Physica A 509 (2018) 121–139

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Predicting global temperature anomaly: A definitive investigation using an ensemble of twelve competing forecasting models☆

Hossein Hassani [a], Emmanuel Sirimal Silva [b,*], Rangan Gupta [c], Sonali Das [d,e,f]

[a] Research Institute of Energy Management and Planning, University of Tehran, Iran
[b] Fashion Business School, London College of Fashion, University of the Arts London, UK
[c] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa
[d] Advanced Mathematical Modelling, Modelling and Digital Science, Council for Scientific and Industrial Research, P.O. Box 395, Pretoria, 0001, South Africa
[e] The School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, 2000, South Africa
[f] Department of Statistics, Nelson Mandela University, Port Elizabeth, 6031, South Africa

HIGHLIGHTS

- A comprehensive comparison of models for forecasting Global Temperature (GT).
- Introductory application of SSA and Multivariate SSA (MSSA) for forecasting GT.
- MSSA outperforms competing models with statistically significant results.
- Conclusive evidence that CO_2 can predict GT.
- The MSSA models report the best direction of change predictions when forecasting GT.
- It is important to model nonlinearity into the relationship between GT and CO_2.

ARTICLE INFO

Article history:
Received 14 January 2018
Received in revised form 13 April 2018
Available online 15 June 2018

JEL classification:
C22
C32
C53
Q53
Q54

Keywords:
CO_2 emissions
Forecasting
Global temperature anomaly
Univariate and multivariate models

ABSTRACT

In this paper we analyse whether (anthropometric) CO_2 can forecast global temperature anomaly (GT) over an annual out-of-sample period of 1907–2012, which corresponds to an initial in-sample of 1880–1906. For our purpose, we use 12 parametric and nonparametric univariate (of GT only) and multivariate (including both GT and CO_2) models. Our results show that the Horizontal Multivariate Singular Spectral Analysis (HMSSA) techniques (both Recurrent (-R) and Vector (-V)) consistently outperform the other competing models. More importantly, from the performance of the HMSSA-V model we find conclusive evidence that CO_2 can forecast GT, and also predict its direction of change. Our results highlight the superiority of the nonparametric approach of SSA, which in turn, allows us to handle any statistical process: linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian.

☆ We would like to thank five anonymous referees for many helpful comments. However, any remaining errors are solely ours.
* Corresponding author.
E-mail addresses: hassani.stat@gmail.com (H. Hassani), e.silva@fashion.arts.ac.uk (E.S. Silva), rangan.gupta@up.ac.za (R. Gupta), sdas@csir.co.za (S. Das).

https://doi.org/10.1016/j.physa.2018.05.147
0378-4371/

1. Introduction

The climate change debate consists of players who, in the absence of accessible evidence-based and objective information, may resort to decisions based on perceptions and even possibly political agendas [1]. The debate gets even more aggressive when global warming is discussed in relation to anthropometric carbon dioxide (CO_2) emissions [2–5].

Global warming has been accepted as a reality. For example, see [6] and more recently, Dergiades et al. [7] for a detailed discussion based on the time-series relationship between global temperatures and emissions. The purpose of this paper is to undertake a rigorous investigation of well-established datasets for global temperature (GT) and CO_2 using a suite of forecasting models in an attempt to identify, possibly, a single model that can be prescribed for forecasting GT. Specifically, we consider 12 time-series models for forecasting GT from both parametric and nonparametric paradigms. These 12 models include 7 univariate models and 5 multivariate models namely: Random Walk (RW), Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS), Neural Networks (NN), Fractionalized ARIMA (ARFIMA), Box–Cox transformation, ARMA errors, Trend and Seasonal components (BATS), Bayesian Autoregression (BAR), Vector Autoregression (VAR), Bayesian Vector Autoregression (BVAR), Singular Spectrum Analysis (SSA) and Multivariate Singular Spectrum Analysis (MSSA). Note that, our primary focus in this paper is forecasting GT using the information contained in CO_2 using the SSA approach, due to its modelling flexibility advantages discussed in the next segment. However, given that a forecasting exercise requires competing models, we choose a wide array of linear and nonlinear versions of univariate and multivariate models most commonly used in the forecasting literature.

Our sample covers the entire period of 1880–2012, with an out-of-sample period of 1907–2012 (based on an initial in-sample period of 1880 to 1906 — note that this initial in-sample period is incremented by one year and models are re-estimated except in the case of RW and SSA models). The start and end points of the analysis are purely driven by data availability at the time of writing this paper, but the choice of the out-of-sample period is determined by the earliest possible break date detected (based on the Bai and Perron [8] tests of multiple structural breaks) in the relationship between GT and CO_2, which happened to be 1906. Since, we estimate our linear models recursively over the out-of-sample period, we are able to accommodate the change in the parameter estimates of the model while producing our forecasts. This approach makes the linear and nonlinear approaches comparable, with both types of model being now able to correct misspecification due to breaks either by structural design or through recursive estimation.

This paper makes several important contributions: (i) this paper makes the first attempt to provide a comprehensive comparison of models for forecasting GT available in the literature; (ii) it marks the introductory application of models such as BATS, SSA and MSSA for forecasting GT; and, (iii) finally, we use a new and automated MSSA forecasting algorithm for generating out-of-sample forecasts for GT. The SSA and MSSA models contribute to the literature on forecasting GT by decomposing GT for denoising and signal extraction prior to forecasting, in addition to the multivariate analysis of both GT and CO_2 in the case of MSSA. Moreover, this automated MSSA forecasting algorithm is optimized based on minimizing a loss function. As such, it ensures the overall MSSA process is less labour intensive as it does away with the need for having to analyse eigenvectors individually in order to differentiate between signal and noise. Also, it can show the best MSSA model for forecasting any given data set in-sample and can therefore be used in future forecasting studies as well. To the best of our knowledge, barring the two studies of Fildes and Kourentzes [3], and McMillan and Wohar [4], all other papers in this area involving time series analysis have analysed these two variables separately in univariate settings.[1]

Our results show that the nonparametric approach of SSA consistently outperform the other competing models. More importantly, from the performance of the SSA models, we find conclusive evidence that CO_2 can forecast GT, and also predict its direction of change. The remainder of the paper is organized as follows: in Section 2, we discuss the literature review, while Section 3 presents a detailed description of the 12 forecasting models investigated with extra emphasis on the SSA and MSSA techniques. In Section 4, we outline the datasets used and the metrics used for the evaluation of the models, with estimations conducted in either R or RATS. In Section 5, we present an in-depth analyses of the results in our quest to identify the best model for our purpose, while in Section 6, we present a discussion. Finally, we present some concluding remarks on the investigation undertaken in Section 7.

2. Literature review

Note, the relationship between CO_2 and GT is primarily analysed using Coupled Atmospheric-Ocean General Circulation Models (AOGCMs) of the United Nations Intergovernmental Panel on Climate Change (IPCC). While the AOGCM is the primary framework, smaller models focusing on certain aspects of the world's climate are also used. Besides the small- and large-scale AOGCMs, there is also a statistical approach, whereby models are estimated directly from observations. But the latter as indicated above, is limited.

The AOGCMs comprise of systems of partial differential equations based on the basic laws of physics, fluid motion and chemistry, aimed at capturing the dynamics of the atmosphere and oceans. Given a set of initial and boundary conditions (which include emissions of atmospheric gases such as CO_2 and volcanic eruptions), these models use a three-dimensional

[1] The reader is referred to McMillan and Wohar [4] for a detailed literature review in this regard, which primarily involves analysis of unit root and persistence properties of these two variables.

grid to simulate the dynamics and to produce time-varying outputs such as global temperature (and precipitation) [3].[2] The advantage of these types of models is that, if the model represents the physical laws accurately, it should be possible to simulate scenarios which have not been seen in the past [9]. However, there are serious limitations to this approach: (a) Attempting to describe the planet using a 200 km grid implies a limited ability to simulate the subgrid-scale dynamics of clouds, aerosols and turbulence adequately; (b) Climate modellers use parameterization techniques in an attempt to make up for the problems of resolution, but the varying approaches yield different results [10]; (c) Uncertainties also arise from an incomplete understanding of the interchange of CO_2 between the oceans, atmosphere and biosphere [9], and; (d) Uncertainty results from the judgemental approaches employed in constructing these models [11].

Given the above-mentioned issues, papers like that of Fildes and Kourentzes [3], and McMillan and Wohar [4], have pursued a statistical approach using historical observations, and provide a means of representing patterns in the data based on linear, nonlinear and nonparametric methods. The forecasting performances of statistical models obviously depend on the extent to which the future looks like the past. So, whereas AOGCMs are governed by laws of physics, fluid motion and chemistry, the statistical models are dependent on the quality of the data. Hence, there is an argument that statistical models have little value in forecasting global temperatures, because the rising CO_2 emissions mean that the future is unlikely to be similar to the past [9].

Instead of completely dismissing statistical approaches, [3], based on established encompassing tests, show that there are advantages of combining a specific AOGCM (namely United Kingdom Meteorological Office decadal climate prediction system, known as DePreSys, as developed by Smith et al. [12]) with linear and nonlinear time series models (namely, the combined forecast obtained from random walk; single, Holt and dampened trend exponential smoothings; autoregressive; univariate and multivariate (including CO_2 emissions) neural networks). While DePreSys provides superior short-term forecasting (one- to four-years-ahead), gains in forecast accuracy were achieved for up to 10-year forecasts using an univariate Holt exponential smoothing model and a multivariate neural network model which included CO_2 as a predictor variable.[3] McMillan and Wohar [4], based on VAR and Generalized Method of Moments (GMM) approaches conclude that CO_2 has a weak relationship with GT, both in terms of in-sample causality tests and an out-of-sample exercise of predicting the last observation in their sample. This could, however, be a result of uncaptured nonlinearity in the relationship between these two variables, which [3] emphasized upon, and which we show to exist in the discussion that follows in the results segment of the paper.[4]

Against this backdrop, in this paper we basically aim to understand better the underlying form of nonlinearity in the relationship between GT and CO_2 emissions by conducting an out-of-sample forecasting exercise, using univariate and multivariate versions of linear and nonlinear models. In the process, we look at a nonparametric approach using the SSA technique, over and above the neural network model. We also follow the recommendations of McSharry [9], who suggests that while the neural network approach is a flexible one, it is possible to improve the underlying nonlinearity by using alternative model specifications. Note that our approach is purely based on a time series analysis, and excludes climate models. Hence, as part of future research, as in [3], it would be interesting to pursue forecast combinations of the SSA models that we develop and AOGCMs, if and when forecast data from the latter set of models could be accessed. Note that the decision to rely on an out-of-sample rather than an in-sample predictability exercise to gauge the relationship between CO_2 and global temperature is motivated out of the belief that: "The ultimate test of any predictive model is its out-of-sample performance" [14]. Since we only use a part of the data to fit various models, and produce predictions for the remaining part of the sample period. Using full-sample information allows even linear models to pick up nonlinearity in the data. However, this is not the case, when we split the data into in- and out-of-sample periods, since good in-sample predictions does not guarantee the same for forecasting over an out-of-sample period. This is because, a particular model might be misspecified over the out-of-sample period due to limited in-sample information.

