Supplementary Materials:

The Combined Effect of Atmospheric and Solar Activity Forcings on the Hydroclimate in Southeastern Europe

Ileana Mares *, Venera Dobrica *, Crisan Demetrescu and Constantin Mares

Institute of Geodynamics, Romanian Academy, 020032 Bucharest, Romania; crisan@geodin.ro (C.D.); constantin_mares_ro@yahoo.com (C.M.)
* Correspondence: ileana_mares@geodin.ro (I.M.); venera@geodin.ro (V.D.)

Canonical Correlation Analysis (CCA)

Statistical analysis for data size reduction was first introduced by Hotelling [1]. Then, this analysis was applied by Lorenz [2] for linear processes of weather and climate and by Hasselmann [3] and Hsieh [4,5] for nonlinear climate processes. Hsieh [6] considers the nonlinear connection of two data sets as a generalization of Canonical Correlation Analysis (CCA), which he calls nonlinear CCA (NLCCA) and solves it with a neural network method. Widmann [7], when a time series is linearly estimated from a time-dependent vector, considers that CCA is equivalent to multiple linear regression (MLR).

Canonical correlation analysis assesses the relationship between two sets of variables or two vectors (X and Y) whose components are the predictor random variables and the predictand random variables:

X=(X1,... , Xn);

Y=(Y1,... , Ym) (in general n<= m).

If there are correlations between these variables, the CCA does nothing but find linear combinations of Xi and Yj, for which the correlations (with each other) are maximal.

Nowadays, several CCAs have been used in various applications, especially for linear relationships between phenomena. The computational algorithm is given in MATLAB routines by the routine "canoncorr", which computes the linear combinations (varieties) U and V of the sets of variables X and Y [8, 9]; thus:

U=(X-mean(X))*A;

V=(Y-mean(Y))*B.

Where A and B are sample canonical coefficients for X input variables and Y input variables, respecively.

In [6], the foundations of nonlinear linkages in CCA by making neural predictions with this technique were described.

Here, we use the algorithm [5] only to the extent that we can specify the nature of the relationships between the vectors of the data sets and their components. Hsieh states that if the nonlinear approach has a significant advantage over the linear approach, it is seen in the nature of the relationships between the data sets. The nonlinear approach is generally inefficient if the data sets are short and noisy or when the relationships between data sets are linear. But it can be said that both forms of linkages are found in nature, and when we roughly approximate nonlinear linkages by linear ones, nature is mutilated.

In the following, the type of curve connecting two time series was determined. Some links are linear others curvilinear as in Figures S1-S5. Testing to see if the nonlinear link between the two series is chaotic is performed [10] following Papoulis [11] by calculating the mutual information size (see Equation 10 in the paper). Mutual information is the normalized entropy of the connection of the phenomena described by the two chronological series in question, and a priori is a measure of indeterminacy (uncertainty) and a posteriori is a measure of information [12].

Therefore, in the following section, we show by testing that the linkages between components of predictor–predictand sets can be of both forms, such as in figures S1-S5 below, where the relevant combinations of any two variables are shown in plane (a,b,c) and in (d) the overall combination of the three variables, x1,x2, and x3, which are represented in space. In these figures, the blue points are the data, and those shown by a curve made up of red small circles are a kind of fit of the data.

Linear links are usually found among the components of vectors describing geophysical phenomena, such as in Figure S3 (a) and Figure S4. In other cases, the existing nonlinearity is evident as can be seen in Figures S1, S2, and S5.

In conclusion, the need to test in advance the nature of the linkage between the components of the predictor–predictand sets is a condition of interest not only theoretically but also practically in order not to reach false conclusions.

Figures S1 and S2 show the linkages between the components of the predictor–predictand sets in the seasonal series (SPR= spring; SUM= summer), denoted by X1=PHDI, X2=SSN, and X3=Q, for the spring and summer seasons, respectively, and the linkages are clearly nonlinear.

