because for such time series the nonlinear approach cannot give better results than
28 what is obtained within the linear approximation. This latter statement holds, in particular, for
29 time series extrapolation within the framework of the Kolmogorov-Wiener theory. Also, as
30 shown by Choi and Cover (1984), the random process that maximizes the entropy rate under

1 constrains on p first values of the correlation function is a Gaussian autoregressive process of
2 order p .

3 Many time series, especially at climatic time scales, are short, that is, their length does not
4 exceed the time scales of interest by orders of magnitude. Therefore, the non-parametric
5 analysis in the frequency domain may not be efficient because, with short time series, it would
6 produce less reliable results. Besides, the nonparametric approach does not allow one to obtain
7 explicit stochastic models in the time domain. (These are two more reasons to prefer the
8 parametric modeling.)

9 A parametric (first of all, autoregressive) analysis in time and frequency domains is effective
10 because it results in relatively accurate estimates due to the postulation of a stochastic model
11 for the time series. In particular, the frequency domain estimates obtained with the properly
12 selected order satisfy, in the Gaussian case, the requirements of the maximum entropy spectral
13 analysis. However, it is not correct to say that any autoregressive spectral estimate has this
14 important property. The number of parameters to be estimated should always be much smaller
15 than the time series length. If, for example, the length $N\Delta t$ of a bivariate time series is 128 years,
16 one should hardly expect statistically reliable results for models with AR orders higher than 5
17 (20 AR coefficients plus 3 elements of the noise covariance matrix to be estimated). This is one
18 of the reasons why it is not possible to make any realistic conclusions about the presence of 60-
19 70 years long “periods”, “cycles”, “oscillations”, or about any other features at such large time
20 scales unless the time series is at least 300 - 400 years long. For parametric models, it is strongly
21 recommended to determine the model’s order and, consequently, the number of parameters to
22 be estimated, on the basis of order selection criteria; an improper selection of the order
23 invalidates the results of analysis.

24 **References**

- 25 Bendat, J. and Piersol A.: Measurement and Analysis of Random Data. Wiley, New York, 1966.
- 26 Bendat, J. and Piersol A.: Random Data. Analysis and Measurements Procedures, 4th ed. Wiley,
27 Hoboken, 2010.
- 28 Box G. and Jenkins J.: Time Series Analysis. Forecasting and Control. Holden-Day, San
29 Francisco, 1970.
- 30 Box, G., Jenkins, G. and Reinsel, G.: Time Series Analysis. Forecasting and Control, 4th ed.
31 John Wiley and Sons, Hoboken, NJ, 2008.

1 Choi, B. and Cover, T.: An information-theoretic proof of Burg's maximum entropy spectrum.
2 Proceedings of the IEEE, 72, 1094-1096, 1984.

3 Emery, W. and Thomson, R.: Data Analysis Methods in Physical Oceanography, 2nd ed.
4 Elsevier, Amsterdam, 2004.

5 Enfield D., Mestas-Nunez A. and Trimble P.: The Atlantic Multidecadal Oscillation and its
6 relationship to rainfall and river flows in the continental U.S., Geophys. Res. Letters 28, 2077-
7 2080, 2001.

8 Gelfand, I. and Yaglom, A.: Calculation of the amount of information about a random function
9 contained in another such function, Uspekhi Matematicheskikh Nauk, 12:3-52, 1957. English
10 translation: American Mathematical Society Translation Series 2, 12,199-246, 1959.

11 Ghanbari, R., Bravo, H., Magnuson, J., Hyzer, W. and Benson, B.: Coherence between lake ice
12 cover, local climate and teleconnections (Lake Mendota, Wisconsin). J. Hydrol. 374, 282-293,
13 2009.

14 Granger, C.W.J. and Hatanaka, M. Spectral Analysis of Economic Time Series. Princeton
15 University Press, Princeton, New Jersey, 1964.

16 Granger, C.W.J. Investigating causal relations by econometric models and crossspectral
17 methods. 37, 424-438, 1969.

18 Hannan, E. and Quinn B.: The determination of the order of an autoregression, J. R. Statist.
19 Soc., B-41, 190-195, 1979.