At this stage, it is important to emphasize the reasons to use the SSA technique in forecasting GT. SSA has recently evolved as a powerful technique in the field of time series analysis [15], besides the other standard forecasting approaches indicated above. SSA is a nonparametric technique that works with arbitrary statistical processes, whether linear or non-linear, stationary or non-stationary, Gaussian or non-Gaussian. Given that the dynamics of real time series, in our case GT, has usually gone through structural changes during the time period under consideration, one needs to make certain that the method of prediction is sensitive to the dynamical variations. Moreover, even though some might argue that

[2] It must be pointed out that, AOGCMs actually use CO_2 concentration as a forcing variable, which in turn, depends on the pattern of emissions, with the representative concentration pathways (RCPs) used for forecasts being based on different emission scenarios.

[3] Fildes and Kourentzes [3], also compared local temperature forecasts at six locations to show that the ten-year-ahead forecasts of the AOGCMs were worse than forecasts from a random walk benchmark, with the exponential smoothing providing the best results. In addition, at a very long-run horizon of twenty years ahead, the multivariate neural network with CO_2 emissions as an explanatory variable produced superior forecasts compared to other time series models. Finally, using the CO_2 concentration in the atmosphere, which is believed to be physically relevant for determining global temperatures, rather than CO_2 emissions, did not improve the forecasts.

[4] In a somewhat connected line of research, Kaufmann and Stern [13], and Beenstock et al. [5] tested the historic tracking capabilities of climate models. Climate modellers use in-sample correlations to confirm the ability of climate models to track the global temperature historically. But, a high correlation though necessary, is not a sufficient condition for such confirmation, given that global temperature tends to be nonstationary. In addition, the tracking errors must also be stationary. Using cointegration tests applied to hindcast data for global temperatures generated by multiple climate change models, both Kaufmann and Stern [13], and Beenstock et al. [5] found that global temperature and their hindcasts, generally fail to be cointegrated. This result meant that the climate models fail to track global temperature historically in the longer run, because their tracking errors are nonstationary.

the nonparametric nature of SSA/MSSA suggests little about the dynamic nature of the model, the exploitation of Takens' theorem [16] in the embedding approach helps overcome this issue. In particular, MSSA succeeds in providing insights into the dynamics of the underlying system via the decomposition of delay-coordinate phase space of given multivariate time series into a set of data-adaptive orthonormal components such as trend, oscillatory patterns and noise [17–19].

In addition, contrary to the standard methods of time series forecasting that assume normality and stationarity of the series (though the latter is not an issue for BVAR models), as a nonparametric method, SSA makes no prior assumptions about the data, with forecasts being obtained through bootstrapping. Furthermore, SSA decomposes a series into its component parts, and reconstructs the series by leaving out the random (noise) component. Clearly then, SSA is a much more general approach that allows us to handle issues of non-stationarity, non-normality, non-linearity, and even seasonality, though the latter is not an issue in our annual data set. In what follows, the predictive accuracy of all forecasts are evaluated via the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) loss functions, the direction of change criterion, and the Hassani–Silva (HS) test for predictive accuracy [20] and the modified Diebold–Mariano test [21], thus enabling a sound analysis and derivation of reliable conclusions.

3. Forecasting models

3.1. Random walk (RW)

The random walk model is used as a benchmark. This is because it is widely accepted that when introducing forecasting techniques for a particular purpose, it is vital that the introduced techniques are able to outperform the RW.[5] In brief, the RW model states that today's GT is the best forecast for tomorrow's GT.

3.2. Autoregressive integrated moving average (ARIMA)

In this paper we use an optimized version of ARIMA (auto-ARIMA), and provided through the forecast package[6] in *R*. A more detailed description of the algorithm underlying auto-ARIMA can be found in [22] whilst a summarized version is available in [23]. The process for obtaining point forecasts using the *R* software is concisely presented in [24]. The initial model structure and parameters for forecasting GT with ARIMA before re-estimation was *ARIMA*(0, 0, 1) with non-zero mean.

3.3. Exponential smoothing (ETS)

The ETS technique in the forecast package overcomes the limitations of the Makridakis et al. [25] algorithm pertaining to the calculation of prediction intervals. Whilst a detailed description of the ETS technique can be found in [24], in brief this algorithm considers the error, trend and seasonal components along with over 30 possible options for choosing the best exponential smoothing model via optimization of initial values and parameters using Maximum Likelihood Estimator and selecting the best model based on the Akaike information criterion (AIC). In the case of GT, the model structure and parameters for forecasting GT with ETS before re-estimation was an additive *ETS*(A, N, N) model which is equivalent to an *ARIMA*(0, 1, 1) model as linear ETS models are special cases of ARIMA model [24].[7]

3.4. Neural networks (NN)

The NN model used in this paper is popularly referred to as *nnetar* and is provided through the forecast package in *R*. For a detailed explanation on how the *nnetar* model operates, see [24]. It may be noted that in all cases the selected neural network model has only $k = 1$ hidden node, $p = 2$ lags, with the specifications being based on annual differences of the data. The neural network takes the form

$$\hat{y}_t = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \psi(x_t'.\hat{\gamma}_j), \tag{1}$$

where x_t consist of p lags of y_t and the function ψ has the logistic form

$$\psi(x_t'.\hat{\gamma}_j) = [1 + exp(-\hat{\gamma}_{j0} + \sum_{i=1}^{p} \hat{\gamma}_{ji}.y_{t-1})]^{-1} j = 1, \ldots, k \tag{2}$$

In the neural network literature, this form is often referred to as a one hidden layer feed forward neural network model. As can be seen, the nonlinearity arises through the lagged y_t entering in a flexible way through the logistic functions of

5 http://robjhyndman.com/hyndsight/benchmarks/.

6 It should be noted that the automated algorithms in the forecast package generates a new and improved re-estimated model each time a new observation is added to the data. As such the models reported for the forecast package are the initial ones.

7 Note that, the initial ETS (along with the RW) model is I(1), while the ARIMA model is I(0).

Eq. (1). The number of logistic functions included, namely k, is known as the number of hidden nodes. When analysing the NN forecasting results, it is important to bear in mind that the NN results reported here are not standard as there exists no standard formulation of a NN model.

3.5. Fractionalized ARIMA model (ARFIMA)

The ARFIMA algorithm used is also from the forecast package in *R*, and it automatically estimates and selects the p and q for an ARFIMA(p, d, q) model based on the Hyndman and Khandakar [22] algorithm whilst d and parameters are selected based on the Haslett and Raftery [26] algorithm. The ARFIMA model may be written as

$$(1-\sum_{i=1}^{p}\phi_i B^i)(1-B)^d y_t = (1+\sum_{i=1}^{q}\phi_i B_i)e_t, \tag{3}$$

where B the backshift operator, p is the order of the autoregressive part, q is the number of lagged forecast errors, and d is allowed to take the form of non-integer values. The ARFIMA model structure and parameters for forecasting GT (before re-estimation) were $p = 2$, $d = 0$, and $q = 0$.

3.6. Box–Cox transformation, ARMA errors, Trend and Seasonal components (BATS)

Initially, an Exponential smoothing state space model with Box–Cox transformation, ARMA errors, Trend and Seasonal components (TBATS) model which aims to provide accurate forecasts for time series with complex seasonality was pursued via the forecast package in R. However, during the modelling process we noticed that a BATS model was found to have a better fit for this data, which is non-seasonal. A detailed description of the model and the underlying formula can be found in [27] and is therefore not reproduced here. The model structure and parameters used by the BATS model for forecasting GT before re-estimation was *BATS*(1, {0, 1}), where 1 is the Box–Cox parameter and {0, 1} is the damping parameter. Note that we are permitting (additive) trend and no seasonal component.

See Appendix A.1 for a discussion on the Classical Autoregressive (AR), Bayesian Autoregressive (BAR), Vector Autoregressive (VAR), and Bayesian Vector Autoregressive (BVAR) Models.

3.7. Singular Spectrum Analysis (SSA)

The SSA technique is a filtering and forecasting technique which is exploited in a variety of fields (see for example, [28–32]). In brief, SSA seeks to filter the noise in a time series and reconstruct a less noisy signal which is then used for forecasting future data points [31] using only the SSA choices of window-length (L) and the number of eigenvalues r [33]. SSA also has its multivariate form which is referred to as Multivariate SSA (MSSA) and can be used for modelling and forecasting using multiple series. In comparison to SSA, MSSA is relatively new with few applications (see for example, [34–37] and [38,39]).

There are several benefits of using SSA and MSSA models. As these models are nonparametric they are not bound by the parametric assumptions of normality, stationarity and linearity [38]. As such, one is able to model the data without any transformations which in turn ensures there is no loss of information [39]. In addition, as SSA is a filtering technique, it enables users to decompose a time series in order to obtain a richer understanding of the underlying dynamics. Moreover, once the associated signal is extracted, SSA enables users to forecast the signal alone. For example, if we are interested in the trend component alone we have the option of extracting the trend from the data and then forecasting the trend.

Those interested in a detailed account of the theory underlying SSA are referred to Sanei and Hassani [15]. Given that many readers are likely to be unfamiliar with the SSA, via Fig. 1 we show the basic SSA process in a visual form.

In brief, the SSA technique is made up of two stages known as Decomposition and Reconstruction. Below we present the decomposition (or filtering) and reconstruction stages of SSA, and in doing so we mainly follow [15].

Stage 1: Decomposition

Consider the real-valued nonzero time series $Y_N = (y_1, \ldots, y_N)$ of sufficient length N. The only choice at this stage is of the Window Length L, an integer such that $2 \leq L \leq N$.

