Theor Appl Climatol (2011) 103:103–107
DOI 10.1007/s00704-010-0285-8

ORIGINAL PAPER

Detecting human influence on climate using neural networks based Granger causality

A. Attanasio · U. Triacca

Received: 17 August 2009 / Accepted: 30 March 2010 / Published online: 25 April 2010

Abstract In this note we observe that a problem of linear approach to Granger causality testing between CO_2 and global temperature is that such tests can have low power. The probability to reject the null hypothesis of non-causality when it is false is low. Regarding non-linear Granger causality, based on multi-layer feed-forward neural network, the analysis provides evidence of significant unidirectional Granger causality from CO_2 to global temperature.

1 Introduction

There is little doubt that much of the Earth has been undergoing a pronounced warming since about the start of the twentieth century. An important question is: Has this global warming been primarily due to natural fluctuations or anthropogenic influences? Among a number of man-made 'greenhouse' gases which absorb and emit thermal radiation, such as nitrous oxide and methane, CO_2 is widely believed to be the major 'greenhouse' gas which has the greatest influence on the global climate. Consequently, the relationships between global temperature and CO_2 concentration have been the subject of intense research in the last two decades.

In particular Triacca (see Triacca 2005) finds that there is no detectable Granger causality (see Granger 1969) from radiative forcing due to CO_2 concentration to global surface temperature anomalies. In other terms, past observations of radiative forcing due to CO_2 do not significantly improve the predictability of current temperature. In this note we show that this (negative) result is due to the inappropriateness of the linear version of the Granger non-causality test utilized by the author. The problem of linear approach to causality testing is that such tests can have low power detecting certain kinds of non-linear causal relations. Regarding non-linear Granger causality based on multi-layer feed-forward (MLF) neural network between CO_2 and global temperature, the analysis provides a quite different story. It provides evidence of significant unidirectional non-linear Granger causality from CO_2 to global temperature.

The paper is organized as follows: A short review of neural networks considering their usage in this work is given in the next section; The results of non-linear Granger causality analysis are presented in Section 3; Some concluding remarks are given in Section 4.

2 Methodology and data

We are interested to test the null hypothesis that CO_2 concentration does not Granger cause global surface temperature. The linear framework is too simple to catch the complex non-linear behavior of the temperature with respect CO_2 concentration. To overcome this problem we use a non-linear out-of-sample analysis of the dynamic relationship between these variables based on artificial neural network models.

Climate change studies represent a field in which neural networks can be applied successfully (see Pasini

A. Attanasio · U. Triacca (✉)
Università di L'Aquila, L'Aquila, Italy
e-mail: umberto.triacca@ec.univaq.it

A. Attanasio
e-mail: alessandro_attanasio@yahoo.it

2009; Pasini et al. 2006). Neural network capability for non-linear modeling and forecasting has been established in the literature both theoretically and empirically (see for example Cybenko 1989; Pasini and Ameli 2003; Pasini et al. 2001; Yi and Prybutok 1996; Wu 1995). A brief description of how neural networks are defined and applied is presented in this section. More detailed material about the topic can be found in Bishop (see Bishop 1995).

MLF neural networks are networks of processing elements (called nodes or neurons), that are ordered in layers. The first layer is called the input layer, the last layer is called the output layer, and the layers in the middle are hidden layers. Each neuron in the network is able to receive input signals, to process them and to send an output signal. Each neuron in a particular layer is connected with all neurons in the next layer.