20 Morice, C., Kennedy, J., Rayner, N. and Jones, P.: Quantifying uncertainties in global and
21 regional temperature change using an ensemble of observational estimates: the HadCRUT4
22 dataset. J. Geophys. Res. 117, D08101, [doi:10.1029/2011JD017187](https://doi.org/10.1029/2011JD017187), 2012.

23 Oladipo, E.O.: Power spectra and coherence of drought in the Interior Plains. J. Climatol. 7,
24 477-491, 1987.

25 Park, J. and Dusek G.: ENSO components of the Atlantic multidecadal oscillation and their
26 relation to North Atlantic interannual coastal sea level anomalies. Ocean Sci., 9, 535-543, 2013.

27 Parzen, E.: Multiple time series: determining the order of autoregressive approximating
28 schemes. In Multivariate Analysis-IV, North Holland Publishing Company, Amsterdam, 283-
29 295, 1977.

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Don't adjust space between Latin and Asian text,
Don't adjust space between Asian text and numbers

Formatted: Font: (Default) Times New Roman, 12 pt

1 Percival, D.D, and Walden, A.T. Spectral Analysis for Physical Applications. Multitaper and
2 Conventional Univariate Techniques. Cambridge University Press, 1993.

3 Philander, S.: El Niño, La Niña, and the Southern Oscillation, Academic Press, San Diego,
4 1989.

5 Privalsky, V.: Stochastic models and spectra of interannual variability of mean annual surface
6 temperature in the North Atlantic. Dynamics of Atmospheres and Oceans, 12, 1-18, 1988.

7 Privalsky, V., Protsenko, I., and Fogel, G.: The sampling variability of autoregressive spectral
8 estimates for two-variate hydrometeorological processes, Proc. 1st World Congress of the
9 Bernoulli Society on Mathematical Stat. Theory and Applications, Vol. 2. VNU Science Press,
10 Utrecht, 651-654, 1987.

11 Privalsky, V. and Jensen, D.: Assessment of the influence of ENSO on annual global air
12 temperature. Dynam. Atmosph. Oceans, 22, 161-178, 1995.

13 Privalsky, V. and Muzylev, S.: An Experimental Stochastic Model of the El Niño – Southern
14 Oscillation System at Climatic Time Scales, Universal J. Geoscience, 1, 28-36, 2013.

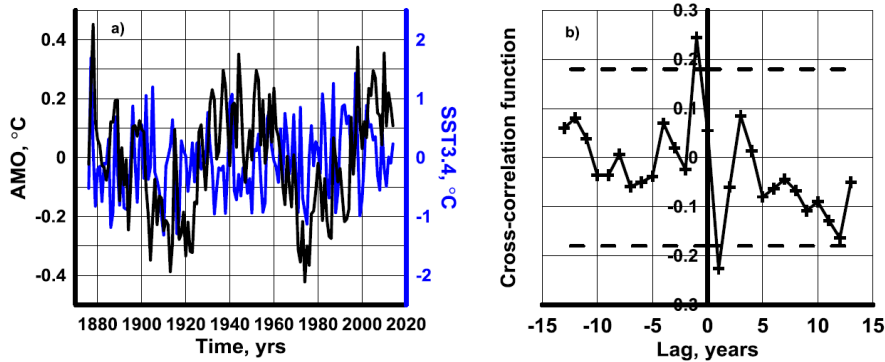
15 Robinson, E.: Multichannel Time Series Analysis with Digital Computer Programs. Holden-
16 Day, San Francisco, 1967.