Step 1: Embedding

Embedding is a mapping that transforms a one-dimensional time series $Y_N = (y_1, \ldots, y_N)$ into the multi-dimensional series $X_1, \ldots, X_K$ with vectors $X_i = (y_i, \ldots, y_{i+L-1})' \in \mathbf{R}^L$, where $K = N - L + 1$. The embedding step results in the trajectory matrix $\mathbf{X} = [X_1, \ldots, X_K] = \left(x_{ij}\right)_{i,j=1}^{L,K}$, which is a Hankel matrix. Accordingly, all the elements along the diagonal $i + j = const$ are constant (i.e., a matrix in which the (i, j)th entry depends only on $i + j$, that is, each skew-diagonal is constant).

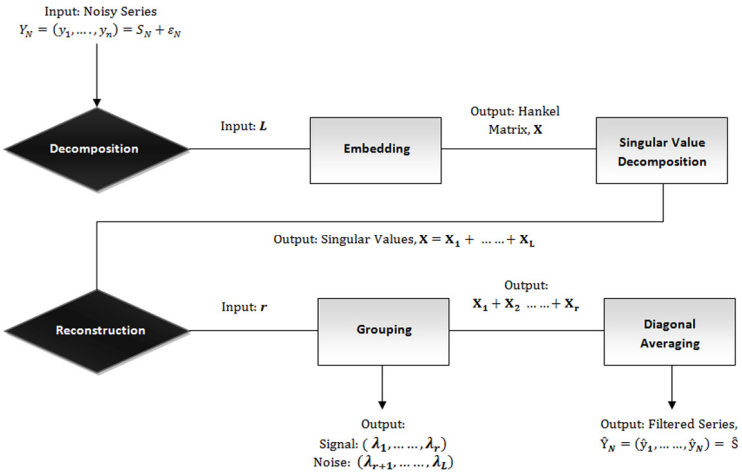

Fig. 1. The basic SSA process.

Step 2: Singular Value Decomposition (SVD)

In this step we obtain the singular value decomposition of the trajectory matrix X. Denote by $\lambda_1, \ldots, \lambda_L$ the eigenvalues of $\mathbf{X}\mathbf{X}'$ in decreasing order of magnitude ($\lambda_1 \geq \cdots \lambda_L \geq 0$) and by $U_1, \ldots, U_L$ the orthonormal system (that is, $(U_i, U_j) = 0$ for $i \neq j$ (the orthogonality property) and $\|U_i\| = 1$ (the unit norm property)) of the eigenvectors of the matrix $\mathbf{X}\mathbf{X}'$ corresponding to these eigenvalues. If we denote $V_i = \mathbf{X}'U_i/\sqrt{\lambda_i}$, then the SVD of the trajectory matrix can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d, \tag{4}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i'$ $(i = 1, \ldots, d)$.

Stage 2: Reconstruction

The second and final choice in SSA, i.e. the number of eigenvalues, r is required at this stage.

Step 1: Grouping

In the grouping step, we split the elementary matrices $\mathbf{X}_i$ into several groups and sum the matrices within each group. Let $I = \{i_1, \ldots, i_p\}$ be a group of indices $i_1, \ldots, i_p$. Then the matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_p}$. The split of the set of indices $J = 1, \ldots, d$ into the disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m}. \tag{5}$$

The procedure of choosing the sets $I_1, \ldots, I_m$ is called the eigentriple grouping.

Step 2: Diagonal Averaging

Diagonal averaging transfers each matrix I into a time series, which is an additive component of the initial series Y_N. This procedure is called *diagonal averaging*, or Hankelization of the matrix $\mathbf{Z}$. The result of the Hankelization of a matrix $\mathbf{Z}$ is the Hankel matrix $\mathcal{H}\mathbf{Z}$, which is the trajectory matrix corresponding to the series obtained as a result of the diagonal averaging. By applying the Hankelization procedure to all matrix components of (5), we obtain another expansion:

$$\mathbf{X} = \widetilde{\mathbf{X}}_{I_1} + \cdots + \widetilde{\mathbf{X}}_{I_m} \tag{6}$$

where $\widetilde{\mathbf{X}}_{I_1} = \mathcal{H}\mathbf{X}_{I_1}$. This is equivalent to the decomposition of the initial series $Y_N = (y_1, \ldots, y_N)$ into a sum of m series:

$$y_n = \sum_{k=1}^{m} \widetilde{y}_n^{(k)} \tag{7}$$

where $\widetilde{Y}_N^{(k)} = (\widetilde{y}_1^{(k)}, \ldots, \widetilde{y}_N^{(k)})$ corresponds to the matrix $\mathbf{X}_{I_k}$.

3.8. Multivariate Singular Spectrum Analysis (MSSA)

Those interested in an in-depth explanation of the theory underlying MSSA are directed to Hassani and Mahmoudvand [33]. We begin by presenting the Horizontal MSSA Recurrent (HMSSA-R) optimal forecasting algorithm which is followed by

the Horizontal MSSA Vector (HMSSA-V) optimal forecasting algorithm. In presenting these two algorithms we mainly follow and rely on the notations in [33].

3.8.1. HMSSA-R optimal forecasting algorithm

1. Consider M time series with identical series lengths of N_i, such that $Y_{N_i}^{(i)} = (y_1^{(i)}, \ldots, y_{N_i}^{(i)})\,(i = 1, \ldots, M)$.
2. In this case we use 25 observations of GT and CO_2 data to train and test the HMSSA models.
3. Beginning with a fixed value of $L = 2\,(2 \leq L \leq \frac{N}{2})$ and in the process, evaluating all possible values of L for Y_{N_i}, using the training data construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \ldots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L,K_i}$ for each single series $Y_{N_i}^{(i)}\,(i = 1, \ldots, M)$ separately.
4. Then, construct the block trajectory matrix $\mathbf{X}_H$ as follows:
$$\mathbf{X}_H = \left[\mathbf{X}^{(1)} : \quad \mathbf{X}^{(2)} : \quad \cdots \quad : \mathbf{X}^{(M)}\right].$$
5. Let vector $U_{H_j} = (u_{1j}, \ldots, u_{Lj})^T$, with length L, be the jth eigenvector of $\mathbf{X}_H\mathbf{X}_H^T$ which represents the SVD.
6. Evaluate all possible combinations of $r\,(1 \leq r \leq L-1)$ step by step for the selected L and construct $\widehat{\mathbf{X}}_H = \sum_{i=1}^{r} U_{H_i} U_{H_i}^T \mathbf{X}_H$ as the reconstructed matrix obtained using r eigentriples:
$$\mathbf{X}_H = \left[\widehat{\mathbf{X}}^{(1)} : \quad \widehat{\mathbf{X}}^{(2)} : \quad \cdots \quad : \widehat{\mathbf{X}}^{(M)}\right].$$
7. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathcal{H}\widehat{\mathbf{X}}^{(i)}\,(i = 1, \ldots, M)$ as the result of the Hankelization procedure of the matrix $\widehat{\mathbf{X}}^{(i)}$ obtained from the previous step for each possible combination of SSA choices.
8. Let $U_{H_j}^{\triangledown}$ denote the vector of the first $L-1$ coordinates of the eigenvectors U_{H_j}, and π_{H_j} indicate the last coordinate of the eigenvectors $U_{H_j}\,(j = 1, \ldots, r)$.
9. Define $\upsilon^2 = \sum_{j=1}^{r} \pi_{H_j}^2$.
10. Denote the linear coefficients vector $\mathcal{R}$ as follows:
$$\mathcal{R} = \frac{1}{1-\upsilon^2} \sum_{j=1}^{r} \pi_{Hj} U_{Hj}^{\triangledown}. \tag{8}$$
11. If $\upsilon^2 < 1$, then the h-step ahead HMSSA forecasts exist and is calculated by the following formula:
$$\left[\hat{y}_{j_1}^{(1)}, \ldots, \hat{y}_{j_M}^{(M)}\right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \ldots, \tilde{y}_{j_M}^{(M)}\right], & j_i = 1, \ldots, N_i, \\ \mathcal{R}^T \mathbf{Z}_h, & j_i = N_i + 1, \ldots, N_i + h, \end{cases} \tag{9}$$
where, $\mathbf{Z}_h = \left[Z_h^{(1)}, \ldots, Z_h^{(M)}\right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i-L+h+1}^{(i)}, \ldots, \hat{y}_{N_i+h-1}^{(i)}\right]\,(i = 1, \ldots, M)$.
12. Seek the combination of L and r which minimizes a loss function, $\mathcal{L}$ and thus represents the optimal HMSSA-R choices for decomposing and reconstructing in a multivariate framework.
13. Finally use the selected optimal L to decompose the series comprising of the validation set, and then select r singular values for reconstructing the less noisy time series, and use this newly reconstructed series for forecasting the remaining $\frac{1}{3}^{\text{rd}}$ observations.

3.8.2. HMSSA-V optimal forecasting algorithm

1. Begin by following the steps in 1–9 of the HMSSA-R optimal forecasting algorithm above.
2. Consider the following matrix
$$\boldsymbol{\Pi} = \mathbf{U}^{\triangledown}\mathbf{U}^{\triangledown T} + (1-\upsilon^2) R R^T, \tag{10}$$
where $\mathbf{U}^{\triangledown} = [U_1^{\triangledown}, \ldots, U_r^{\triangledown}]$. Now consider the linear operator
$$\mathcal{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbb{R}^L, \tag{11}$$
where
$$\mathcal{P}^{(v)} Y = \begin{pmatrix} \boldsymbol{\Pi} Y_{\triangle} \\ R^T Y_{\triangle} \end{pmatrix},\ Y \in \mathfrak{L}_r, \tag{12}$$
and $Y_{\triangle}$ is vector of last $L-1$ elements of Y.
3. Define vector $Z_j^{(i)}\,(i = 1, \ldots, M)$ as follows:
$$Z_j^{(i)} = \begin{cases} \widetilde{X}_j^{(i)} & \text{for } j = 1, \ldots, k_i \\ \mathcal{P}^{(v)} Z_{j-1}^{(i)} & \text{for } j = k_i + 1, \ldots, k_i + h + L - 1 \end{cases} \tag{13}$$

Table 1
MSSA models for forecasting Global Temperature.

h	HMSSA-R	HMSSA-V
1	MSSA(3, 1)	MSSA(4, 1)
2	MSSA(3, 1)	MSSA(3, 1)
3	MSSA(3, 1)	MSSA(5, 1)
4	MSSA(4, 1)	MSSA(4, 1)
5	MSSA(3, 1)	MSSA(4, 1)
6	MSSA(3, 1)	MSSA(3, 1)
7	MSSA(3, 1)	MSSA(2, 1)
8	MSSA(3, 1)	MSSA(3, 1)
9	MSSA(2, 1)	MSSA(2, 1)
10	MSSA(2, 1)	MSSA(5, 3)

Note: Shown here in brackets are the combinations of L and r as MSSA(L, r).

where, $\widetilde{X}_j^{(i)}$'s are the reconstructed columns of trajectory matrix of the ith series after grouping and leaving noise components.

