



University of Pretoria
Department of Economics Working Paper Series

Predicting Global Temperature Anomaly: A Definitive Investigation Using an Ensemble of Twelve Competing Forecasting Models

Hossein Hassani

Bournemouth University

Emmanuel Sirimal Silva

Bournemouth University

Rangan Gupta

University of Pretoria

Sonali Das

Council for Scientific and Industrial Research and Nelson Mandela Metropolitan University

Working Paper: 2015-61

August 2015

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Predicting Global Temperature Anomaly: A Definitive Investigation Using an Ensemble of Twelve Competing Forecasting Models

Hossein Hassani

Statistical Research Centre, The Business School, Bournemouth University, UK. Email: hassani.stat@gmail.com

Emmanuel Sirmal Silva

Statistical Research Centre, The Business School, Bournemouth University, UK. Email: esilva@bournemouth.ac.uk

Rangan Gupta

Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: rangan.gupta@up.ac.za.

Sonali Das

Advanced Mathematical Modelling, Modelling and Digital Science,

Council for Scientific and Industrial Research, P.O. Box 395, Pretoria, 0001, South Africa. Email: sdas@csir.co.za.

and

Department of Statistics, PO Box 77000, Nelson Mandela Metropolitan University,

Port Elizabeth, 6031, South Africa.

August 27, 2015

Abstract

In this paper we analyze whether (anthropometric) CO_2 can forecast global temperature anomaly (GT) over an annual out-of-sample period of 1907-2012, using an in-sample of 1880-1906. For our purpose, we use 12 parametric and non-parametric univariate (only comprising of GT) and multivariate (including both GT and CO_2) models. Our results show that the Horizontal Multivariate Singular Spectral Analysis (HMSSA) models (both Recurrent (-R) and Vector (-V)) consistently outperform the other competing models. More importantly, from the performance of the HMSSA-R model, we find conclusive evidence that CO_2 can forecast GT, and predict its direction of change. Our results 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.

JEL Codes: C22; C32; C53; Q53; Q54.

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

1 Introduction

The climate change debate comprises 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 (Khandekar et al., 2005). The debate gets even more aggressive when global warming is discussed in relation to anthropometric carbon di-oxide (CO_2) emissions (Solomon et al., 2009; McMillan and Wohar, 2013). Global warming is popularly quantified using global temperature anomaly (GT) measures which is the difference between a reference long-term average value

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 (Hansen et al., 2010). Similarly global CO₂ is estimated from energy statistics published by the United Nations (2014) involving a complex system comprising of, again not limited to, questionnaires, official statistics and other supplementary information (Boden et al., 1995).

While global warming has been accepted as happening (Hansen et al., 1981), the purpose of this paper is to undertake a rigorous investigation of well established datasets for GT and CO₂ 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), Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (TBATS), Bayesian Autoregression (BAR), Vector Autoregression (VAR), Bayesian Vector Autoregression (BVAR) and Multivariate Singular Spectrum Analysis (MSSA). Our sample covers the annual period of 1880-2012, with an out-of-sample period of 1907-2012 (based on an in-sample period of 1880 to 1906). 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 (2003) tests of multiple structural breaks) in the relationship between GT and CO₂, which happened to be 1906. Since, we estimate our models recursively over the out-of-sample period, we are able accommodate for the change in the parameter estimates of the model while producing our forecasts. This approach, in some sense makes the linear and nonlinear approaches comparable, by allowing all the structural breaks to be in the out-of-sample, over which all types of models are re-estimated over an expanding window.

This paper makes several important contributions: (i) this paper makes the first attempt to provide a comprehensive comparisons of models for forecasting GT available in the literature; (ii) it marks the introductory application of models such as TBATS, SSA and MSSA for forecasting GT; and, (iii) the approach presented in this paper is a rare occasion in which univariate SSA's filtering capabilities are combined with the modelling and forecasting capabilities of MSSA in order to find a solution to the problems associated with forecasting GT. Summarily, in our investigation, we seek to isolate the trend in GT via SSA, and then propose using the extracted GT trend along with CO₂ data via MSSA for generating out-of-sample forecasts for GT. Finally, we use a new and automated MSSA forecasting algorithm for generating out-of-sample forecasts for GT. To the best of our knowledge, the only other paper to have analyzed the relationship between CO₂ and GT is that of McMilan and Wohar (2013), barring which all other papers in this literature have analyzed these two variables separately in univariate settings.¹ McMilan and Wohar (2013), based on VAR and Generalized Method of Moments (GMM) approaches conclude that CO₂ has a weak relationship with GT. This could, however, be a result of uncaptured nonlinearity in the relationship between these two variables, which is what we aim to understand better in this paper by conducting an out-of-sample forecasting exercise, using univariate and multivariate versions of linear and nonlinear models. The decision to rely on an out-of-sample rather than an in-sample predictability exercise to gauge the relationship between CO₂ and

¹The reader is referred to McMilan and Wohar (2013) for a detailed literature review in this regard.

global temperature is motivated out of the belief that: “The ultimate test of any predictive model is its out-of-sample performance” (Campbell, 2008).

At this stage, it is important to emphasize the decision to use the SSA technique in forecasting GT, which has recently evolved as a powerful technique in the field of time series analysis (Hassani, 2007; Hassani et al., 2009), besides the other standard forecasting approaches indicated above. SSA is a non-parametric 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 not sensitive to the dynamical variations. Moreover, 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), SSA method is non-parametric and makes no prior assumptions about the data, with forecasts being obtained through bootstrapping. Additionally, SSA method 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. The rest of the paper is organised as follows: in Section 2 we present a detailed description of the 12 forecasting models investigated with extra emphasis on the SSA and MSSA models; in Section 3 we detail the datasets used and the metrics used for the evaluation of the models, with estimations conducted in either R or RATS; in Section 4 we present an in-depth analyses of the results in our quest to identify the best model for our purpose; and finally we present some concluding remarks on the investigation undertaken in Section 5.