17 Robinson, E. and Treitel, S.: Geophysical Signal Analysis. Prentice-Hall: Englewood Cliffs,
18 NJ, 1980.

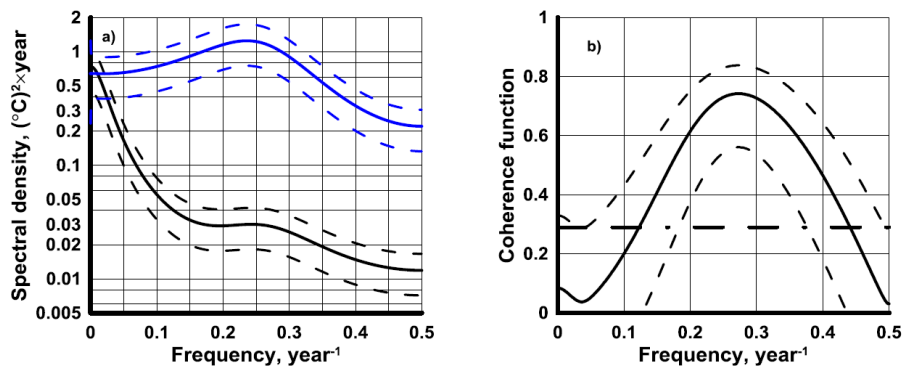
19 Schneider, U. and Schönwiese, C.-D.: Some statistical characteristics of El Niño-Southern
20 Oscillation and the North Atlantic Oscillation indices. *Atmósfera*, 2, 167-180, 1989.

21 von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research. Cambridge University
22 Press. Cambridge, 1999.

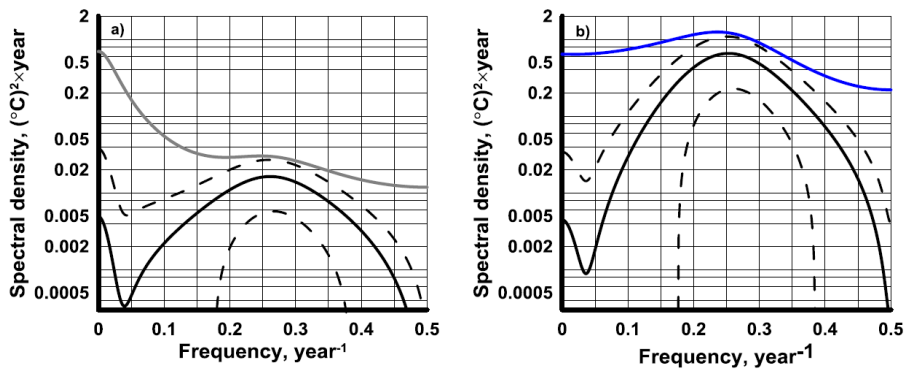
23



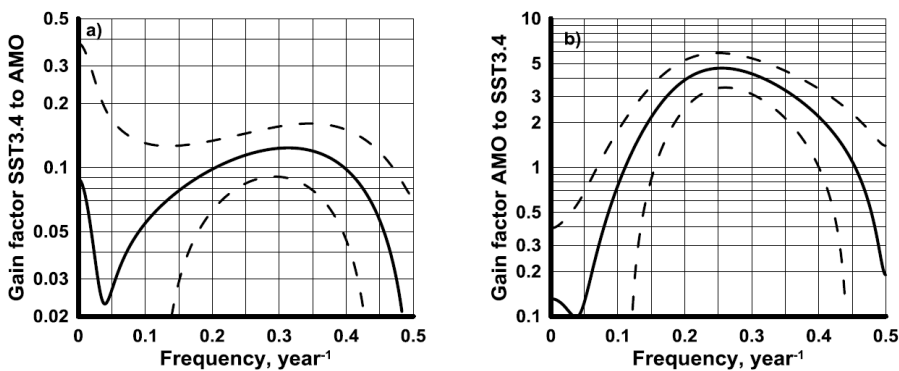
1
 2 Figure 1. Observed AMO and SST3.4, 1876-2014 (a) and AMO–SST3.4 crosscorrelation
 3 function estimate (b); the dashed lines show approximate 90% confidence bounds.



4
 5
 6 Figure 2. (a) spectra of AMO (black lines) and SST3.4 (blue lines) with approximate 90%
 7 confidence bounds (dashed lines); (b) coherence function AMO – SST3.4 (solid line) with
 8 approximate 90% confidence bounds (dashed lines). The dot-dashed line shows the
 9 approximate upper 90% confidence bound for the true zero coherence. For confidence limits see
 10 Privalsky et al. (1987).



1
 2 Figure 3. Coherent spectra: contribution of SST3.4 to the AMO spectrum (a), and contribution
 3 of AMO to the SST3.4 spectrum (b). Dashed, ~~and~~ grey and blue lines show approximate 90%
 4 confidence bounds and AMO and SST3.4 spectra, respectively.



5
 6 Figure 4. Gain factors SSR3.4-AMO (a) and AMO-SST3.4 (b) with approximate 90%
 7 confidence bounds (dashed lines).