Although many types of MLF neural networks have been proposed, the most popular one for time series forecasting is the so-called three-layer feed-forward neural network. The MLF neural networks used in this study are of this kind. The use of the three-layer feed-forward neural network models is motivated by the fact that one hidden layer feed-forward network can represent some continuous functions to an arbitrary accuracy given enough hidden neurons. A three-layer feed-forward neural network with a single output neuron and n hidden neurons is a non-linear, non-parametric regression model specified as:

$$O = f\left(\sum_{i=1}^{n} w_i f\left(\sum_{j=1}^{m} w_{ij} I_j + \theta_i\right) + \theta\right)$$

where w_i is the connecting weight from the ith hidden neuron to the output neuron, w_{ij} from the ith input neuron to the jth hidden one, θ_i between the constant input and the ith hidden neuron and θ is the threshold coefficient associated with the output neuron; O is the output of the network, $I_1, \ldots, I_m$ are the input signals of the network and $f : \mathbb{R} \to (0, 1)$ is the so-called activation function. In this paper f is a sigmoid function. The hidden layer(s) in the network and the non-linear activation function of neurons allow non-linearity to be introduced into neural networks.

We denote with z_t the radiative forcing due to CO_2 concentration and with v_t the global surface temperature anomalies. z_t is in the range [0.10594, 1.6291], while v_t is in the range [−0.519, 0.580]. The data are described in Triacca (see Triacca 2005). The mechanisms underlying the relationship between CO_2 concentrations and temperature are well understood. Essentially, they have to do with the radiative trapping properties of CO_2. The physical climate system primarily responds to changes in radiative forcing. For calculate the radiative forcing (RC) due to CO_2 we have used the following simplified expression

$$RC = 5.35 \ln\left(\frac{CO_2}{C_0}\right)$$

with $C_0 = 280$, where 280 is the assumed pre-1750 CO2 atmospheric concentration. This formula is proposed in Myhre (see Myhre et al. 1998). The time frame of the study is 1856 to 2005, with T = 150 observations. Figure 1 shows the time series of radiative forcing over the 150-year period as described. The temperature data over the same period are shown in Fig. 2.

The raw data must be scaled between the upper and lower bounds of the sigmoid activation functions. We use the linear transformations

$$y_t = \frac{z_t - (\min_z - 0.05)}{(\max_z + 0.05) - (\min_z - 0.05)}$$

where $\min_z = 0.10594$, $\max_z = 1.6291$ and

$$x_t = \frac{v_t - (\min_v - 0.05)}{(\max_v + 0.05) - (\min_v - 0.05)}$$

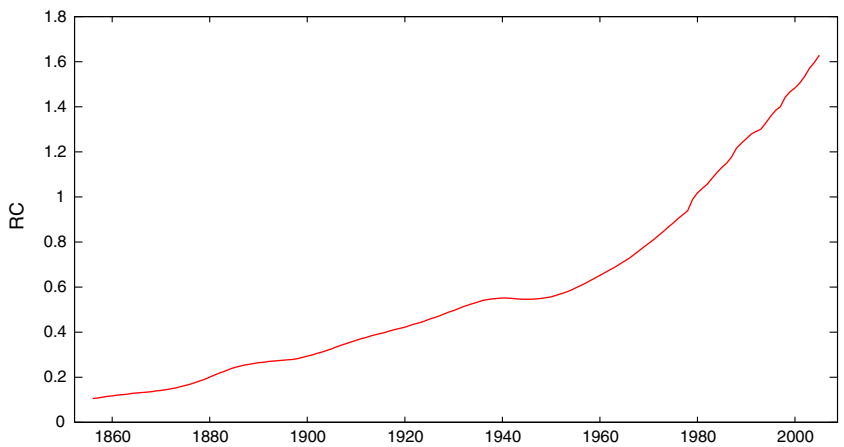

Fig. 1 Radiative forcing (1856–2005)