2 Methodology

2.1 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 RW². In brief, the RW model states that today’s GT is the best forecast for tomorrow’s GT.

2.2 Autoregressive Integrated Moving Average (ARIMA)

In this paper we use an optimized version of ARIMA which is referred to as auto-ARIMA, and provided through the forecast package in R. A more detailed description of the algorithm underlying auto-ARIMA can be found in Hyndman and Khandakar (2008) whilst a summarized version is available in Hassani et al. (2015b). The process for obtaining point forecasts using the R software is concisely presented in Hyndman and Athanasopoulos (2013).

2.3 Exponential Smoothing (ETS)

The ETS technique in the forecast package overcomes the limitations of the Makridakis et al. (1998) algorithm pertaining to the calculation of prediction intervals. Whilst a detailed description of the ETS technique can be found in Hyndman and Athanasopoulos (2013), in brief this algorithm considers the error, trend and seasonal components along with over 30 possible

²<http://robjhyndman.com/hyndsight/benchmarks/>

options for choosing the best exponential smoothing model via optimization of initial values and parameters using the MLE and selecting the best model based on the AIC.

2.4 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 Hyndman et al. (2013). It may be noted that in all cases the selected neural network model has only $k=1$ hidden node, $p=2$ lags and we adopt annual difference specifications.

2.5 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 (2008) algorithm whilst d and parameters are selected based on the Haslett and Raftery (1989) algorithm.

2.6 Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (TBATS)

The TBATS model is a technique aimed at providing accurate forecasts for time series with complex seasonality. A detailed description of the model can be found in De Livera et al. (2011) and is therefore not reproduced here.

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 (1980), can be written as follows:

$$y_t = C + A(L)y_t - 1 + \epsilon_t \quad (1)$$

where: y : ($n \times 1$) vector of variables (global temperatures and global CO₂ emissions) being forecasted; $A(L)$: ($n \times p$) polynomial matrix in the backshift operator L with lag length p , i.e., $A(L) = A_1L + A_2L^2 + \dots + A_pL^p$; C : ($n \times 1$) vector of constant terms, and ϵ : ($n \times 1$) 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 Littermann (1981, 1986), Doan et al (1984), Todd (1984), and Spencer (1993), 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 (1981) 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:

$$S(i, j, m) = [w \times g(m) \times f(i, j)] \frac{\sigma_i}{\sigma_j} \quad (2)$$

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 extant 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, which in our case, happened to be the BAR1 ($w=0.3, d=0.5, k_{ij}=0.001$) and BVAR5 ($w=0.1, d=2.0, k_{ij}=0.5$).³

The BVAR model is estimated using Theil’s (1971) 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. (1990) 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.

2.7 Singular Spectrum Analysis (SSA)

The SSA technique is now a popular filtering and forecasting technique which is exploited in a variety of fields (see for example, Hassani et al. (2009), Hassani et al. (2010a,b), Hassani et al. (2015a), Silva and Hassani (2015)). 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 (Hassani et al. 2015a). SSA also has its multivariate form which is referred to as Multivariate SSA (MSSA) which can be used for modelling and forecasting using multiple series. In comparison to SSA,

³Complete details of the results from the other BAR and BVAR models are available upon request from the authors.

MSSA is relatively new with few applications (see for example, Hassani et al. (2010c), Patterson et al. (2011), Oropeza and Sachchi (2011), Hassani et al. (2013a,b)).

There are several benefits of using SSA and MSSA models. As these models are non-parametric they are not bound by the parametric assumptions of normality, stationarity and linearity (Hassani et al. (2013a)). As such, one is able to model the data without any transformations which in turn ensures there is no loss of information (Hassani et al. 2013b). 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 (2015). In brief, the SSA technique is made up of two stages known as Decomposition and Reconstruction. The rationale behind using SSA here is not for forecasting, but for extracting the trend in GT. Below we present the filtering and reconstruction stages of SSA, and in doing so we mainly follow Sanei and Hassani (2015).

Stage 1: Decomposition

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

Step 1: Embedding

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

Step 2: Singular Value Decomposition (SVD)

In this step we obtain the singular value decomposition of the trajectory matrix. Denote by $\lambda_1, \dots, \lambda_L$ the eigenvalues of $\mathbf{X}\mathbf{X}'$ in decreasing order of magnitude ($\lambda_1 \geq \dots \lambda_L \geq 0$) and by U_1, \dots, 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:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad (3)$$

where $\mathbf{X}_i = \sqrt{\lambda_i}U_iV_i'$ ($i = 1, \dots, d$).

Stage 2: Reconstruction

The second and final parameter in SSA, i.e. the number of eigenvalues, r is required at this stage. Note that when L is sufficiently large the first eigenvalue corresponds to the trend of a given series.