4. Now, by constructing matrix $\mathbf{Z}^{(i)} = [Z_1^{(i)}, \ldots, Z_{k_i+h+L-1}^{(i)}]$ and performing diagonal averaging we obtain a new series $\hat{y}_1^{(i)}, \ldots, \hat{y}_{N_i+h+L-1}^{(i)}$, where $\hat{y}_{N_i+1}^{(i)}, \ldots, \hat{y}_{N_i+h}^{(i)}$ provides the h step ahead HMSSA-V forecast for the selected combination of L and r.
5. Finally, follow steps 12–13 in the HMSSA-R optimal forecasting algorithm to find the optimal L and r for obtaining HMSSA-V forecasts.

Finally, in Table 1 we present the MSSA models for obtaining the forecasts for GT at each horizon.

4. Data and metrics

4.1. Data

The data we investigate here consists of two variables, namely the global temperature anomaly (GT) and global carbon dioxide (CO_2) emissions, and spans the annual period of 1880–2012, with the start and end of the period being purely driven by the availability of data. GT measures the difference between a reference long-term average value (which happens to be 1951–1980) and the actual value. The calculation of GT is in itself a complex process adjusting for aspects such as, though not limited to, unequal distance between measuring stations, difference due to the richly observed northern hemisphere versus the poorly observed southern hemisphere, ocean versus terrestrial measurements, polar regions, into one representative number for the whole earth [40]. Similarly global CO_2 is estimated from energy statistics published by the United Nations [41] involving a complex system comprising of, again not limited to, questionnaires, official statistics and other supplementary information [42]. The GT were obtained from the *National Aeronautics and Space Administration*'s (NASA) and the *Goddard Institute for Studies* (GISS).[8] The data on CO_2 was obtained from the *Carbon Dioxide Information Analysis Centre*[9] and is measured in thousand metric tons of carbon. While GT remains untransformed in our analysis, we use the natural logarithm of CO_2 emissions. The log transformation allows us to make the highly skewed distribution of CO_2 less skewed, which in turn, is valuable both for making patterns in the data more interpretable and for helping to meet the assumptions of inferential statistics.

4.2. Forecasting exercise

We separate the entire data period into an initial in-sample period spanning 1880–1906, and an out-of-sample period of 1907–2012. Note that, since we use a recursive estimation for the models (barring RW and SSA), we have an expanding in-sample period, while producing the out-of-sample forecasts for the various horizons. This separation of the period was determined by the Bai and Perron [8] tests of multiple structural breaks applied to the global temperatures equation of the VAR model, which, in turn, detected the following break dates: 1907, 1945, 1974, 1992, and 2004. Since our linear models are estimated recursively over the out-of-sample period in which all the break dates fall, this separation of the period is ideal as it helps us to accommodate changes in the parameter estimates of the model in the out-of-sample period. In other words, if we would have a split in the initial in-sample and out-of-sample periods different from the one based on the structural break tests, with some of the breaks in the initial in-sample period, the linear models would be misspecified. This is because, the linear models would not be capturing the change in the response of GT to CO_2 emissions that took place due to breaks in the

[8] The data is available for download from: http://data.giss.nasa.gov/gistemp/.
[9] The data is available for download from: http://cdiac.ornl.gov/trends/emis/tre_glob_2013.html.

Table 2
Descriptive statistics for GT.

	Min	Max	Mean	Med.	SD	IQR	SW(p)
Overall (1880–2012)	−47	66	−1.47	−7	28.96	35	<0.01
In-Sample (1880–1906)	−44	−10	−25.44	−25	8.09	12.5	0.95*
Out-of-Sample (1907–2012)	−47	66	4.63	2.50	29.20	36.75	<0.01

*Indicates data is normally distributed based on the Shapiro–Wilk (SW) test at $p = 0.01$.

initial in-sample. This way, we aim to make the comparison between linear and nonlinear models fair, by ensuring that the superior predictability of the nonlinear models, if and when it exists, is not because of misspecification due to regime changes but due to the inherent nonlinear relationship of GT with CO_2.[10] Finally, we produce one- to ten-year-ahead forecasts based on this recursive estimation scheme to forecast in the short- and long-term horizon. Except for the MSSA models where the initial model remains constant and the data is updated, all other models update the data and re-estimate model parameters each time a new observation is introduced. Accordingly, we generate forecasts at each horizon such that we have $n - h + 1$ observations where $n = 106$ and h refers to the horizon of interest.

4.3. Descriptive statistics

Table 2 summarizes the descriptive statistics for the entire GT data set, and the in-sample and out-of-sample periods separately. This analysis uncovers some interesting insights relating to the series. Firstly, the structure of the series overall and out-of-sample appears to be closely associated as expected (given the large out-of-sample forecasting period). Both these series are skewed and report an almost identical interquartile range (IQR), and the out-of-sample data captures the overall minimum and maximum GT values. In comparison, the initial in-sample series captures a period where the GT values are negative and normally distributed.

4.4. Metrics

To evaluate the 12 competing models, we use the popular Root Mean Squared Error (RMSE) loss function and the direction of change (DC) criteria for comparing their forecasting performances. All outcomes relating to forecast accuracy are tested for statistical significance using the Kolmogorov–Smirnov Predictive Accuracy (KSPA) test [20], whilst the DC results are tested for statistical significance using a Student's t-test.

Root Mean Squared Error (RMSE)

The RMSE is now a standard quantitative technique for evaluating forecasting accuracy of alternate models. It is also popular as one of the most frequently cited measures in forecasting literature (see, for example, [43] and [23]). In this paper we consider both the RMSE and Ratio of the RMSE (RRMSE). Here, in order to save space we present only the RRMSE.

$$\text{RRMSE} = \left(\frac{\sum_{i=1}^{N} (\hat{Y}_{T+h,i} - Y_{T+h,i})^2}{\sum_{i=1}^{N} (\tilde{Y}_{T+h,i} - Y_{T+h,i})^2} \right)^{1/2},$$

where $\hat{Y}_{T+h}$ represents the h-step ahead forecast obtained by the benchmark model, $\tilde{Y}_{T+h}$ represents the h-step ahead forecast obtained by a competing model, and N is the number of forecasts. If RRMSE < 1 then the benchmark model outperforms the competing model by 1-RRMSE percent.

Direction of Change (DC)

The DC metric is an equally important measure as the RMSE because it is able to show whether the forecast is correctly predicting the actual direction of change. A model is said to have a better DC prediction than a random walk if it records a DC greater than 50% [44]. The DC calculation is presented below and in doing so we mainly follow [38].

Consider a univariate time series Y_T. Let Z_{Yi} take the value 1 if the forecast correctly predicts the direction of change and 0 otherwise. Then $\bar{Z}_X = \sum_{i=1}^{N} Z_{Yi}/N$ shows the proportion of forecasts that correctly predict the direction of the series.

5. Empirical results

Forecasting results covering all horizons from 1–10 can be found in Tables 7 and 8. The presentation of the results which considers both short and long term forecasts of GT enables stakeholders to select the best model at each forecast horizon

[10] Given that, the structural break test is performed on the full-sample of the data, which in turn, is not available to a forecaster while making a forecast, we reconducted our analysis using an initial in-sample period of 1880–1946 i.e., essentially a 50 percent split of the data. Our revised analyses, which are available upon request, produced qualitatively similar results.

of interest. Here in Tables 3 and 4, we present a summarized version of the out-of-sample forecasting results for GT which considers horizons 1, 5, and 10, and thereby represents the short, medium and long term forecasts. We initially discuss the overall results. The first observation is that the HMSSA-V model can provide the best forecast for GT across all horizons. Also, it is important to note that the two MSSA techniques with their various fitted models succeed in reporting the lowest RMSE and MAE in comparison to all models evaluated here, across all horizons. As such, in relation to all models considered in this study, we are able to identify clearly that MSSA has the potential to provide the most accurate forecasts for GT when CO_2 is used as auxiliary information. Secondly, given that these superior MSSA results have been attained by using CO_2 as an indicator variable, we are able to conclude that CO_2 can indeed help in predicting GT. If one is interested in a single model for obtaining the best possible forecasts for GT, then based on the average lowest RMSE and MAE we can propose that specifically the HMSSA-V technique, as used in this study, is best for this purpose.

The results also enable a more detailed analysis and differentiation between univariate and multivariate models. In terms of the univariate models we see that across all horizons, ETS forecasts are best for GT (Table 7). Interestingly as well, on most occasions the RW forecasts are seen outperforming the univariate models (except ETS), and also outperforming the multivariate BVAR5 and VAR forecasts with the exception of $h = 1$ step ahead results from VAR and BVAR5 models. Another point to note is that forecasts for GT from the univariate BAR1 model is almost as bad as the worst performing NN model according to the average RMSE and MAE criteria. Furthermore, as both univariate and multivariate models are considered in this study, practitioners also have the option of selecting the best univariate or multivariate model for forecasting GT at a particular horizon of interest.

The RMSE and MAE results reported in Table 7 clearly indicates how the forecasting accuracy of all models deteriorates as the forecasting horizon expands from 1-step ahead to 10-steps ahead. A closer look at these results show that ETS and HMSSA models incur a deterioration in accuracy levels at a comparatively slower rate than rest of the models. This in turn indicates that the univariate ETS model is performing considerably well in relation to other univariate models considered here, and also that the MSSA models are performing better than all other models in terms of the RMSE criteria.

In Table 4 we consider HMSSA-V as a benchmark model and present the RRMSE results for GT forecasts for horizons 1, 5, and 10 steps-ahead with Table 8 reporting all results. The choice of HMSSA-V as a benchmark model was influenced by the fact that the HMSSA-V model reports the lowest RMSE and lowest average MAE across all ten forecasting horizons. Here we exploit the two-sided HS test [20] and the modified Diebold–Mariano test to find statistically significant differences between forecasts across all 10 forecasting horizons. Evidence from the two-sided KSPA test (results indicated on the 'Avg.' row in Table 4) suggests that on average, across all 10 horizons, the only instance in which HMSSA-V fails to produce a forecast that is not significantly better than another model's forecast is in comparison to HMSSA-R and ETS forecasts at a 90% confidence level. This provides added justification for selecting the HMSSA-V model proposed in this study as a benchmark for forecasting GT. At the same time, it also shows the power of the ETS model for those wishing to rely on a univariate forecasting approach when modelling GT.

