Global Warming Is Driven by Anthropogenic Emissions: A Time Series Analysis Approach

Pablo F. Verdes

Heidelberg Academy of Sciences, c/o Institute of Environmental Physics, Im Neuenheimer Feld 229, D-69120 Heidelberg, Germany

(Received 12 May 2006; published 24 July 2007)

The solar influence on global climate is nonstationary. Processes such as the Schwabe and Gleissberg cycles of the Sun, or the many intrinsic atmospheric oscillation modes, yield a complex pattern of interaction with multiple time scales. In addition, emissions of greenhouse gases, aerosols, or volcanic dust perturb the dynamics of this coupled system to different and still uncertain extents. Here we show, using two independent driving force reconstruction techniques, that the combined effect of greenhouse gases and aerosol emissions has been the main external driver of global climate during the past decades.

DOI: 10.1103/PhysRevLett.99.048501 PACS numbers: 92.40.Aa, 05.45.Tp, 92.60.Ry, 92.70.−j

Global climate has changed rapidly over recent decades, as reflected by patterns of tropical circulation, cyclone intensity, and other climatological parameters [1]. In particular, global temperatures have increased by $(0.6 \pm 0.2)\,^\circ$C since 1860. A question that has become increasingly crucial is to determine the responsibility of several potentially explaining factors, most notably the emission of heat-trapping greenhouse gases (GHGs), the direct and indirect cooling effect of aerosols, the Sun's increased radiative output, the lack of volcanic activity, etc. [2,3]. On one hand, the thesis of warming by increased GHGs is supported by extensive simulations with atmosphere-ocean general circulation models (GCMs) [4–6]. On the other hand, a number of paleoclimatic reconstructions of global temperatures show a conspicuous correlation to the secular behavior of solar irradiance during the Holocene [7–9]. Furthermore, recent empirical studies claim that the physics of global response to solar variations is misrepresented in current theoretical models [10], making this subject one of the most actively debated in the recent literature.

The intellectually most satisfactory analysis of any dynamical system, and, in particular, of the climate system, corresponds to the case in which its underlying mechanisms can be modeled from first principles. As such, the GCM approach is preferable in that it allows a deeper understanding of causes and effects. However, the conclusions drawn from GCM simulations will be continuously revised insofar as these models are refined and their performances improved. This constitutes a formidable task, for climate dynamics is extremely complex and not fully understood: Among other delicate issues, models of the atmosphere must account for its couplings with the land, oceans, and associated nonlinear feedbacks, together with their chemical responses to increased radiation at all wavelengths [11]. For such complex systems, an alternative modeling path is given by data-driven (statistical) techniques. According to this philosophy, the evolution of the system is studied by recording time series, which are sets of observations x_t collected at regular intervals of time. In this framework, the modeling task consists in estimating the function f that is supposed to generate the data via the equation $x_{t+\tau} = f(\mathbf{x}_t)$, where $\mathbf{x}_t = (x_t, x_{t-\tau}, \ldots, x_{t-(d-1)\tau})$ for some properly chosen time delay τ and embedding dimension d [12] (the state space may include delayed values of x_t only or, in the multivariate case, further available magnitudes as well). The standard approach is to choose a nonlinear model family whose free parameters typically carry no meaningful physical interpretation but instead provide a suitable (flexible) functional form. In a second step, this model f is fit to the observed data according to a maximum-likelihood or least-squares procedure, a strategy that then allows us to make concrete predictions on the system under study. Furthermore, once such an empirical model has been built, the issue of whether the dynamics is externally perturbed can be assessed and the profile of the forcing agent can be accurately estimated, as we summarize in the following.

We first describe the specific time-series data and framework considered. We have focused on the history of the past 150 years of (i) instrumental temperature anomalies (T) from the current Intergovernmental Panel on Climate Change (IPCC) global time series [13], (ii) the solar total irradiance (STI) reconstruction by Lean [14], and (iii) the index of volcanic activity (V) as from Ref. [5]. These magnitudes, depicted in Figs. 1(a)–1(c), represent the main natural (as opposed to anthropogenic) components of the global warming problem. Closing our system of interest to encompass only the interaction among them, we have then considered the question of whether this system is autonomous or externally driven. To this end, we have resorted to the theory of nonlinear time-series analysis and applied two independent state-of-the-art data-driven techniques for driving force reconstruction [15–18]. To describe the general frame more precisely, suppose we have recorded measurements from a forced dynamical system. Its nonstationarity can be parametrized by a scalar α in a dynamics of the form $x_{t+1} = f(\mathbf{x}_t, \alpha_t)$. The evolution of α_t accounts for the effects of either an external driving force or other internal degrees of freedom varying on large time scales. It is a remarkable result from