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 $I = \{i_1, \dots, i_p\}$ be a group of indices i_1, \dots, i_p . Then the matrix \mathbf{X}_I corresponding to the group I is defined as $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. The split of the set of indices $J = 1, \dots, d$ into the disjoint subsets I_1, \dots, I_m corresponds to the representation

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (4)$$

The procedure of choosing the sets I_1, \dots, 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 (4), we obtain another expansion:

$$\mathbf{X} = \tilde{\mathbf{X}}_{I_1} + \dots + \tilde{\mathbf{X}}_{I_m} \quad (5)$$

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

$$y_n = \sum_{k=1}^m \tilde{y}_n^{(k)} \quad (6)$$

where $\tilde{Y}_N^{(k)} = (\tilde{y}_1^{(k)}, \dots, \tilde{y}_N^{(k)})$ corresponds to the matrix \mathbf{X}_{I_k} . Figure 1 below shows the extracted GT trend using the SSA procedure explained above. For this purpose we use $L = 18$, $r = 1$. In the next step, we use this extracted GT trend and actual CO₂ data with MSSA for forecasting GT.

2.8 Multivariate Singular Spectrum Analysis (MSSA)

Those interested in an in-depth explanation of the theory underlying MSSA are directed to Hassani and Mahmoudvand (2013). 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 Hassani and Mahmoudvand (2013).

2.8.1 HMSSA-R Optimal Forecasting Algorithm

1. Consider M time series with identical series lengths of N_i , such that $Y_{N_i}^{(i)} = (y_1^{(i)}, \dots, y_{N_i}^{(i)})$ ($i = 1, \dots, M$).
2. Split each time series into three parts leaving $\frac{2}{3}^{rd}$ for model training and testing, and $\frac{1}{3}^{rd}$ for validation. In this case we use 25 observations from extracted GT trend and CO₂ 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 $Y_{N_i}^{(i)}$, using the training data construct the trajectory matrix $\mathbf{X}^{(i)} = [X_1^{(i)}, \dots, X_K^{(i)}] = (x_{mn})_{m,n=1}^{L, K_i}$ for each single series $Y_{N_i}^{(i)}$ ($i = 1, \dots, M$) separately.

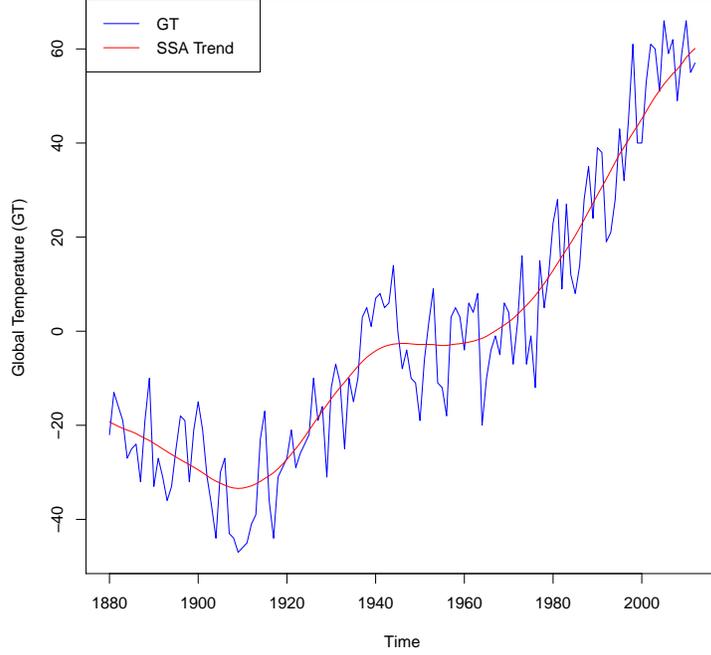


Figure 1: SSA trend for GT (1880–2012).

4. Then, construct the block trajectory matrix \mathbf{X}_H as follows:

$$\mathbf{X}_H = \left[\mathbf{X}^{(1)} : \mathbf{X}^{(2)} : \dots : \mathbf{X}^{(M)} \right].$$

5. Let vector $U_{H_j} = (u_{1j}, \dots, u_{Lj})^T$, with length L , be the j^{th} 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:

$$\mathbf{X}_H = \left[\widehat{\mathbf{X}}^{(1)} : \widehat{\mathbf{X}}^{(2)} : \dots : \widehat{\mathbf{X}}^{(M)} \right].$$

7. Consider matrix $\widetilde{\mathbf{X}}^{(i)} = \mathcal{H}\widehat{\mathbf{X}}^{(i)}$ ($i = 1, \dots, M$) 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 $U_{H_j}^\nabla$ 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} ($j = 1, \dots, r$).

9. Define $v^2 = \sum_{j=1}^r \pi_{H_j}^2$.

10. Denote the linear coefficients vector \mathcal{R} as follows:

$$\mathcal{R} = \frac{1}{1 - v^2} \sum_{j=1}^r \pi_{Hj} U_{Hj}^\nabla. \quad (7)$$

11. If $v^2 < 1$, then the h -step ahead HMSSA forecasts exist and is calculated by the following formula:

$$\left[\hat{y}_{j_1}^{(1)}, \dots, \hat{y}_{j_M}^{(M)} \right]^T = \begin{cases} \left[\tilde{y}_{j_1}^{(1)}, \dots, \tilde{y}_{j_M}^{(M)} \right], & j_i = 1, \dots, N_i, \\ \mathcal{R}^T \mathbf{Z}_h, & j_i = N_i + 1, \dots, N_i + h, \end{cases} \quad (8)$$

where, $\mathbf{Z}_h = \left[Z_h^{(1)}, \dots, Z_h^{(M)} \right]^T$ and $Z_h^{(i)} = \left[\hat{y}_{N_i - L + h + 1}^{(i)}, \dots, \hat{y}_{N_i + h - 1}^{(i)} \right]$ ($i = 1, \dots, M$).

12. Seek the combination of L and r which minimises 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 $\frac{1}{3}^{rd}$ observations.