The RRMSE results in Table 4 are further tested for statistical significance via the two-sided KSPA test which seeks to ascertain the existence of a statistically significant difference between the distributions of two forecast errors. Leaving aside the HMMSA-R and HMSSA-V models which do not report any statistically significant differences between forecast errors at the 90% confidence level, we can conclude that beyond $h = 1$, the HMSSA-V forecasts are indeed significantly better than all other competing forecasts on most instances (with the exception of ETS). Interestingly, HMSSA-V forecasts only outperformed ETS at $h = 2$ and $h = 10$ steps ahead. Also, we failed to find evidence of statistically significant differences between forecasts of HMSSA-V and RW models at $h = 1,\ 4,\ and\ 8$ steps ahead. The beauty of the RRMSE criterion is that it enables us to further quantify the performance of a given forecasting model and show by what percentage it outperforms forecasts from another model. Accordingly, based on the RRMSE we are able to conclude that on average, across all ten horizons, the HMSSA-V forecasts are better in the range of 1% (relative to the HMSSA-R) and 29% (relative to the NN).

Presented in Table 5 (summarized output) and Table 9 (all output) are the DC prediction results from the GT forecasts. It is important to evaluate the accuracy of a model in its ability to correctly predict the actual direction of change in the time series.[11] The DC results indicate that on average the univariate BAR1 model is worst in predicting the actual direction of change in GT. If we consider only the univariate models, then the AR model has the worst DC prediction whilst the ETS model reports the best. In terms of multivariate models, the MSSA models report the best DC predictions on average and are almost identical. However, if one was to suggest a single model with the best average DC prediction then it would be the HMSSA-V model with an average accuracy of 66%. The inclusion of the DC results at each horizon enables practitioners to select the best model for a particular forecasting horizon not only based on the RMSE criterion but also based on its ability at providing a good DC prediction for GT. In line with good statistical practice we test all DC predictions for statistical significance using the Student's t-test as in [45]. Accordingly, we see that the DC predictions reported for AR, ARIMA, ARFIMA, NN, BATS, BAR1, VAR, and BVAR5 models at all horizons are likely to be a result of chance occurrences. However, the DC prediction for ETS at $h = 5$ steps ahead is found to be statistically significant whilst half of the DC predictions from the HMSSA-V model report statistically significant outcomes. As such, we are able to conclude with 90% confidence that the HMSSA-V DC predictions are significantly greater than 50% across forecasting horizons 2–6.

[11] Note that it is not possible to calculate the DC metric for RW forecasts as it results in the DC statistic going to infinity.

Table 3
RMSE for out-of-sample forecasts for Global Temperature.

h		RW	AR	ARIMA	ETS	NN	ARFIMA	BATS	BAR1	VAR	BVAR5	HMSSA-R	HMSSA-V
1	RMSE	10.66	10.68	10.21	10.00	12.25	10.76	10.28	10.70	10.65	10.34	10.25	**9.99**
	MAE	8.62	8.33	8.31	8.31	10.00	8.96	8.60	8.38	*7.62*	7.91	8.46	8.22
5	RMSE	14.61	17.55	14.88	13.28	17.51	17.14	15.34	18.14	16.12	15.54	12.80	**12.68**
	MAE	11.68	13.60	12.28	10.93	13.52	13.62	12.69	14.06	12.62	12.01	10.39	*10.38*
10	RMSE	16.92	23.89	19.43	16.87	27.29	23.21	18.76	25.45	20.26	19.31	16.22	**15.39**
	MAE	14.21	19.90	16.04	14.31	18.64	18.99	16.13	21.27	16.40	15.57	13.38	*12.61*

Note: Shown in bold font is the model reporting the lowest RMSE. The model reporting the lowest MAE is italicized. The multivariate models use CO_2 at lag 2.

Table 4
RRMSE for out-of-sample forecasts for Global Temperature.

h	$\frac{HMSSA-V}{RW}$	$\frac{HMSSA-V}{AR}$	$\frac{HMSSA-V}{ARIMA}$	$\frac{HMSSA-V}{ETS}$	$\frac{HMSSA-V}{NN}$	$\frac{HMSSA-V}{ARFIMA}$	$\frac{HMSSA-V}{BATS}$	$\frac{HMSSA-V}{BAR1}$	$\frac{HMSSA-V}{VAR}$	$\frac{HMSSA-V}{BVAR5}$	$\frac{HMSSA-V}{HMSSA-R}$
1	0.94	0.94	0.98	0.99	0.82*,†	0.93	0.97	0.93	0.94	0.97	0.97
5	0.87*,†	0.72*,†	0.85*,†	0.95	0.72*,†	0.74*,†	0.83*,†	0.70*,†	0.79*,†	0.82*,†	0.99
10	0.91*	0.64*,†	0.79*,†	0.91*	0.56*,†	0.66*,†	0.82*,†	0.60*,†	0.76*,†	0.80	0.95
Avg.	0.91*	0.75*	0.87*	0.96	0.71*	0.76*	0.86*	0.72*	0.81*	0.84*	0.99

*Indicates a statistically significant difference between the distribution of forecasts based on the two-sided KSPA test in [20] at $p = 0.10$.
†Indicates the existence of a statistically significant difference between the forecasts based on the modified Diebold–Mariano test in [21].

Table 5
DC predictions from out-of-sample forecasts for Global Temperature.

h	AR	ARIMA	ETS	NN	ARFIMA	BATS	BAR1	VAR	BVAR5	HMSSA-R	HMSSA-V
1	0.42	0.54	0.60	0.57	0.56	0.58	0.42	0.38	0.41	0.60	0.61
5	0.49	0.51	0.66*	0.55	0.53	0.50	0.47	0.56	0.55	0.70*	0.67*
10	0.38	0.42	0.55	0.49	0.39	0.41	0.35	0.54	0.51	0.60	0.64
Avg.	0.45	0.49	0.60	0.58	0.51	0.48	0.43	0.52	0.50	0.65	0.66

*Indicates the DC prediction is statistically significant based on a Student's t test at $p = 0.05$.

Table 6
RMSE results from out-of-sample forecasts for Global Temperature using Radiative Forcing.

h	RW	AR	BAR1	VAR	BVAR5	HMSSA-R	HMSSA-V	RF:HMSSA-R	RF:HMSSA-V
1	10.66	10.68	10.7	10.65	10.34	10.25	9.99	11.53	10.58
2	12.61	13.1	13.2	12.61	12.32	10.99	10.98	12.12	11.04
3	13.22	14.51	14.75	13.75	13.43	11.71	11.68	12.77	11.18
4	12.89	15.64	16.03	14.57	13.97	12.22	12.08	13.48	11.95
5	14.61	17.55	18.14	16.12	15.54	12.80	12.68	13.73	12.83
6	14.82	18.77	19.54	17.07	16.37	13.11	13.06	14.6	13.40
7	14.89	19.69	20.59	17.86	17.23	13.62	13.61	15.58	14.09
8	15.44	21.11	22.22	18.82	18.03	14.68	14.56	16.65	15.35
9	17.05	22.72	24.08	19.69	18.83	15.51	15.50	16.51	17.22
10	16.92	23.89	25.45	20.26	19.31	16.22	15.39	17.04	18.13
Avg.	14.31	17.77	18.47	16.14	15.54	13.11	**12.95**	14.40	13.577

5.1. Robustness analysis

Though we followed the works of Fildes and Kourentzes [3], and McMillan and Wohar [4] in using CO_2 emissions as a predictor variable for GT, based on the suggestion of the Editor, we also evaluated the performance of our SSA models using radiative forcing instead of CO_2. Note, radiative forcing is defined as the difference of insolation (sunlight) absorbed by the Earth and energy radiated back to space. We observed that just like the MSSA models with CO_2, the same models with radiative forcing outperformed the univariate and multivariate linear models and the neural network model. However, the HMSSA-V model based on total radiative forcing, could only outperform the HMSSA-R and HMSSA-V based on CO_2 emissions at horizons of three- and four-year-ahead. Also, on average, the HMSSA-R and HMSSA-V using CO_2 emissions continued to perform the best. A selection of these results are reported via Table 6 for the benefit of the reader.

6. Discussion

The discussion begins with a graphical illustration of the best out-of-sample forecasts for GT in the very short run ($h = 1$) and the very long run ($h = 10$) as shown via Fig. 2. All forecasts shown are from the HMSSA-V model. Interestingly, the forecasts from the other models showed clear signs of difficulties in modelling and providing an accurate forecast for GT amidst the variation visible in the series, and in most cases the competing models were seen picking up the variations too late in time and reflecting on the comparatively poor DC predictions reported in Table 5. The use of MSSA models enabled to overcome the issues pertaining to modelling the volatility in GT and provide forecasts which are comparatively more accurate and reliable for decision making.

However, as evident via Fig. 2, even the best forecasts appear to have difficulty in capturing the levelling which occurs post 2000 until recently. As such, it is pertinent to briefly comment on this. For this purpose, shown via Fig. 3 are the best performing one, five, and ten-steps ahead forecasts from 2000 onwards based on the lowest RMSE criteria. All forecasts appearing here are from the HMSSA-V model. Notice how some of the forecasts appear to fluctuate whilst the others are represented as comparatively smoother lines. There are two possible explanations for this. Firstly, MSSA is sensitive to the choice of window length L and the number of eigenvalues r, and the automated MSSA algorithm introduced in this paper is programmed to find the best possible forecast by minimizing a loss function. Therefore, instead of seeking out components which will capture most of the fluctuations, more weighting is given to the forecast line which can reduce the error. Secondly, the MSSA algorithm does not update and fit a new model to the data each time a new observation is introduced. Whilst this approach does provide more stability to the forecasting model, it does however make it difficult for the model to ensure the levelling which occurs towards the end of the out-of-sample period is accurately captured. Yet, it is noteworthy that in comparison to the other models which are all recursive, these MSSA forecasts with a fixed model fitted at the training stage is able to outperform the rest. Given the non-recursive nature of the MSSA forecasts, as the horizon increases beyond 1-step ahead, we notice a deterioration in the forecast accuracy (it still deteriorates at a lower rate than the recursive models). It would be interesting to see whether programming recursive MSSA models can enable MSSA to capture the levelling post 2000 more accurately.

Given that GT data is now available till 2017, below, we also show the ex ante forecasts from the best MSSA models using CO_2 and total radiative forcing (TRF) as auxiliary information. As expected, based on the forecasting exercise carried out in this study, we see that the HMSSA-V model using CO_2 as auxiliary information performs the best in tracking more closely the trend of GT. (See Fig. 4.)