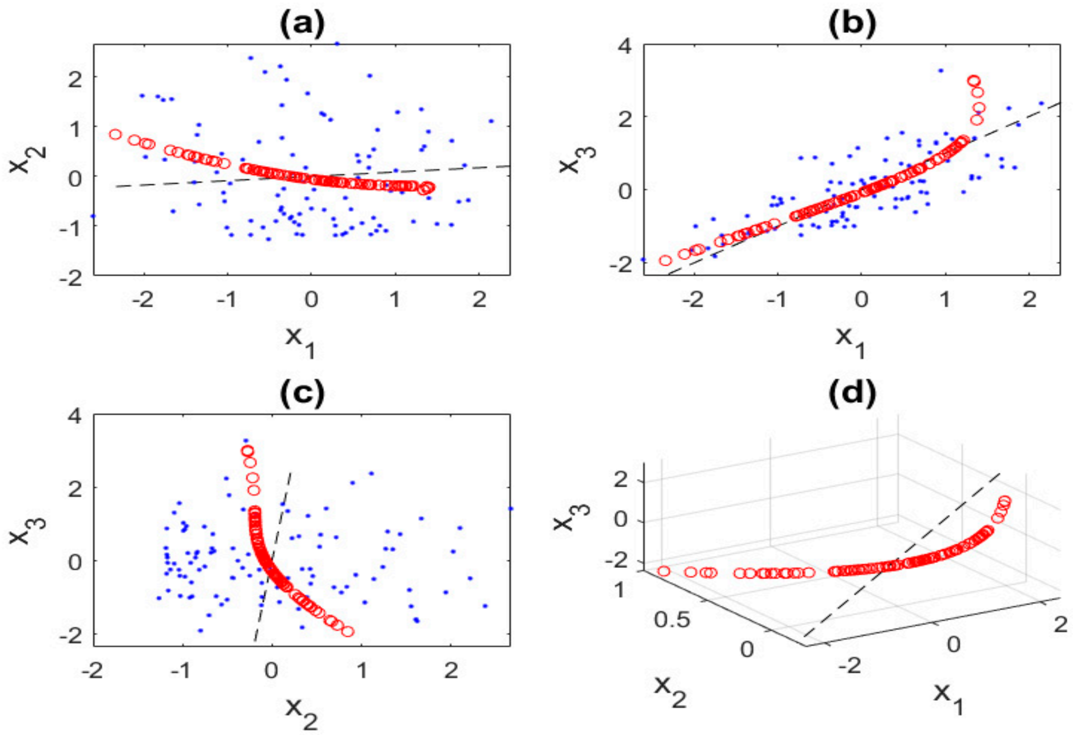

Figure S1. Relationship between (a) the PHDI (X1) and SSN (X2); (b) the PHDI (X1) and Q (X3); (c) the SSN (X2) and Q (X3); (d) the spatial representation of the overall combination of the three variables in the spring season. The blue points are the data, and those shown by a curve made up of red small circles are a kind of fit of the data.

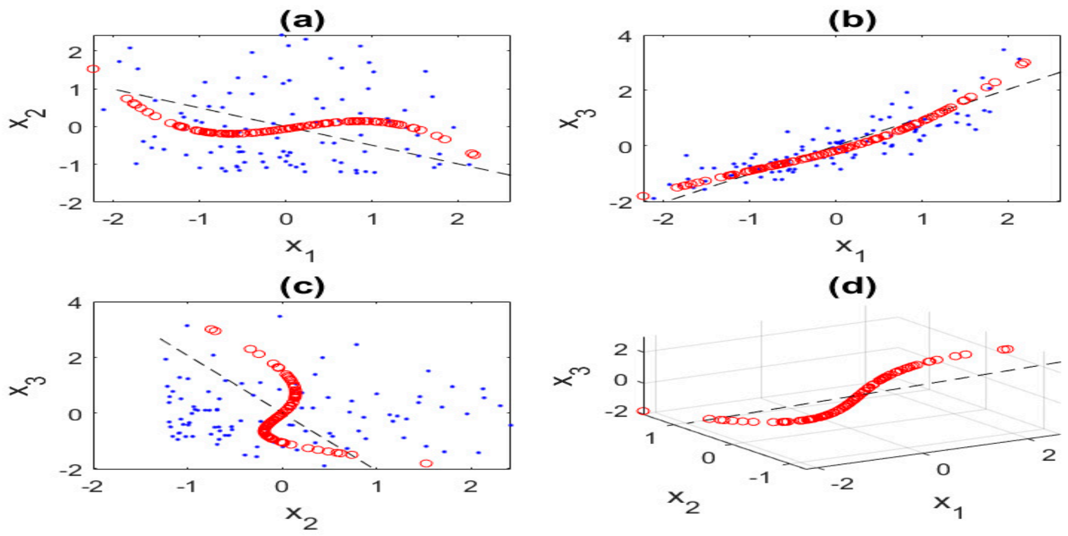

Figure S2. Same variables as in Figure S1 but for the summer season.