2.8.2 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

$$\mathbf{\Pi} = \mathbf{U}^\nabla \mathbf{U}^{\nabla T} + (1 - v^2) \mathbf{R} \mathbf{R}^T, \quad (9)$$

where $\mathbf{U}^\nabla = [U_1^\nabla, \dots, U_r^\nabla]$. Now consider the linear operator

$$\mathcal{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbb{R}^L, \quad (10)$$

where

$$\mathcal{P}^{(v)} Y = \begin{pmatrix} \mathbf{\Pi} Y_\Delta \\ \mathbf{R}^T Y_\Delta \end{pmatrix}, \quad Y \in \mathfrak{L}_r, \quad (11)$$

and Y_Δ is vector of last $L - 1$ elements of Y .

3. Define vector $Z_j^{(i)}$ ($i = 1, \dots, M$) as follows:

$$Z_j^{(i)} = \begin{cases} \tilde{X}_j^{(i)} & \text{for } j = 1, \dots, k_i \\ \mathcal{P}^{(v)} Z_{j-1}^{(i)} & \text{for } j = k_i + 1, \dots, k_i + h + L - 1 \end{cases} \quad (12)$$

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

4. Now, by constructing matrix $\mathbf{Z}^{(i)} = [Z_1^{(i)}, \dots, Z_{k_i + h + L - 1}^{(i)}]$ and performing diagonal averaging we obtain a new series $\hat{y}_1^{(i)}, \dots, \hat{y}_{N_i + h + L - 1}^{(i)}$, where $\hat{y}_{N_i + 1}^{(i)}, \dots, \hat{y}_{N_i + h}^{(i)}$ 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 model parameters for obtaining the forecasts for GT at each horizon.

Table 1: MSSA models for forecasting Global Temperature

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

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

3 Data and Metrics

3.1 Data

The data we investigate here consists of two variables, namely the global temperatures anomaly (GT) and global carbon dioxide (CO₂) 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. The GT were obtained from the *National Aeronautics and Space Administrations* (NASA) and the *Goddard Institute for Studies* (GISS). As mentioned earlier, the GT data relates to temperature anomalies relative to the base period 1951-1980. The data on CO₂ was obtained from the *Carbon Dioxide Information Analysis Centre*, and is measured in thousand metric tons of carbon. While GT remains untransformed in our analysis, we use the natural logarithm of CO₂ emissions. Further, we separate the entire data period into an in-sample period spanning 1880-1906, and an out-of-sample period of 1907-2012. This separation of the period was determined by the Bai and Perron (2003) 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 models are estimated recursively over the out-of-sample period in which all the remaining break dates fall, this separation of the period is ideal as it helps us to accommodate for changes in the parameter estimates of the model in the out-of-sample period. Finally, we produce one- to ten-year-ahead forecasts based on this recursive estimation scheme to forecast in the short- and long-term horizon.

3.2 Metrics

To evaluate the 12 competing models, we use the popular 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 (Hassani and Silva, 2015), 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, Altavilla and De Grauwe (2010) and Hassani et al. (2015b)). Here, we mainly follow Altavilla and Grauwe (2010) in defining the RMSE.

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^N e_{t+h+i}^2 \right)^{1/2},$$

where, $e_{t+k} = \chi_{t+k} - \hat{\chi}_{t+k}$ is the forecast error where $h \geq 1$, and $\hat{\chi}_{t+h}$ represents the h -step-ahead forecast.

In addition to the RMSE, we also consider the Ratio of the RMSE (RRMSE) criterion. In order to save space we only present an example of the RRMSE criterion.

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

where, \hat{y}_{T+h} is the h -step ahead forecast obtained by RW, \tilde{y}_{T+h} is the h -step ahead forecast from the ETS model, and N is the number of the forecasts. If $\frac{RW}{ETS}$ is less than 1, then the RW outperforms ETS by $1 - \frac{RW}{ETS}$ 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% (Altavilla and de Grauwe, 2010). A detailed description of the DC metric can be found in Altavilla and Grauwe (2010). In brief,

$$\text{DC} = \left(\frac{1}{T} \sum_{t=1}^T \phi \Delta_{t+h}^e = \Delta_{t+h} \right),$$

where, Δ_{t+h} and Δ_{t+h}^e are the actual and predicted direction of change in GT h steps ahead, and ϕ equals 1 if $\Delta_{t+h} = \Delta_{t+h}^e$ and 0 otherwise.

4 Empirical Results

Table 2 presents the out-of-sample forecasting RMSE results for GT. 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 of interest. We first discuss the overall results. The first observation is that no single model can provide the best forecast for GT across all horizons. However, it is moot to note that the two MSSA models succeed in reporting the lowest RMSE in comparison to all models evaluated 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. Secondly, given that these superior MSSA results have been attained by using CO₂ as an indicator variable, we are able to conclude that CO₂ 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 we can propose that specifically the HMSSA-R model, 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. Interestingly as well, on most occasions the RW forecasts are seen to outperform the univariate models (except ETS), and also the BAR1, 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 multivariate BAR1 model is almost as bad as the worst performing NN model according to the average RMSE criterion. 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.