At this stage, it is important to compare our results with the existing evidence of predictability of GT emanating from CO_2 emissions. As discussed in the introduction, available studies based on time series analysis that do aim to forecast GT based on CO_2 emissions are limited to two, with bulk of the papers using climate models. Just as in [4], we find no evidence of the ability of CO_2 emissions in predicting GT based on linear models. This weak evidence is not surprising given that the Brock et al. [46] test indicated strong evidence of nonlinearity in the relationship between the two variables.[12] However,

12 Complete details of these results are available upon request from the authors.

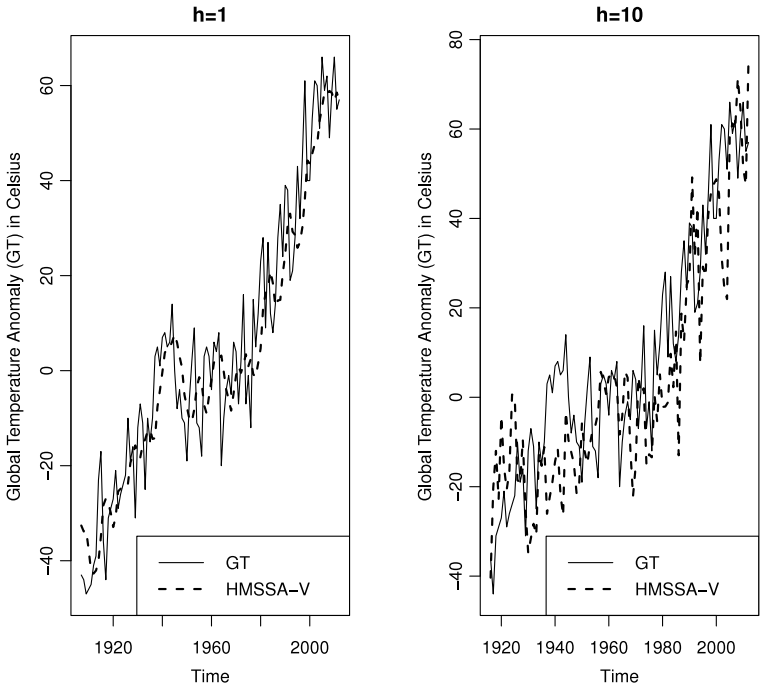

Fig. 2. Out-of-sample MSSA forecasts for GT.

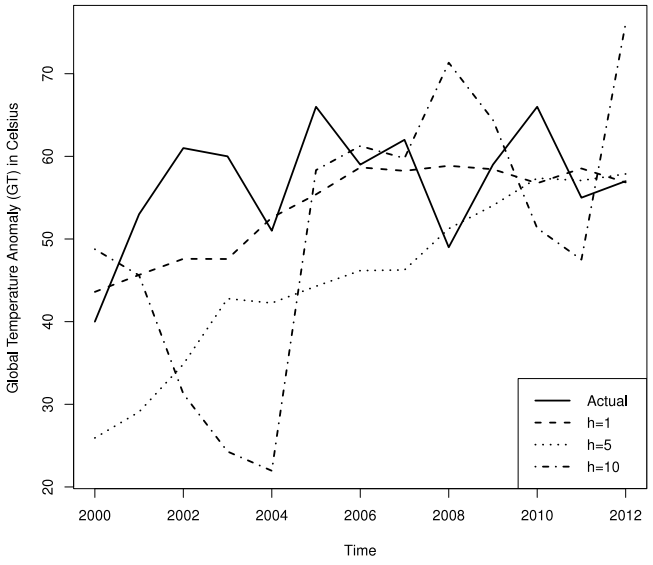

Fig. 3. The best out-of-sample forecasts (HMSSA-V model) post 2000 at selected horizons.

unlike in [3] to certain extent, we found that, even though nonlinearity exists, a NN model cannot seem to capture this underlying nonlinearity well enough, as is indicated in its poor forecasting performance. The superiority of the SSA models in forecasting however, tend to suggest that these models are able to better capture this unknown nonlinear relationship via a nonparametric formulation.[13]

[13] This line of thinking is vindicated by the superior fit of the SSA approach relative to the NN method. Complete details of the goodness-of-fit results are available upon request from the authors.

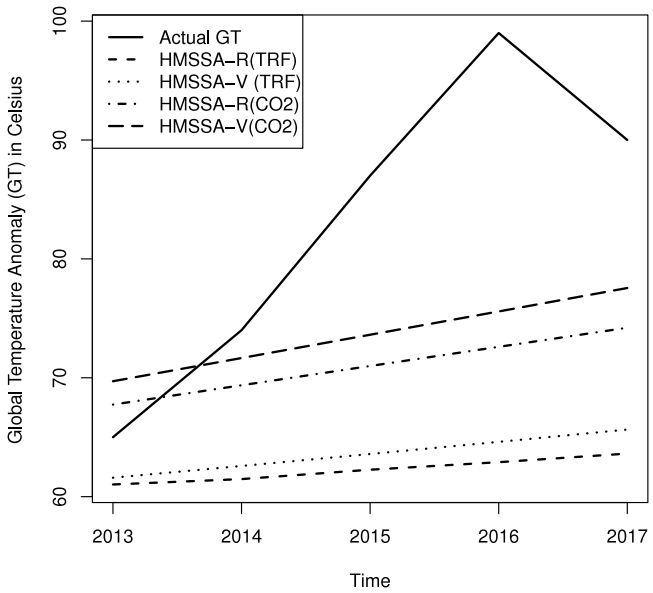

Fig. 4. The best MSSA forecasts for GT 2013–2017.

7. Conclusions

The two popular and passionate debates about climate change are (i) Is it possible to predict global temperature anomaly (GT) reliably?; and (ii) Is there definitive causal evidence of (anthropometric) CO_2 being the driver of GT? This paper is an exercise to contribute to this debate by providing objective analyses of relevant data using an ensemble comprising of 12 parametric and nonparametric out-of-sample forecasting techniques using only GT data (univariate), as well as using both GT and CO_2 data (multivariate). The significance of this paper lies in that, using the well established datasets we have identified the 'best' model, from 12 candidate models, for forecasting GT both in the short- and long-term horizons. Specifically, our results have identified that the Horizontal Multivariate Singular Spectral Analysis (HMSSA) models (both Recurrent (-R) and Vector (-V)) consistently outperform the other competing models in a statistically significant fashion. Further, from the performance of the HMSSA-R model, we have conclusive evidence that CO_2 can predict GT. We also evaluated the models in their ability to predict the direction of change (DC) for the GT forecasts. Although in the univariate setup the exponential smoothing (ETS) model performs best in forecasting DC followed by the HMSSA models, in the multivariate setup, the HMSSA models once again score best when evaluated using the metrics considered. Thus, from our investigation and from the analyses of the findings, if we have to recommend *a* model for forecasting GT, HMSSA-R is a clear winner. Our results also highlight the superiority of the nonparametric approach of the SSA, which in turn, allows us to handle any statistical process: linear or nonlinear, stationary or non-stationary, Gaussian or non-Gaussian. Hence, from the perspective of a climate modeller relying on AOGCMs, it would make sense to compare the ex ante forecasts (i.e., future path) of GT generated from the SSA model with that from the AOGCMs. In this regard, as part of future analysis, it would be interesting to go back and compare the future paths generated in historical reports of IPCC on GT, with that generated from our SSA model. Using historical comparison will allow us to determine how far or close the paths generated by the climate models and that by time series models match up with what actually materialized in terms of global temperature levels. Also note, using the SSA model, given that it is based on bootstrapping, we will be able to generate interval and density forecasts for global temperatures, thus allowing us to determine the uncertainty associated with such forecasts.

Though the objective of this paper was to pursue a forecasting exercise to determine whether CO_2 emissions can predict the future movements in GT, using a large number of linear and nonlinear models, our results can be utilized to draw important policy conclusions as well. We show that it is important to model nonlinearity into the relationship between GT and CO_2, otherwise, the latter does not seem to contain any predictive information about the future paths of GT. Once this is done, we find strong evidence of the role of CO_2 in increasing GT, i.e., we show that continued CO_2 emissions will cause further warming.[14]

[14] Based on the suggestion of an anonymous referee, we estimated a time-varying model of cointegration as outlined in [47]. Based on this framework, we observed that the response of GT to CO_2 has steadily decreased post World War II, with the average impact suggesting that one unit increase in CO_2 would increase GT by 0.0486 unit. Complete details of these results are available upon request from the authors. As discussed in the Fifth Climate Change Synthesis Report of the IPCC [48], continuous global warming will have long-lasting changes in all components of the climate system, increasing the likelihood of severe, pervasive and irreversible impacts for both people and ecosystems. In addition, climate change will amplify existing risks and create new risks

Our findings also open up several avenues for future research. It would be interesting to analyse the emissions from various countries including the top emitters, to determine which countries can predict global temperatures the best, and also to find out whether the top emitters can predict global temperatures better than world emissions. The findings from such analysis can help design better policies. Also, it is important to point out that, while we only concentrate on CO_2 emissions in this paper, there is also a large literature that has related global temperatures to sunspots, but primarily based on in-sample analysis (see [49] for a detailed review). Given this, in the future it would be interesting to extend this literature to out-of-sample forecasting, and also include sunspot along with CO_2 emissions to check the robustness of our results.

Appendix

A.1. Classical Autoregressive (AR), Bayesian Autoregressive (BAR), Vector Autoregressive (VAR), and Bayesian Vector Autoregressive (BVAR) models

The Vector Autoregressive (VAR) model, though "atheoretical" is particularly useful for forecasting purposes. Note an unrestricted VAR model, as suggested by Sims [50], can be written as follows:

$$y_t = C + A(L)y_{t-1} + \epsilon_t \tag{14}$$

where: y: a vector of variables (global temperatures and global CO_2 emissions) being forecasted; $A(L)$: a polynomial matrix in the backshift operator L with lag length p, i.e., $A(L) = A_1L + A_2L^2 + \cdots + A_pL^p$; C: a vector of constant terms, and ε: vector of white-noise error terms. In our case $p = 2$ based on the Akaike Information Criterion (AIC).

The VAR model uses equal lag length for all the variables of the model, and leads to the problem of overparameterization. This, in turn, results in multicollinearity and loss of degrees of freedom leads to inefficient estimates and large out-of-sample forecasting errors.