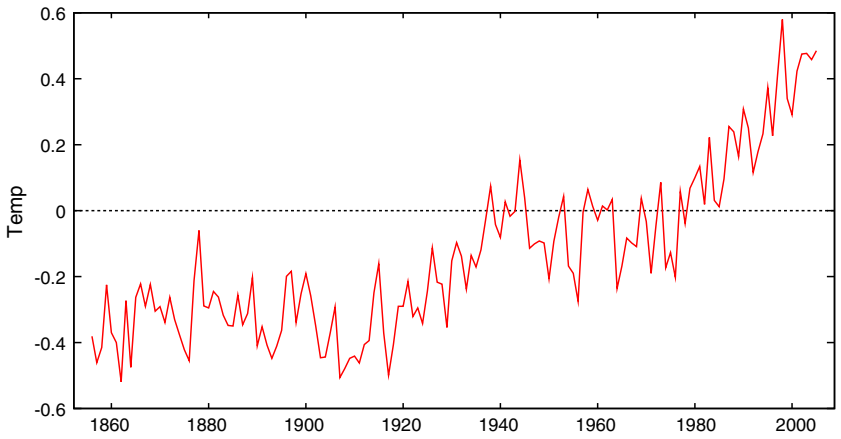

Fig. 2 Global surface temperature anomalies (1856–2005)

where $\min_v = -0.519$, $\max_v = 0.580$. In this way the data are normalized within the scale of 0 to 1.

In the sequel we will use the notations $(n_x;n)$ and $(n_x, n_y;n)$ to denote, respectively, the MLF neural network with inputs $x_{t-1}, ..., x_{t-n_x}$ and with n neurons in the hidden layer and the MLF neural network with inputs $x_{t-1}, ..., x_{t-n_x}, y_{t-1}, ..., y_{t-n_y}$ and with n neurons in the hidden layer. The number of lagged observations used (n_x and n_y) and number of nodes (n) in the hidden layer must be chosen. Although there are many rules of thumb proposed, none of them can be universally applied. Guidelines are either heuristic or obtained from limited empirical studies.

In this paper we employ the following procedure to determine the network structures and estimate their unknown parameters.

The T observations of our sample are divided into a training part and a test part. The test part is the last P-annual observations while the training sample consists of all previous $R = T - P$ observations. Often, the testing set, ranges in size from 10% to 30% of the training set. Thus we have posed $P = 30$. The weights to be used in the MLF neural networks are estimated from the data by minimizing the sum of squares

$$E = \sum_{t=1}^{R} (x_t - \hat{x}_t)^2$$

where $\hat{x}_t$ is the output of the network. The Levenberg-Marquardt algorithm is used in order to obtain a numerical solution for this minimization problem. We consider 100, 500 and 1000 epochs. To avoid getting stuck in local minima, we adopt the common practice of multiple starts in neural network training. In particular, based on the training subsample we train each neural network 150 times using 150 different sets of initial random weights. In our simulation we use an empirical method in order to prevent the overfitting. We estimate the neural network's weights and then we observe the MSE on the test set. When this MSE is very large with respect to the MSE of the ARIMA (3,1,0) model we discard the neural network. We choose among the selected neural network models the best. We prefer using this approach and not, as usual, an early stopping method, because our neural network models are compared with an ARIMA(3,1,0) model that is estimated without consider the validation set.

Let us consider the univariate model $(n_x;n)$. We have identified an ARIMA(3,1,0) model for the series x_t. Since feed-forward neural networks per se are generalized non-linear AR models, the choice $n_x = 4$ seems reasonable (see Chatfield and Faraway 1998). The number of hidden nodes (n) is another important factor for the neural network model. Hidden nodes play a major role for the non-linear modeling of neural networks. Since there is no systematic method to determine this parameter, it is varied with 4 levels ranging from 2 to 5. In this way we obtain five MLF neural networks. The training sample is used to estimate the parameters for any specific model architecture. Then the out-of-sample testing data are used to select the best model among all models considered. We use the mean squared error (MSE) as performance measures. The best univariate model is the one which gives the least MSE in the test set.

The bivariate neural network models, $(n_x, n_y;n)$, $n_y = 1, 2, 3, 4$, are obtained by adding to the architecture of the univariate models the input nodes: $\{y_{t-1}\}$, $\{y_{t-1}, y_{t-2}\}$, $\{y_{t-1}, y_{t-2}, y_{t-3}\}$ and $\{y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}\}$, respectively. In order to consider the possibility of a linear relationship, we estimate also the first equation of a VAR model for $(\Delta x_t, \Delta^2 y_t)$. Again, the best bivariate model in the class of linear and non-linear models is the one which gives the best results in the test set.