Prior to discussing the results over the forecasting horizon, we test the out-of-sample forecasts for statistically significant differences by using the one-sided Kolmogorov-Smirnov Predictive Accuracy (KSPA) test (Hassani and Silva, 2015). The KSPA test evaluates whether the model with the lowest RMSE also reports a lower stochastic error in comparison to a competing model. The results from the one-sided KSPA test are indicated on Table 2. Based on the test results, we find no statistically significant differences between the forecasts from HMSSA-R and HMSSA-V models for any horizon at the 10% significance level. This was expected as there is a very meagre difference between the RMSE's from the two MSSA models. In relation to the other models, we notice a high number of statistically significant outcomes at the 10% significance level beyond the $h = 1$ step-ahead horizon, suggesting that on most occasions, the HMSSA-R model is indeed reporting a lower stochastic error than the other models. In fact, at horizons 2-3 and 5-10 the HMSSA-R model reports a lower stochastic error in comparison to all other univariate models (except HMSSA-V). Therefore, we are able to suggest the HMSSA-R model as the best model from those evaluated here for forecasting GT at these horizons. However, at $h = 1$ step-ahead, where the HMSSA-V model reports the lowest RMSE, we can only suggest that the HMSSA-V model outperforms NN, ARFIMA and TBATS in terms of reporting the comparatively lowest stochastic error. This suggests that at $h = 1$ step-ahead, there is a high probability of chance occurrences causing differences in forecasting results between the MSSA models and RW, AR, ARIMA, ETS, BAR1, VAR and BVAR5 models. As such, in the very short run ($h = 1$) we do not find sufficient evidence to suggest that any particular model evaluated here is superior to another for forecasting GT based on the one-sided KSPA test.

In Table 3 we consider HMSSA-R as a benchmark model and present the RRMSE results for GT forecasts. The choice of HMSSA-R as a benchmark model was influenced by the fact that the one-sided KSPA test reported a high number of statistically significant outcomes proving the HMSSA-R forecasts tend to report stochastically smaller errors than majority of the competing models at majority of the forecasting horizons, and as the HMSSA-R model reports the lowest average RMSE across all ten forecasting horizons. Here, once again we exploit the KSPA test which is based on the cumulative distribution functions (c.d.f) of the forecast errors (Hassani and Silva, 2015). In this case we use it to test the overall performance of the models across all 10 forecasting horizons in order to ascertain whether on average one model is significantly better than another. The empirical c.d.f. of squared out-of-sample forecasting errors across all horizons are shown in Figure 2. According to Hassani et al. (2009), if the c.d.f. of forecast errors from one model lies strictly above and towards the left of the c.d.f. of forecast errors from

another model, then the model lying on the left reports a lower stochastic error. However, it is not possible to draw conclusions from Figure 2 alone, and as such we call on the one-sided KSPA test. Evidence from the one-sided KSPA test (results indicated on the ‘Avg.’ row in Table 3) suggests that on average, across all 10 horizons, the only instance in which HMSSA-R fails to report a stochastically smaller error than another model is in comparison to HMSSA-V forecasts at a 90% confidence level. This provides added justification for selecting the HMSSA-R model proposed in this study as a benchmark for forecasting GT.

The RRMSE results in Table 3 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 HMSSA-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-R forecasts are indeed significantly better than the HMSSA-V forecasts. 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-R forecasts are 27%, 49%, 40%, 33%, 52%, 48%, 40%, 51%, 44%, 42% and 2% better than forecasts from RW, AR, ARIMA, ETS, NN, ARFIMA, TBATS, BAR1, VAR, BVAR5 and HMSSA-V models respectively.

Presented in Table 4 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.⁴ The DC results indicate that on average the multivariate 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-R model with an average accuracy of 78%. 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 Hassani et al. (2009). Accordingly, we see that the DC predictions reported for AR, ARIMA, ARFIMA, TBATS, BAR1, VAR, and BVAR5 models at all horizons are likely to be a result of chance occurrences. However, some of the DC predictions for ETS and NN models are found to be statistically significant whilst all the DC predictions from the two MSSA models report statistically significant outcomes across all horizons. As such, we are able to conclude with 95% confidence that the MSSA DC predictions are significantly greater than 50% across all horizons.

⁴Note that it is not possible to calculate the DC metric for RW forecasts as it results in the DC statistic going to infinity because the RW is simply today’s value considered to be tomorrow’s forecast.

Table 2: RMSE for out-of-sample forecasts for Global Temperature.

h	RW	AR	ARIMA	ETS	NN	ARFIMA	TBATS	BARI	VAR	BVAR5	HMSSA-R	HMSSA-V
1	10.66	10.68	10.21	10.00	12.25*	10.76*	10.28*	10.70	10.65	10.34	8.76	8.74
2	12.61*	13.10*	12.14*	11.44*	13.78*	12.98*	12.86*	13.20*	12.61*	12.32*	8.77	9.17
3	13.22*	14.51*	13.10*	12.01*	16.25*	14.51*	13.92*	14.75*	13.75*	13.43*	8.71	9.44
4	12.89*	15.64*	13.62*	12.27	17.20*	15.51*	13.77*	16.03*	14.57*	13.97*	9.87	9.87
5	14.61*	17.55*	14.88*	13.28*	17.51*	17.14*	15.34*	18.14*	16.12*	15.54*	8.64	8.65
6	14.82*	18.77*	15.53*	13.64*	19.13*	18.17*	15.73*	19.54*	17.07*	16.37*	8.69	8.69
7	14.89*	19.69*	16.31*	14.29*	21.76*	19.11*	15.80*	20.59*	17.86*	17.23*	8.86	8.87
8	15.44*	21.11*	17.06*	15.20*	21.08*	20.52*	16.50*	22.22*	18.82*	18.03*	9.23	9.28
9	17.05*	22.72*	18.72*	16.29*	20.53*	22.27*	18.40*	24.08*	19.69*	18.83*	9.34	9.54
10	16.92*	23.89*	19.43*	16.87*	27.29*	23.21*	18.76*	25.45*	20.26*	19.31*	9.29	9.77
Avg.	14.31*	17.77*	15.10*	13.53*	18.68*	17.42*	15.14*	18.47*	16.14*	15.54*	9.02	9.20

Note: Shown in bold font is the model reporting the lowest RMSE. The multivariate models use CO₂ at lag 2. * indicates that the model shown in bold font reports a lower stochastic error than the indicated model based on the one-sided KSPA test in Hassani and Silva (2015) at $p = 0.10$.