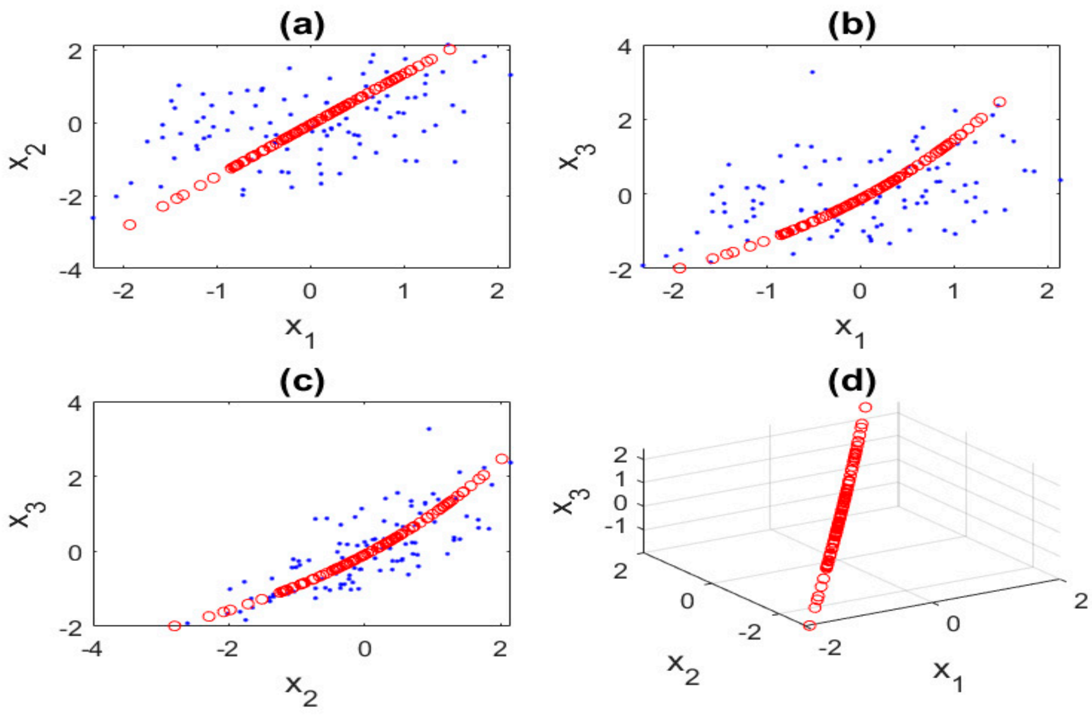

Figure S3. Relationship between the GBOI (X1), PHDI (X2), and Q (X3) during the spring season.

In Figure S3 (a), between the GBOI (X1) and PHDI (X2), the link is linear. The water component Q is closely related to the nonlinear PHDI (Figure S3 (c)), which has a relatively obvious scedasticity (increasing spread around the curve formed by red circles as one goes from small to large deviations). In S3 (d), the overall combination of the three variables also seems to be linear, which means that the combination (X1,X2) is very close.

Figure S4, in which the variables formed by the annual series of the GBOI, PHDI, and Q (river discharge), show links with linear structures.