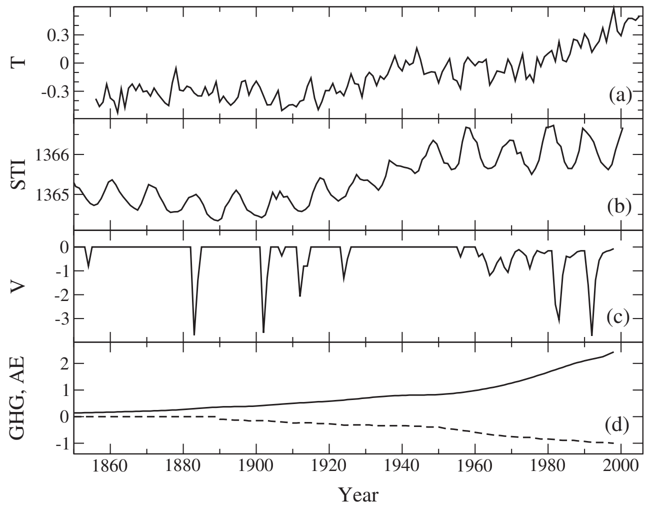

FIG. 1. Time-series data employed in this study. Yearly averages of (a) temperature anomalies (°C), (b) solar total irradiance (W/m^2), (c) equivalent radiative forcing for ice core volcanism (W/m^2), and (d) anthropogenic components (W/m^2). Solid line: Equivalent radiative forcing for CO_2 and other well-mixed trace gases (methane, nitrous oxides, and chlorofluorocarbons). Dashed line: Direct radiative forcing effect of tropospheric aerosols (no cloud feedback).

the theory of nonlinear time-series analysis that it is possible to reconstruct the variation of a slow driving force solely from the observations x_t, even without any prior knowledge on the internal dynamics (more precisely, it was shown in Refs. [17,18] that α and f can be estimated simultaneously if they develop on different time scales). In our particular case, we have represented the dynamics according to $T_t = f(\mathbf{x}_t, \alpha_t)$, with $\mathbf{x}_t = (\langle \text{STI} \rangle_t, \langle \text{STI} \rangle_{t-5}, \langle V \rangle_t)$, where a bracket $\langle \, \rangle_t$ indicates an average of the enclosed variable over the interval $(t - 5, t]$ to account for possible delayed effects [19].

The first methodology that we have employed for the reconstruction of α_t can be summarized as follows: (i) Consider initially that the system is stationary, and set $\alpha_t = 0 \ \forall \ t$. Construct from the original records the data set $\mathcal{D}$ whose elements are patterns $(T_t, \mathbf{x}_t, \alpha_t = 0)$, and randomly split it in disjoint learning $\mathcal{L}$ and validation $\mathcal{V}$ sets. (ii) Choose *any* flexible functional form for f and use *any* data-driven fitting approach to build from these data a global stationary model $T = f(\mathbf{x}, \alpha = 0)$. (iii) Estimate the smooth α profile that maximally improves the performance of this stationary model by readjusting its internal parameters *and* α_t to minimize the prediction error

$$E_{\mathcal{L}} = \sum_{i \in \mathcal{L}} [T_i - f(\mathbf{x}_i, \alpha_i)]^2 + \lambda \sum_{i \in \mathcal{L}} (\alpha_{i+1} - \alpha_i)^2.$$

Here the first term is the standard mean square error of f, and the second term enforces the basic assumption of a smooth α behavior. (The hyperparameter λ fixes the relative weight of both terms and can be determined as indicated below.) (iv) Stop the simultaneous fitting process for f and α at the minimum of $E_{\mathcal{V}}$, which corresponds to the optimal model's generalization capability. (v) For the determination of λ, repeat the above steps for different values of this hyperparameter and choose $\lambda_{\text{opt}} = \arg \min_\lambda E_{\mathcal{V}}^{\min}(\lambda)$. For a more detailed discussion of this algorithm and an assessment of its very good performance for driving force estimation, please refer to Ref. [17].