3 Testing for non-causuality

We are interested in testing the null hypothesis that y does not Granger cause x. In order to do this we compare one-step-ahead predictive ability of the following competing models:

$$\hat{x}_t = \hat{\phi}_0 + \left(1 + \hat{\phi}_1\right) x_{t-1} + \left(\hat{\phi}_2 - \hat{\phi}_1\right) x_{t-2} + \left(\hat{\phi}_3 - \hat{\phi}_2\right) x_{t-3} - \hat{\phi}_3 x_{t-4} \quad (1)$$

$$\hat{x}_t = f\left(\sum_{i=1}^{4} \hat{w}_i f\left(\sum_{j=1}^{4} \hat{w}_{ij}^{x} x_{t-j} + \sum_{j=1}^{3} \hat{w}_{ij}^{y} y_{t-1} + \hat{\theta}_i\right) + \hat{\theta}\right) \quad (2)$$

The model (1) is an ARIMA(3,1,0). This model is the best in the class of linear (ARIMA) and non-linear (Neural Networks) univariate models (see Table 1).

The model (2) is the best in the class of linear (VAR) and non-linear (Neural Networks) bivariare models (see Table 2). Following the approch described in Pasini et al. (2001) and Pasini and Ameli (2003), we analyze the robustness of this model . We perform ten runs of the neural model NN(4, 3; 4) with distinct initial random weights. The MSE out of sample is calculated for every run. We obtain the mean ($\hat{\mu}$) and the standard deviation ($\hat{\sigma}$) of these values. The spread of the results of our runs is evaluated considering the interval $\hat{\mu} \pm 2\hat{\sigma}$ (see Table 3). It is important to note that the MSE out of sample of ARIMA model (0.008063) belongs to the interval [0.005628, 0.011464]. This means that certain initial random weights could provide an estimated neural model NN(4, 3; 4) with a MSE out of sample not significantly different from 0.008063. Thus a possible Granger causality from radiative forcing due to CO_2 concentration to global surface temperature anomalies could not emerge.

For y to be Granger-causal for x we require that the out-of-sample forecasting performance of Eq. 2 including y is better than Eq. 1 excluding y. The criterion for forecasting performance is the mean squared error of the forecast. This approach is close to Granger's original idea of causality testing (see Ashley et al. 1980), and avoids problems of within-sample linear tests.

We estimate Eqs. 1 and 2 using observations $t = 1, 2, ..., R$, and computes

$$\hat{e}_{R+k}^{(x)} = x_{R+k} - \hat{\phi}_0 - \left(1 + \hat{\phi}_1\right) x_{R+k-1} - \left(\hat{\phi}_2 - \hat{\phi}_1\right) x_{R+k-2} - \left(\hat{\phi}_3 - \hat{\phi}_2\right) x_{R+k-3} - \hat{\phi}_3 x_{R+k-4}$$

and

$$\hat{e}_{R+k}^{(xy)} = x_{R+k} - f\left(\sum_{i=1}^{4} \hat{w}_i f\left(\sum_{j=1}^{4} \hat{w}_{ij}^{x} x_{R+k-j} + \sum_{j=1}^{3} \hat{w}_{ij}^{y} y_{R+k-1} + \hat{\theta}_i\right) + \hat{\theta}\right)$$

for $k = 1, ..., P$.

In this way we obtain two sequences of P one-step-ahead forecast errors, with $P + R = T$, where T is the sample size. Finally, we calculate the mean square prediction errors ($MSPE$) defined as:

$$MSPE(ARIMA(3, 1, 0)) = \frac{1}{P} \sum_{k=1}^{P} \left(\hat{e}_{R+k}^{(x)}\right)^2$$

$$MSPE(4, 3; 4) = \frac{1}{P} \sum_{k=1}^{P} \left(\hat{e}_{R+k}^{(xy)}\right)^2$$