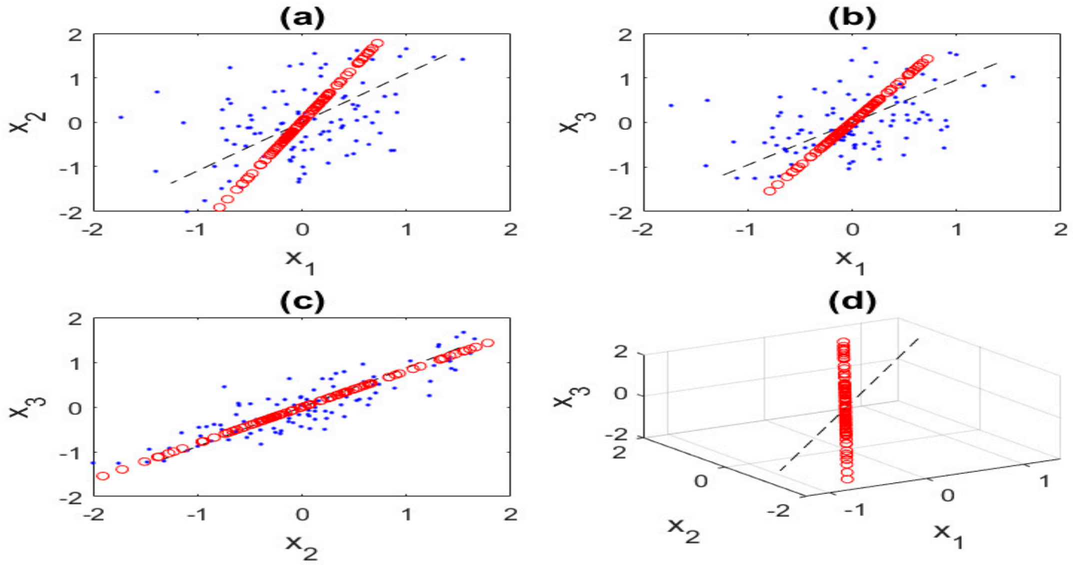

Figure S4. Relationship between the GBOI (X1), PHDI (X2), and Q (X3) for annual values.