In Fig. 2, we show how the considered dynamical system has been perturbed during the past 150 years according to this driving force reconstruction methodology. We have employed feedforward artificial neural networks (ANNs) with a 4:10:1 architecture (3 + 1 input, 10 hidden, and 1 output units), with sigmoidal activation functions and trained according to the standard backpropagation rule. The reported results are robust with respect to architecture variations (number of hidden units) and average 20 independent realizations of this algorithm. Notice that, since the methodology formulation is univariate in the forcing, then multiple driving factors, if present, will be represented by a single scalar. In such a case, α_t must be interpreted as an *effective* univariate representation of these external drivers. We have therefore constructed a combined record that accounts for the positive GHG and negative aerosol radiative forcing effects [20]. These individual time series are taken from Ref. [5] and depicted in Fig. 1(d); their sum is plotted in Fig. 2 for comparison against the reconstructed driving force. The coincidence is remarkable: The estimated perturbation shadows, within error bands, the evolution of this combined profile. In particular, the steep rise of the most recent part of this record is well reproduced and shows that gases originated in anthropogenic activities are

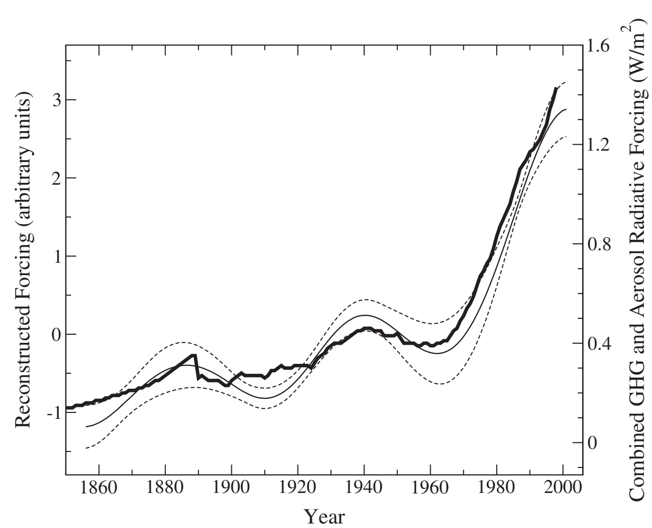

FIG. 2. Reconstruction of external forcing (thin line) using the first estimation methodology. Dashed lines embrace 1 standard deviation. For comparison, the evolution of combined GHG and aerosol radiative forcing levels is indicated with a thick line.

responsible for affecting the dynamics of global climate during the past decades.

As a test of this conclusion, we have also employed a second, independent methodology for driving force reconstruction [18]. In the following, we summarize its essential aspects. If we split the available recordings into $N_{\rm int}$ intervals of M points each, then we can assume that α is constant within each interval and write

$$T_t^{(m)} \simeq f(\mathbf{x}_t^{(m)}, \alpha^{(m)}),$$

where $\alpha^{(m)}$ is the mean value of the driving force in the mth interval considered, and now t runs only on points of this interval. For a smooth dependence of f with α, we can employ a first-order expansion on its second argument centered in $\alpha^{(k)}$. Rearranging terms, we obtain

$$T_t^{(m)} - f(\mathbf{x}_t^{(m)}, \alpha^{(k)}) \simeq \frac{\partial f}{\partial \alpha}(\mathbf{x}_t^{(m)}, \alpha^{(k)})[\alpha^{(m)} - \alpha^{(k)}],$$

which should be valid for intervals k and m close enough (in practice, they can be taken with a substantial overlap to fulfill this condition). Some simplifications transform these equations into a system of the form

$$E_k^{k+1} = A^{k+1}\Delta\alpha_k^{k+1}, \qquad (1)$$

where E_k^{k+1} are cross-prediction errors, the coefficients $A^{k+1} = \langle \frac{\partial f}{\partial \alpha}(\mathbf{x}_t^{(k+1)}, \alpha^{(k+1)})\rangle_t$ are unknown, and $\Delta\alpha_k^{k+1} = \alpha^{(k+1)} - \alpha^{(k)}$ are the quantities of interest. Estimates of E_k^{k+1} can be obtained by using any data-driven modeling approach, and the system (1) can be solved up to the scale factor A^1 and offset $\alpha^{(1)}$. According to the considerations above, the steps to reconstruct α are (i) choose a data segment k containing M iterates and use them to build a model $f(\bullet, \alpha^{(k)})$. (ii) Compute $E_k^{k\pm1}$, the average prediction errors on the first-neighboring intervals. (iii) Repeat the previous steps for all possible k. Finally, (iv) reconstruct the driving parameter by solving system (1). It has been previously shown that this algorithm is able to produce accurate estimates of the driving force profile—for a more detailed discussion, please refer to Refs. [18,21].