Table 3: RRMSE for out-of-sample forecasts for Global Temperature.

h	$\frac{HMSSA-R}{RW}$	$\frac{HMSSA-R}{AR}$	$\frac{HMSSA-R}{ARIMA}$	$\frac{HMSSA-R}{ETS}$	$\frac{HMSSA-R}{NN}$	$\frac{HMSSA-R}{ARFIMA}$	$\frac{HMSSA-R}{TBATS}$	$\frac{HMSSA-R}{BARI}$	$\frac{HMSSA-R}{VAR}$	$\frac{HMSSA-R}{BVAR5}$	$\frac{HMSSA-R}{HMSSA-V}$
1	0.82	0.82	0.86	0.88	0.72*	0.81	0.85	0.82	0.82	0.85	1.01
2	0.70*	0.67*	0.72*	0.77*	0.64*	0.68*	0.68*	0.66*	0.70*	0.71*	0.96
3	0.66*	0.60*	0.66*	0.73*	0.54*	0.60*	0.63*	0.59*	0.63*	0.65*	0.92
4	0.77	0.63*	0.72*	0.80	0.57*	0.64*	0.72*	0.62*	0.68*	0.71*	1.00
5	0.59*	0.49*	0.58*	0.65*	0.49*	0.50*	0.56*	0.48*	0.54*	0.56*	0.99
6	0.59*	0.46*	0.56*	0.64*	0.45*	0.48*	0.55*	0.44*	0.51*	0.53*	1.00
7	0.60*	0.45*	0.54*	0.62*	0.41*	0.46*	0.56*	0.43*	0.50*	0.51*	0.99
8	0.60*	0.44*	0.54*	0.61*	0.44*	0.45*	0.56*	0.42*	0.49*	0.51*	0.99
9	0.55*	0.41*	0.50*	0.57*	0.45*	0.42*	0.51*	0.39*	0.47*	0.50*	0.98
10	0.55*	0.39*	0.48*	0.55*	0.34*	0.40*	0.50*	0.37*	0.46*	0.48*	0.95
Avg.	0.63*	0.51*	0.60*	0.67*	0.48*	0.52*	0.60*	0.49*	0.56*	0.58*	0.98

Note: * indicates a statistically significant difference between the distribution of forecasts based on the two-sided KSPA test in Hassani and Silva (2015) at $p = 0.10$.