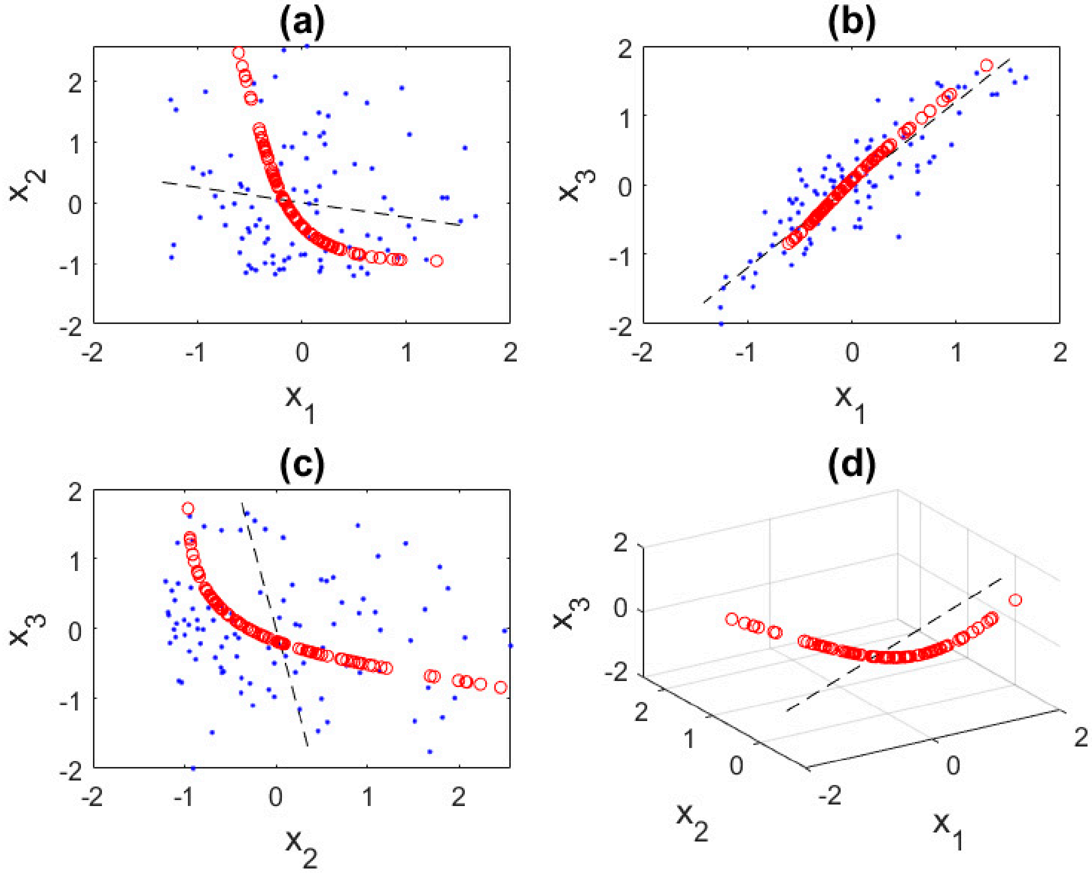

Figure S5. Relationship between the Q (X1), SSN (X2), and PHDI (X3) for annual values.

Figure S5 highlights the link between the annual solar activity SSN and the other parameters designating the terrestrial phenomena: the Danube discharge (Q) and the basin moisture through the PHDI.

In fact, the links between the analyzed parameters, either seasonal or annual, are nonlinear if solar activity is included.

In the end, we must conclude that it is difficult to give physical explanations of the nonlinear links between phenomena when the intimate processes in one of them are not known in detail. But with mutual information and entropy transfer [13], the causal connection can be specified somewhat.

If we think about Saltzman's postulate [14], under an initial impulse, a system like the climatic one, then we have "free variations due to internal instabilities and feedbacks, usually involving nonlinear interaction among different components of the climatic system, that can occur even if there are no forcing changes". The effect of even a weak solar signal on the Earth's climate system can lead to significant climate variations if the system is nonlinear [15].

In fact, only a stochastic–dynamic model of the terrestrial climate can provide us with truthful information for physical explanations.

References

1. Hotelling, H. Relations Between Two Sets of Variates. *Biometrika* **1936**, *28*, 321–377. https://doi.org/10.1093/biomet/28.3-4.321.
2. Lorenz, E.N. Empirical orthogonal functions and statistical weather prediction. In *Statistical Forecasting Project;* Department of Meteorology, Massachusetts Institute of Technology: Cambridge, MA, USA, 1956; p 49.
3. Hasselmann, K.PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.* **1988**, *93*, 11015–11021. https://doi.org/10.1029/JD093iD09p11015.
4. Hsieh, W.W. Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *J. Clim.* **2001**, *14*, 2528–2539.
5. Hsieh, W.W. Nonlinear principal component analysis of noisy data. *Neural Netw.* **2007**, *20*, 434–443.
6. Hsieh, W.W. Nonlinear multivariate and time series analysis by neural network methods. *Rev. Geophys.* **2004**, *42*, RG1003.
7. Widmann, M. One-Dimensional CCA and SVD, and Their Relationship to Regression Maps. *J. Clim.* **2005**, *18*, 2785–2792.
8. Krzanowski, W.J. *Principles of Multivariate Analysis: A User's Perspective*; Oxford University Press: New York, NY, USA, 1988.
9. Seber, G.A.F. *Multivariate Observations*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1984.
10. Moore, J.; Grinsted, A.; Jevrejeva, S. Is there evidence for sunspot forcing of climate at multi-year and decadal periods? *Geophys. Res. Lett.* **2006**, *33*, L17705.
11. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1984.
12. Guiasu, S. *Information Theory with Applications*; McGraw-Hill Inc.: London, UK, 1977.
13. Timme, N.M.; Lapish, C. A tutorial for information theory in neuroscience. *Eneuro* **2018**, *5*, 1–40. https://doi.org/10.1523/ENEURO.0052-18.2018.
14. Saltzman, B. A survey of statistical–dynamical models of the terrestrial climate. In *Advances in Geophysics*; Academic Press: New York, NY, USA; San Francisco, CA, USA; London, UK, 1978; Volume 20, pp. 183–304.
15. Ogurtsov, M.G.; Raspopov, O.M.; Oinonen, M.; Jungner, H.; Lindholme, M. Possible Manifestation of Nonlinear Effects When Solar Activity Affects Climate Changes. *Geomagn. Aeron.* **2010**, *50*, 15–20.