In Fig. 3, we show the external force that, according to this estimation method, has perturbed global climate during the time span under study. In this case, we have considered $N_{\rm int} = 117$ intervals of $M = 30$ points each (97% of overlap), and within them we have built local models of the dynamics using radial basis functions (RBFs) with 3 centers. The reported results are again an average of 20 independent realizations of this driving force reconstruction algorithm and have been smoothed with a 10-point sliding window. The obtained profile is robust with respect to variations in the intervals' length and overlap, but the number of centers employed in the RBF modeling must be kept low to avoid overfitting the reduced

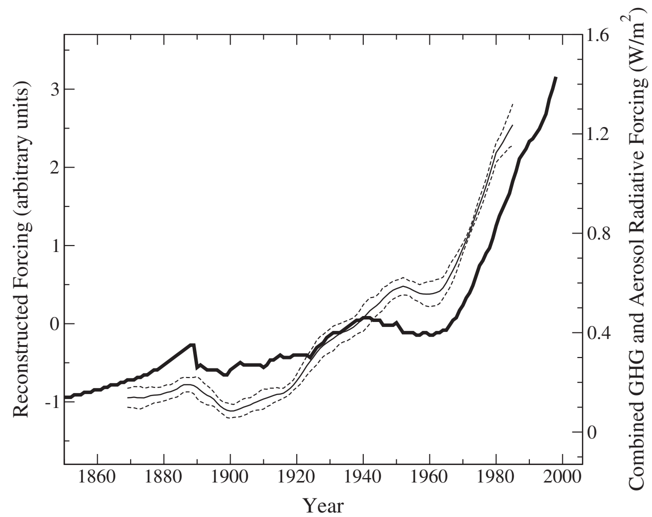

FIG. 3. Estimated external driving force (thin line) acting on global climate using the second reconstruction methodology. The profile of combined GHG and aerosol radiative forcing effects is again represented with a thick line.

databases. In addition, similar results are obtained when using ANNs instead of RBFs. As Fig. 3 indicates, the reconstructed driving force again follows the combined GHG and aerosol radiative forcing profile, although not as closely as in Fig. 2. As shown in Ref. [17], the first methodology produces more accurate estimations than this second one, especially towards the limit of short observational records. The advantage there is the use of a *global* model to reconstruct α, meaning that all available data are employed simultaneously, while the present method is based on developing a series of *local* models on reduced databases. Taken together, these results suggest a convergence to the evolution of combined GHG and aerosol radiative effects as the true agent perturbing global temperature dynamics.

Finally, a simpler attribution analysis can be made along the following lines: Consider the point-by-point prediction errors made by a model f of measured temperature anomalies T_t:

$$T_t - f(\mathbf{x}_t). \qquad (2)$$

(Notice that we have not chosen the usual quadratic error. The reason will be clear below.) If model f is forced by α, these errors can be approximated, to first order in α around its mean value $\bar{\alpha}$, by

$$\frac{\partial f}{\partial \alpha}(\mathbf{x}_t, \bar{\alpha})[\alpha_t - \bar{\alpha}].$$

For strongly nonlinear functions f, the above partial derivative prefactor can largely fluctuate. However, with the hypothesis of a mild variation of $\partial f/\partial\alpha$, errors (2) can be then approximated as roughly proportional to the oscillations of α_t around its mean value. In Fig. 4, we show how

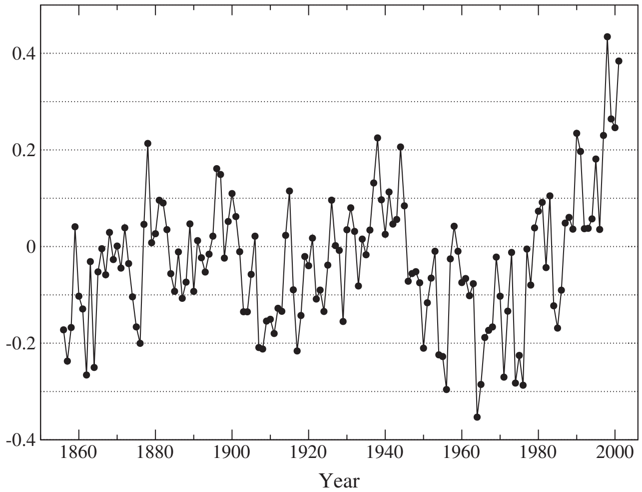

FIG. 4. Point-by-point prediction errors of a 3:10:1 ANN model of temperature anomalies as a function of time.

the prediction errors of a 3:10:1 ANN model of $T_t = f(\langle \mathrm{STI} \rangle_t, \langle \mathrm{STI} \rangle_{t-5}, \langle V \rangle_t)$ are distributed [22]. As this figure illustrates, there is a final rise in this profile that supports the contention that anthropogenic emissions have been the main driver of global climate during the past decades. As a note of caution, it must be emphasized that this linear approximation of point prediction errors represents only a reduced analysis that here allows for a more intuitive approach to the problem under study. Although this strategy can be very useful by quickly giving a hint on the driving force profile, the obtained portrait will be distorted in the general nonlinear case. There, the global or local methodologies described above should be employed as they properly account for varying partial derivatives of the dynamics f.