An approach to overcome this overparameterization, as described in [51,52], is to use a Bayesian VAR (BVAR) model. Instead of eliminating longer lags, the Bayesian method imposes restrictions on these coefficients by assuming that these are more likely to be near zero than the coefficient on shorter lags. However, if there are strong effects from less important variables, the data can override this assumption. The restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients with the standard deviation decreasing as the lags increases. The exception to this is, however, the coefficient on the first own lag of a variable, which has a mean of unity. Note Litterman [51] used a diffuse prior for the constant. This is popularly referred to as the "Minnesota prior" due to its development at the University of Minnesota and the Federal Reserve Bank at Minneapolis.

The standard deviation of the distribution of the prior for lag m of variable j in equation i for all i, j and m, defined as $S(i, j, m)$, can be specified as follows:

$$S(i, j, m) = [w \times g(m) \times f(i, j)]\frac{\sigma_i}{\sigma_j} \tag{15}$$

with $f(i, j) = 1$, if $i = j$ and k_{ij} otherwise, with $(0 \leq k_{ij} \leq 1)$, $g(m) = m^{-d}$, $d > 0$. Note σ_i is the standard error of the univariate autoregression for variable i. The ratio $\frac{\sigma_i}{\sigma_j}$ scales the variables so as to account for differences in the units of measurement and, hence, causes specification of the prior without consideration of the magnitudes of the variables. The term w indicates the overall tightness and is also the standard deviation on the first own lag, with the prior getting tighter as we reduce the value. The parameter $g(m)$ measures the tightness on lag m with respect to lag 1, and is assumed to have a harmonic shape with a decay factor of d, increasing which tightens the prior on increasing lags. The parameter $f(i, j)$ represents the tightness of variable j in equation i relative to variable i, and by increasing the interaction, i.e., the value of k_{ij}, we can loosen the prior. Following the literature on BVAR models, we look at the following combinations of w and d: (0.3, 0.5), (0.2, 1.0), (0.1, 1.0), (0.2, 2.0) and (0.1, 2.0), with k_{ij} set at 0.5. Univariate versions of the BVAR models, which we call Bayesian autoregressive (BAR) models, are estimated for the same values of w and d as above, but with k_{ij} set at 0.001, since a small interaction value basically reduces the multivariate model to its corresponding univariate version. In the results section however, we only report the BAR and BVAR models which produces the most accurate forecasts on average, which in our case, happened to be the BAR1 ($w = 0.3$, $d = 0.5$, $k_{ij} = 0.001$) and BVAR5 ($w = 0.1$, $d = 2.0$, $k_{ij} = 0.5$).[15] In this regard, we follow Gupta and

for natural and human systems; with risks being higher for vulnerable communities around the world, resulting in an unequal society, and hamper socio-economic development. Understandably, limiting climate change would require not only substantial and sustained reductions in CO_2 emissions (mitigation), but also adaptation, to limit negative climate change impacts.

[15] Complete details of the results from the other BAR and BVAR models are available upon request from the authors.

Table 7
RMSE for out-of-sample forecasts for Global Temperature.

h		RW	AR	ARIMA	ETS	NN	ARFIMA	BATS	BAR1	VAR	BVAR5	HMSSA-R	HMSSA-V
1	RMSE	10.66	10.68	10.21	10.00	12.25	10.76	10.28	10.70	10.65	10.34	10.25	**9.99**
	MAE	8.62	8.33	8.31	8.31	10.00	8.96	8.60	8.38	*7.62*	7.91	8.46	8.22
2	RMSE	12.61	13.10	12.14	11.44	13.78	12.98	12.86	13.20	12.61	12.32	10.99	**10.98**
	MAE	10.21	10.57	10.07	9.45	11.04	10.86	10.35	10.65	9.97	9.50	8.95	*8.94*
3	RMSE	13.22	14.51	13.10	12.01	16.25	14.51	13.92	14.75	13.75	13.43	11.71	**11.68**
	MAE	10.73	11.55	10.64	9.95	12.56	12.06	11.59	11.74	11.15	10.77	*9.46*	9.60
4	RMSE	12.89	15.64	13.62	12.27	17.20	15.51	13.77	16.03	14.57	13.97	12.22	**12.08**
	MAE	10.49	12.45	11.10	10.03	13.12	12.76	11.72	12.71	11.57	10.99	9.89	*9.77*
5	RMSE	14.61	17.55	14.88	13.28	17.51	17.14	15.34	18.14	16.12	15.54	12.80	**12.68**
	MAE	11.68	13.60	12.28	10.93	13.52	13.62	12.69	14.06	12.62	12.01	10.39	*10.38*
6	RMSE	14.82	18.77	15.53	13.64	19.13	18.17	15.73	19.54	17.07	16.37	13.11	**13.06**
	MAE	11.86	14.93	12.68	11.10	13.93	14.56	13.00	15.66	13.62	12.96	10.81	*10.73*
7	RMSE	14.89	19.69	16.31	14.29	21.76	19.11	15.80	20.59	17.86	17.23	13.62	**13.61**
	MAE	12.52	16.22	13.78	12.33	15.71	15.63	13.48	17.04	14.47	13.78	*11.53*	11.54
8	RMSE	15.44	21.11	17.06	15.20	21.08	20.52	16.50	22.22	18.82	18.03	14.68	**14.56**
	MAE	12.90	17.33	14.41	12.69	15.67	16.55	14.05	18.34	15.49	14.77	12.06	*11.98*
9	RMSE	17.05	22.72	18.72	16.29	20.53	22.27	18.40	24.08	19.69	18.83	15.51	**15.50**
	MAE	13.77	18.68	15.28	13.42	15.32	17.98	15.38	19.88	15.95	15.15	12.53	*12.52*
10	RMSE	16.92	23.89	19.43	16.87	27.29	23.21	18.76	25.45	20.26	19.31	16.22	**15.39**
	MAE	14.21	19.90	16.04	14.31	18.64	18.99	16.13	21.27	16.40	15.57	13.38	*12.61*
Avg.	RMSE	14.31	17.77	15.10	13.53	18.68	17.42	15.14	18.47	16.14	15.54	13.11	**12.95**
Avg.	MAE	11.70	14.36	12.46	11.25	13.95	14.20	12.70	14.97	12.89	12.34	10.75	*10.63*

Note: Shown in bold font is the model reporting the lowest RMSE. The model reporting the lowest MAE is italicized. The multivariate models use CO_2 at lag 2.

Table 8
RRMSE for out-of-sample forecasts for Global Temperature.

h	$\frac{HMSSA-V}{RW}$	$\frac{HMSSA-V}{AR}$	$\frac{HMSSA-V}{ARIMA}$	$\frac{HMSSA-V}{ETS}$	$\frac{HMSSA-V}{NN}$	$\frac{HMSSA-V}{ARFIMA}$	$\frac{HMSSA-V}{BATS}$	$\frac{HMSSA-V}{BAR1}$	$\frac{HMSSA-V}{VAR}$	$\frac{HMSSA-V}{BVAR5}$	$\frac{HMSSA-V}{HMSSA-R}$
1	0.94	0.94	0.98	0.99	0.82*,†	0.93	0.97	0.93	0.94	0.97	0.97
2	0.87*,†	0.84*,†	0.90*,†	0.96†	0.80*,†	0.85*,†	0.85*,†	0.83*,†	0.87*,†	0.89*,†	0.99
3	0.88*	0.80*,†	0.89*,†	0.97	0.72*,†	0.80*,†	0.84*,†	0.79*,†	0.85*	0.87*	0.99
4	0.94	0.77*,†	0.89*	0.98	0.70*,†	0.78*,†	0.88*,†	0.75*,†	0.83*,†	0.86*	0.99
5	0.87*,†	0.72*,†	0.85*,†	0.95	0.72*,†	0.74*,†	0.83*,†	0.70*,†	0.79*,†	0.82*,†	0.99
6	0.88*,†	0.70*,†	0.84*,†	0.96	0.68*,†	0.72*,†	0.83*,†	0.67*,†	0.77*,†	0.80*,†	0.99
7	0.91*,†	0.69*,†	0.83*,†	0.95	0.63*,†	0.71*,†	0.86*	0.66*,†	0.76*,†	0.79*,†	0.99
8	0.94	0.69*,†	0.85*	0.96	0.69*,†	0.71*,†	0.88*	0.66*,†	0.77*,†	0.81*,†	0.99
9	0.91*	0.68*,†	0.83*	0.95	0.75*,†	0.70*,†	0.84*	0.64*,†	0.79*,†	0.82*	0.99
10	0.91*	0.64*,†	0.79*,†	0.91*	0.56*,†	0.66*,†	0.82*,†	0.60*,†	0.76*,†	0.80	0.95
Avg.	0.91*	0.75*	0.87*	0.96	0.71*	0.76*	0.86*	0.72*	0.81*	0.84*	0.99

*Indicates a statistically significant difference between the distribution of forecasts based on the two-sided KSPA test in [20] at $p = 0.10$.
†Indicates the existence of a statistically significant difference between the forecasts based on the modified Diebold–Mariano test in [21].

Kabundi [53,54], and Gupta et al. [55], as it allows us to provide an objective way of choosing the hyperparameters of the (subjective) prior.

The BVAR model is estimated using Theil's [56] mixed estimation technique, which involves supplementing the data with prior information on the distribution of the coefficients. In an artificial way, the number of observations and degrees of freedom are increased by one, for each restriction imposed on the parameter estimates. The loss of degrees of freedom due to over parameterization associated with a VAR model is, therefore, not a concern in the BVAR model. Further note that, one major advantage of the BVAR and BAR models is that we can use non-stationary data for its estimation. Sims et al. [57] indicate that with the Bayesian approach entirely based on the likelihood function, the associated inferences do not require special treatment for non-stationarity, since the likelihood function exhibits the same Gaussian shape regardless of the presence of non-stationarity. Given this, we mimic AR and VAR models by setting $w = 2.0$, $d = 2.0$, $k_{ij} = 0.001$, and $w = 2.0$, $d = 0$, $k_{ij} = 1.0$, respectively. In other words, we are able to estimate classical versions of AR and VAR models without worrying about ensuring stationarity of the variables under consideration.

A.2. All results

See Tables 7–9.