If

$$MSPE(4, 3; 4) < MSPE(ARIMA(3, 1, 0)),$$

then y Granger causes x. We obtained $MSPE$(ARIMA (3,1,0)) = 0.008063 and $MSPE$(4,3;4) = 0.006481, thus $MSPE$(4,3;4) < $MSPE$(ARIMA(3,1,0)). The models give different MSP errors. Is this difference statistically significant? The encompassing test ENC-T (see Clark and McCracken 2001; Harvey et al. 1997) suggests a statistically significant better performance of model (4,3;4) over model ARIMA(3,1,0). In fact, using

Table 1 Performance of univariate models

Model	MSE in-sample	MSE out-of-sample
ARIMA(3,1,0)	0.006471	0.008063
(4;2)	0.005313	0.015721
(4;3)	0.005295	0.012293
(4;4)	0.006812	0.010464
(4;5)	0.005616	0.016230

Table 2 Performance of bivariate models

Model	MSE in-sample	MSE out-of-sample
VAR(1)	0.007812	0.009288
(4,2;2)	0.005904	0.007950
(4,2;3)	0.006064	0.006880
(4,3;4)	0.006104	0.006481
(4,2;5)	0.006299	0.008299

Table 3 Robustness of NN(4,3;4) model

Model	$\hat{\mu}$	$\hat{\sigma}$	MSE out-of-sample
(4,3;4)	0.008546	0.001459	[0.005628, 0.011464]

a 5% significance level, the value of test statistics, ENC-T = 2.901887, implies the rejection of the null hypothesis of equal forecast accuracy. In sum, it can be stated, in the light of this non-linear out-of-sample Granger causality test, that CO_2 has significant non-linear explanatory power for global temperature.

4 Conclusion

In this study we have used non-fully traditional (for the climatology) method to examine the non-linear interaction between CO_2 concentration and global temperature anomalies: the neural networks. In particular we have combined these models with the concept of Granger causality. Results from non-linear Granger causality analysis are consistent with the hypothesis that CO_2 concentration causes global temperatures. These results reinforce our confidence in the role of anthropogenic forcing given to the recent global warming.

Acknowledgement We are indebted to an anonymous referee for his valuable comments on our previous draft. Any remaining lack of clarity are our own responsibility.

References

Ashley R, Granger CWJ, Schmalensee R (1980) Advertising and aggregate consumption: an analysis of causality. Econometrica 48(5):1149–1167

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Chatfield C, Faraway J (1998) Time series forecasting with neural networks: a comparative study using the airline data. Appl Stat 47:231–250

Clark TE, McCracken MW (2001) Tests of equal forecast accuracy and encompassing for nested models. J Econom 105: 85–110

Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Mathematical Control Signals Systems 2:303–314

Granger CWJ (1969) Investigating causal relations by econometric methods and cross-spectral methods. Econometrica 34:424–438

Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13(2): 281–291

Myhre G, Highwood EJ, Shine K, Stordal F (1998) New estimates of radiative forcing due to well mixed greenhouse gases. Geophys Res Lett 25:2715–2718

Pasini A (2009) In: Haupt SE, Pasini A, Marzban C (eds.) Artificial intelligence methods in the environmental sciences. Springer, New York, pp 235–254

Pasini A, Ameli F (2003) Radon short range forecasting through time series preprocessing and neural network modeling. Geophys Res Lett 30(7):1386. doi:10.1029/2002GL016726

Pasini A, Pelino V, Potestà S (2001) A neural network model for visibility nowcasting from surface observations: results and sensitivity to physical input variables. J Geophys Res 106(D14):14951–14959

Pasini A, Lorè M, Ameli F (2006) Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system. Ecol Model 191: 58–67

Triacca U (2005) Is Granger causality analysis appropriate to investigate the relationship between atmospheric concentration of carbon dioxide and global surface air temperature? Theor Appl Climatol 81:133–135

Yi J, Prybutok R (1996) A neural network model forecasting for prediction of daily maximum ozone concentation in a industrialized urban area. Environ Pollut 92(3): 349–357

Wu B (1995) Model-free forecasting for non-linear time series (with application to exchange rates). Comput Stat Data Anal 19:433–59