Summing up, in this work we have taken advantage of a remarkable result from the theory of nonlinear time-series analysis, namely, that it is in practice possible to estimate the variation of external driving forces acting on complex systems even when their internal mechanisms are unknown, to study the attribution problem in climate change from a different perspective. Using two independent methodologies for accurate driving force reconstruction and different data-driven modeling tools, we have presented evidence that the forcing agent on global climate dynamics can be consistently identified with the combined effect of anthropogenic emissions. The present study is particularly interesting in that it represents a new approach to this important problem. Furthermore, it is purely data-driven and thus avoids any possibly unaccounted mechanisms of first-principles general circulation modeling. We believe this is a distinctive aspect that can enrich the continuing debate on the future of our climate.

We thank P. Grassberger, H. Abarbanel, G. Schmidt, T. Crowley, and D. Pierce for valuable comments. This work has been supported by the Alexander von Humboldt Foundation.

[1] *IPCC Third Assessment Report—Climate Change 2001: The Scientific Basis*, edited by J.T. Houghton *et al.* (Cambridge University Press, Cambridge, England, 2001).
[2] J.L. Lean, J. Beer, and R.S. Bradley, Geophys. Res. Lett. **22**, 3195 (1995).
[3] A. Robertson *et al.*, J. Geophys. Res. **106**, 14 783 (2001).
[4] D.T. Shindell, R.L. Miller, G.A. Schmidt, and L. Pandolfo, Nature (London) **399**, 452 (1999).
[5] T. Crowley, Science **289**, 270 (2000).
[6] P.A. Stott *et al.*, Science **290**, 2133 (2000).
[7] U. Neff *et al.*, Nature (London) **411**, 290 (2001).
[8] G. Bond *et al.*, Science **294**, 2130 (2001).
[9] D.V. Hoyt and K.H. Schatten, *The Role of the Sun in Climate Change* (Oxford University Press, New York, 1997).
[10] N. Scafetta and B. J. West, Geophys. Res. Lett. **33**, L05708 (2006).
[11] J.E. Hansen *et al.*, in *Climate Modeling: Past, Present and Future*, edited by D. Randall (Academic, San Diego, 2000), p. 127.
[12] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series Vol. 7 (Cambridge University Press, Cambridge, England, 1997).
[13] C.K. Folland *et al.*, Geophys. Res. Lett. **28**, 2621 (2001).
[14] J.L. Lean, Geophys. Res. Lett. **27**, 2425 (2000).
[15] M.I. Széliga, P.F. Verdes, P.M. Granitto, and H.A. Ceccatto, Physica (Amsterdam) **327A**, 190 (2003).
[16] M.I. Széliga, P.F. Verdes, P.M. Granitto, and H.A. Ceccatto, Int. J. Neural Syst. **13**, 103 (2003).
[17] P.F. Verdes, P.M. Granitto, and H.A. Ceccatto, Phys. Rev. Lett. **96**, 118701 (2006).
[18] P.F. Verdes, P.M. Granitto, H.D. Navone, and H.A. Ceccatto, Phys. Rev. Lett. **87**, 124101 (2001).
[19] The two methodologies that we have employed for nonstationarity characterization are not restricted to the case of dynamical systems: Their theoretic derivations also hold, without modifications, in the more general case of regressions $y = f(x)$ that change over time.
[20] Following [5], we have not included estimates of indirect aerosol radiative forcing because evaluations by the IPCC indicate that the confidence in them is very low.
[21] P.F. Verdes, P.M. Granitto, and H.A. Ceccatto, Sol. Phys. **221**, 167 (2004).
[22] To illustrate how this modeling approach is relevant to the global warming problem, we have repeated 200 times the following experiment: We have randomly selected 10% of the available data for testing purposes. The remaining record was randomly split in fractions of (i) 80% used to iteratively adjust the internal parameters of these ANN models and (ii) 10% employed for validation. Simultaneously, linear models were adjusted and tested on the same sets. The variances explained by linear and ANN models are 0.61 ± 0.15 and 0.80 ± 0.10, respectively. This 30% improvement shows that ANN models can be advantageously used, replacing standard linear techniques.