Table 9
DC predictions from out-of-sample forecasts for Global Temperature.

h	AR	ARIMA	ETS	NN	ARFIMA	BATS	BAR1	VAR	BVAR5	HMSSA-R	HMSSA-V
1	0.42	0.54	0.60	0.57	0.56	0.58	0.42	0.38	0.41	0.60	0.61
2	0.43	0.50	0.65	0.61	0.58	0.53	0.42	0.42	0.40	0.70*	0.70*
3	0.52	0.52	0.61	0.58	0.57	0.45	0.50	0.54	0.54	0.66*	0.69*
4	0.49	0.49	0.62	0.61	0.52	0.46	0.47	0.55	0.55	0.64	0.67*
5	0.49	0.51	0.66*	0.55	0.53	0.50	0.47	0.56	0.55	0.70*	0.67*
6	0.50	0.51	0.61	0.58	0.54	0.51	0.48	0.56	0.53	0.67*	0.67*
7	0.41	0.47	0.56	0.56	0.45	0.43	0.38	0.54	0.50	0.62	0.62
8	0.43	0.42	0.57	0.57	0.47	0.45	0.42	0.53	0.47	0.63	0.63
9	0.46	0.50	0.60	0.64	0.50	0.45	0.43	0.56	0.53	0.65	0.65
10	0.38	0.42	0.55	0.49	0.39	0.41	0.35	0.54	0.51	0.60	0.64
Avg.	0.45	0.49	0.60	0.58	0.51	0.48	0.43	0.52	0.50	0.65	0.66

*Indicates the DC prediction is statistically significant based on a Student's t test at $p = 0.10$.

References

[1] M.L. Khandekar, T.S. Murty, P. Chittibabu, The global warming debate: a review of the state of science, Pure Appl. Geophys. 162 (2005) 1557–1586.
[2] S. Solomon, G.-K. Plattner, R. Knutti, P. Friedlingstein, Irreversible climate change due to carbon dioxide emissions, Proc. Natl Acad. Sci. USA 106 (6) (2009) 1704–1709.
[3] R. Fildes, N. Kourentzes, Validation and forecasting accuracy in models of climate change, Int. J. Forecast. 27 (2011) 968–995.
[4] D.G. McMillan, M.E. Wohar, The relationship between temperature and CO_2 emissions: evidence from a short and very long dataset, Appl. Econ. 45 (2013) 3683–3690.
[5] M. Beenstock, Y. Reingerwertz, N. Paldor, Testing the historic tracking of climate models, Int. J. Forecast. 32 (4) (2016) 1234–1246.
[6] K.C. Green, J.S. Armstrong, Global warming: forecasts by scientists versus scientific forecasts, Energy Environ. 18 (2007) 997–1021.
[7] T. Dergiades, R. Kaufmann, T. Panagiotidis, Long-run changes in radiative forcing and surface temperature: the effect of human activity over the last five centuries, J. Environ. Econ. Manage. 76 (2016) 67–85.
[8] J. Bai, P. Perron, Computation and analysis of multiple structural change models, J. Appl. Econometrics 18 (2003) 1–22.
[9] P.E. McSharry, Validation and forecasting accuracy in models of climate change: Comments, Int. J. Forecast. 27 (2011) 996–999.
[10] D.A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D.J. Frame, J.A. Kettleborough, S. Knight, A. Martin, J.M. Murphy, C. Piani, D. Sexton, L.A. Smith, R.A. Spicer, A.J. Thorpe, M.R. Allen, Uncertainty in predictions of the climate response to rising levels of greenhouse gases, Nature 433 (2005) 403–406.
[11] D.A. Stainforth, M.R. Allen, E.R. Tredger, L.A. Smith, Confidence, uncertainty and decision-support relevance in climate predictions, Phil. Trans. R. Soc. A Math. Phys. Eng. Sci. 365 (2007) 2145–2161.
[12] D.M. Smith, S. Cusack, A.W. Colman, C.K. Folland, G.R. Harris, J.M. Murphy, Improved surface temperature prediction for the coming decade from a global climate model, Science 317 (2007) 796–799.
[13] R.K. Kaufmann, D.I. Stern, A Statistical Evaluation of Atmosphere-Ocean General Circulation Models: Complexity V Simplicity. Rensselaer Working Papers in Economics, No. 0411, 2004.
[14] J.Y. Campbell, Viewpoint: estimating the equity premium, Canad. J. Econ. 41 (2008) 1–21.
[15] S. Sanei, H. Hassani, Singular Spectrum Analysis of Biomedical Signals, CRC Press, United States, 2015.
[16] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, in: Lecture Notes in Mathematics, vol. 898, Springer-Verlag, 1981, pp. 366–381.
[17] M. Ghil, M.R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, P. Yiou, Advanced spectral methods for climatic time series, Rev. Geophys. 40 (2002) 1–41.
[18] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D 35 (1989) 395–424.
[19] M. Ghil, R. Vautard, Interdecadal oscillations and the warming trend in global temperature time series, Nature 350 (1991) 324–327.
[20] H. Hassani, E.S. Silva, A Kolmogorov–Smirnov based test for comparing the predictive accuracy of two sets of forecasts, Econometrics 3 (2015) 590–609.
[21] D.I. Harvey, S.J. Leybourne, P. Newbold, Testing the equality of prediction mean squared errors, Int. J. Forecast. 13 (1997) 281–291.
[22] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for R, J. Stat. Softw. 27 (2008) 1–22.
[23] H. Hassani, A. Webster, E.S. Silva, S. Heravi, Forecasting U.S. tourist arrivals using optimal singular spectrum analysis, Tourism Manage. 46 (2015) 322–335.
[24] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, OTexts, Australia, 2013, http://OTexts.com/fpp.
[25] S.G. Makridakis, S.C. Wheelwright, R.J. Hyndman, Forecasting: Methods and Applications, third ed., Wiley, New York, 1998.
[26] J. Haslett, A.E. Raftery, Space–time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion), Appl. Stat. 38 (1989) 1–50.
[27] A.M. De Livera, R.J. Hyndman, R.D. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, J. Amer. Statist. Assoc. 106 (2011) 1513–1527.
[28] F. Lisi, A. Medio, Is a random walk the best exchange rate predictor? Int. J. Forecast. 13 (1997) 255–267.
[29] D.D. Thomakos, T. Wang, L.C. Wille, Modeling daily realized futures volatility with singular spectrum analysis, Physica A 312 (2002) 505–519.
[30] C. Beneki, B. Eeckels, C. Leon, Signal extraction and forecasting of the uk tourism income time series: a singular spectrum analysis approach, J. Forecast. 31 (2012) 391–400.
[31] H. Hassani, E.S. Silva, R. Gupta, M.K. Segnon, Forecasting the price of gold, Appl. Econ. 47 (2015) 4141–4152.
[32] E.S. Silva, H. Hassani, On the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recession, Int. Econ. 141 (2015) 34–49.
[33] H. Hassani, R. Mahmoudvand, Multivariate singular spectrum analysis: A general view and new vector forecasting approach, Int. J. Energy Stat. 1 (2013) 55–83.
[34] H. Hassani, A.S. Soofi, A.A. Zhigljavsky, Predicting daily exchange rate with singular spectrum analysis, Nonlinear Anal. RWA 11 (2010) 2023–2034.
[35] K. Patterson, H. Hassani, S. Heravi, A. Zhigljavsky, Multivariate singular spectrum analysis for forecasting revisions to real-time data, J. Appl. Stat. 38 (2011) 2183–2211.
[36] V. Oropeza, M. Sachchi, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis, Geophysics 76 (2011) V25–V32.

[37] A. Groth, M. Ghil, Multivariate singular spectrum analysis and the road to phase synchronization, Phys. Rev. E 84 (2012) 036206.
[38] H. Hassani, H. Heravi, A. Zhigljavsky, Forecasting UK industrial production with multivariate singular spectrum analysis, J. Forecast. 32 (2013) 395–408.
[39] H. Hassani, A.S. Soofi, A. Zhigljavsky, Predicting inflation dynamics with singular spectrum analysis, J. Roy. Statist. Soc. Ser. A 176 (2013) 743–760.
[40] J. Hansen, R. Ruedy, M. Sato, K. Lo, Global surface temperature change, Rev. Geophys. 48 (2010) RG4004, 1–29.
[41] United Nations, 2011 Energy Statistics Yearbook, United Nations Department for Economic and Social Information and Policy Analysis, Statistics Division, New York, 2014.
[42] T.A. Boden, G. Marland, R.J. Andres, Estimates of Global, Regional, and National Annual CO_2 Emissions from Fossil-Fuel Burning, Hydraulic Cement Production, and Gas Flaring: 1950-1992. *ORNL/CDIAC-90, NDP-30/R6*, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 1995.
[43] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, Int. J. Forecast. 14 (1998) 35–62.
[44] C. Altavilla, P. De Grauwe, Forecasting and combining competing models of exchange rate determination, Appl. Econ. 42 (2010) 3455–3480.
[45] H. Hassani, H. Heravi, A. Zhigljavsky, Forecasting European industrial production with singular spectrum analysis, Int. J. Forecast. 25 (2009) 103–118.
[46] W. Brock, D. Dechert, J. Scheinkman, B. LeBaron, A test for independence based on the correlation dimension, Econometric Rev. 15 (1996) 197–235.
[47] H.J. Bierens, L.F. Martins, Time varying cointegration, Econometric Theory 26 (2010) 1453–1490.
[48] United Nations Intergovernmental Panel on Climate Change (IPCC), Fifth Climate Change Synthesis Report, 2014, Available for download from: https://www.ipcc.ch/report/ar5/syr/.
[49] H. Hassani, X. Huang, R. Gupta, M. Ghodsi, Does sunspot numbers cause global temperatures? A reconsideration using non-parametric causality tests, Physica A 460 (C) (2016) 54–65.
[50] C.A. Sims, Macroeconomics and reality, Econometrica 48 (1980) 1–48.
[51] R.B. Litterman, A Bayesian Procedure for Forecasting with Vector Autoregressions. Working Paper, Federal Reserve Bank of Minneapolis, 1981.
[52] R.B. Litterman, Forecasting with bayesian vector autoregressions: Five years of experience, J. Bus. Econom. Statist. 4 (1986) 25–38.
[53] R. Gupta, A. Kabundi, Forecasting macroeconomic variables in a small open economy: a comparison between small- and large-scale models, J. Forecast. 29 (1–2) (2010) 168–185.
[54] R. Gupta, A. Kabundi, A large factor model for forecasting macroeconomic variables in South Africa, Int. J. Forecast. 27 (4) (2011) 1076–1088.
[55] R. Gupta, A. Kabundi, S.M. Miller, J. Uwilingiye, Using large data sets to forecast sectoral employment, Stat. Methods Appl. 23 (2) (2014) 229–264.
[56] H. Theil, Principles of Econometrics, John Wiley, New York, 1971.
[57] C.A. Sims, J.H. Stock, M.W. Watson, Inference in linear time series models with some unit roots, Econometrica 58 (1990) 113–144.