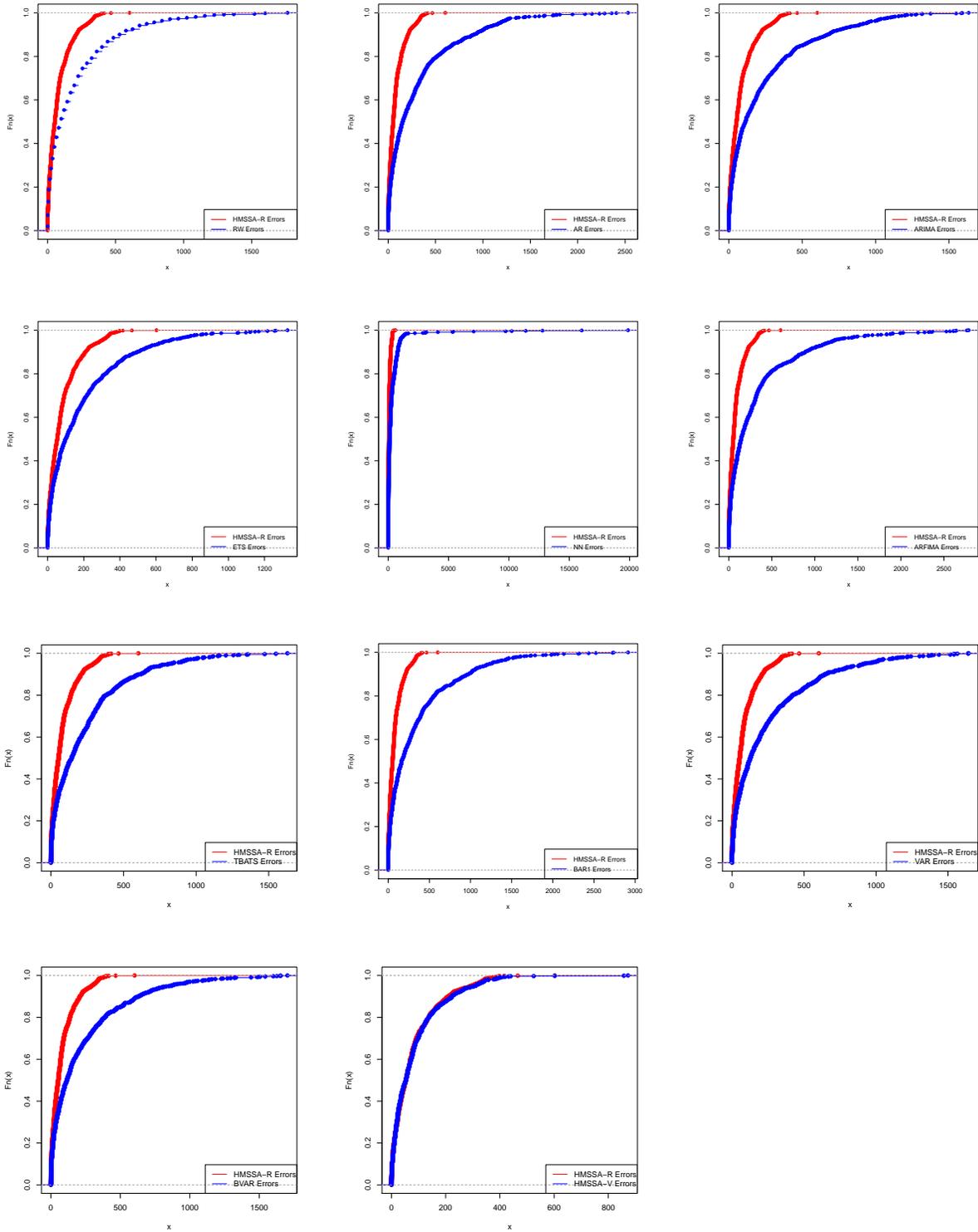


Figure 2: The cumulative distribution functions of the squared out-of-sample errors.

Table 4: DC predictions from out-of-sample forecasts for Global Temperature.

h	AR	ARIMA	ETS	NN	ARFIMA	TBATS	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	0.71*	0.71*
2	0.43	0.50	0.65*	0.61*	0.58	0.53	0.42	0.42	0.40	0.78*	0.78*
3	0.52	0.52	0.61*	0.58	0.57	0.45	0.50	0.54	0.54	0.72*	0.71*
4	0.49	0.49	0.62*	0.61*	0.52	0.46	0.47	0.55	0.55	0.67*	0.67*
5	0.49	0.51	0.66*	0.55	0.53	0.50	0.47	0.56	0.55	0.74*	0.74*
6	0.50	0.51	0.61*	0.58	0.54	0.51	0.48	0.56	0.53	0.82*	0.82*
7	0.41	0.47	0.56	0.56	0.45	0.43	0.38	0.54	0.50	0.82*	0.82*
8	0.43	0.42	0.57	0.57	0.47	0.45	0.42	0.53	0.47	0.82*	0.78*
9	0.46	0.50	0.60	0.64*	0.50	0.45	0.43	0.56	0.53	0.84*	0.83*
10	0.38	0.42	0.55	0.49	0.39	0.41	0.35	0.54	0.51	0.85*	0.86*
Avg.	0.45	0.49	0.60	0.58	0.51	0.48	0.43	0.52	0.50	0.78	0.77

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

Finally, in Figure 3 we present a graphical illustration of the the best out-of-sample forecasts for GT in the very short run ($h = 1$) and the very long run ($h = 10$). It is evident that as the horizon increases the forecasting task appears more difficult even for the MSSA models which are seen performing significantly better than the other models considered here. 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 reflects the comparatively poor DC predictions reported in Table 4. Figure 4 in the Appendix shows the plot of the actual and the forecasts for horizons 1 to 10 attained via the models with the best DC predictions in both univariate and multivariate cases. The approach proposed in this study, whereby we extract the trend in GT with SSA prior to combining the extracted GT trend with CO₂ data in a multivariate framework enables the MSSA models to overcome the issues pertaining to modelling the volatility in GT and provide forecasts which are comparatively more accurate and reliable for decision making.

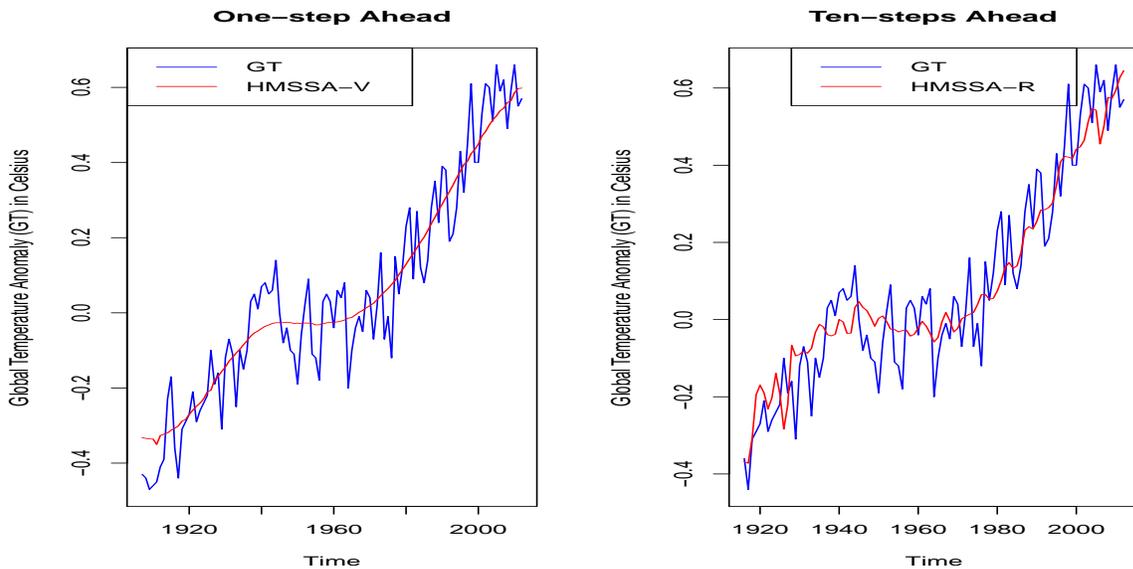


Figure 3: Out-of-sample MSSA forecasts for GT.

5 Conclusions

The two popular and passionate debates in the space of climate change are (i) Is it possible to predict global temperature anomaly (GT) reliably?; and (ii) Is there definitive causal evidence of (anthropometric) CO₂ 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 non-parametric out-of-sample forecasting techniques using only GT data (univariate), as well as using both GT and CO₂ 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 CO₂ 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.

References

- Altavilla, C. and De Grauwe, P. (2010). ‘Forecasting and combining competing models of exchange rate determination’, *Applied Economics*, Vol. 42, pp. 3455–3480.
- Bai J., and Perron P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18, 1–22.
- Boden, T.A., G. Marland, and R. J. Andres. (1995). Estimates of global, regional, and national annual CO₂ 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.
- Campbell, J.Y. (2008). Viewpoint: Estimating the Equity Premium. *Canadian Journal of Economics*, 41, 1-21.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106, 1513–1527.
- Doan, T. A., Litterman, R. B., and Sims, C. A. (1984). Forecasting and Conditional Projections Using Realistic Prior Distributions. *Econometric Reviews*, 3, 1–100.
- Hansen, P. R., Peter, R., Lunde, A., Nason, J. M. and James, M. (2011). The Model Confidence Set. *Econometrica*, 79, 453–497.
- ansen, J. E., M. Sato, R. Ruedy, A. Lacis, J. Glascoe. (1998). Global climate data and models: A reconciliation, *Science*, 281, 930932.
- ansen, J., Ruedy, R., Sato, M., Lo, K. (2010) Global surface temperature change. *Rev Geophys* 48, RG4004, 1-29.
- Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281-291.
- Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5(2), pp. 239-257.
- Hassani, H., Heravi, H., and Zhigljavsky, A. (2009). Forecasting European industrial production with Singular Spectrum Analysis. *International Journal of Forecasting*, 25, 103–118.
- Hassani, H., Mahmoudvand, R. and Yarmohammadi, M. (2010a). Filtering and denoising in the linear regression model. *Fluctuation and Noise Letters*, 9, 343–358.
- Hassani, H., Dionisio, A. and Ghodsi, M. (2010b). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlinear Analysis: Real World Applications*, 11, 492–502.
- Hassani, H., Soofi, A. S., and Zhigljavsky, A. A. (2010c). Predicting daily exchange rate with singular spectrum analysis. *Nonlinear Analysis: Real World Applications*, 11, 2023–2034.
- Hassani, H. and Mahmoudvand, R. (2013). Multivariate Singular Spectrum Analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics*, 1, 55–83.

- Haslett, J., and Raftery, A. E. (1989). Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with discussion), *Applied Statistics*, 38, 1–50.
- Hassani, H., Heravi, H., and Zhigljavsky, A. (2013a). Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, 32, 395–408.
- Hassani, H., Soofi, A. S., and Zhigljavsky, A. (2013b). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society A*, 176, 743–760.
- Hassani, H., and Silva, E. S. (2015). A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts. *Econometrics*, 3, 590–609.
- Hassani, H., Webster, A., Silva, E. S., and Heravi, S. (2015a). Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis. *Tourism Management*, 46, 322–335.
- Hassani, H., Silva, E. S., Gupta, R., and Segnon, M. K. (2015b). Forecasting the price of gold. *Applied Economics*, 47, 4141–4152.
- Hyndman, R. J., and Khandakar, Y. (2008). Automatic time series forecasting: The forecast Package for R. *Journal of Statistical Software*, 27, 1–22.
- Hyndman, R. J., and Athanasopoulos G. (2013). *Forecasting: principles and practice*. OTexts: Australia. OTexts.com/fpp.
- Hyndman, R. J., Athanasopoulos, G., Razbash, S., Schmidt, D., Zhou, Z., Khan, Y., & Bergmeir, C. (2013). Package forecast: Forecasting functions for time series and linear models. Available via: <http://cran.r-project.org/web/packages/forecast/forecast.pdf>. [Accessed: 08 August 2015]
- Khandekar, M.L., Murty, T.S., Chittibabu, P. (2005) The global warming debate: a review of the state of science. *Pure Appl Geophys* 162,15571586.
- Litterman, R. B. (1981). A Bayesian Procedure for Forecasting with Vector Autoregressions. *Working Paper*, Federal Reserve Bank of Minneapolis.
- Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business and Economic Statistics*, 4, 25–38.
- Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1998) *Forecasting: Methods and Applications*. 3rd edn, Wiley: New York.
- McMillan, D.G., Wohar, M. E. (2013). The relationship between temperature and CO₂ emissions: evidence from a short and very long dataset. *Applied Economics*, 45, 3683-3690.
- Oropeza, V., and Sachchi, M. (2011). Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76, V25–V32.
- Patterson, K., Hassani, H., Heravi, S., and Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics*, 38, 2183–2211.
- Sanei, S., and Hassani, H. (2015). *Singular Spectrum Analysis of Biomedical Signals*, CRC Press: United States.

- Silva, E. S., and Hassani, H. (2015). On the use of Singular Spectrum Analysis for Forecasting U.S. Trade before, during and after the 2008 Recession. *International Economics*, 141, 34–49.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48, 1–48.
- Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*, 58, 113–144.
- Solomon, S., Plattner, G.-K., Knutti, R., Friedlingstein, P. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl Acad. Sci. USA* 106(6), 1704-1709.
- Spencer, D. E. (1993). Developing a Bayesian Vector Autoregression Model. *International Journal of Forecasting*, 9, 407–421.
- Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley.
- Todd, R. M. (1984). Improving Economic Forecasting with Bayesian Vector Autoregression. *Quarterly Review*, Federal Reserve Bank of Minneapolis, Fall, 18–29.
- United Nations. (2014). 2011 Energy Statistics Yearbook. United Nations Department for Economic and Social Information and Policy Analysis, Statistics Division, New York.

Appendix

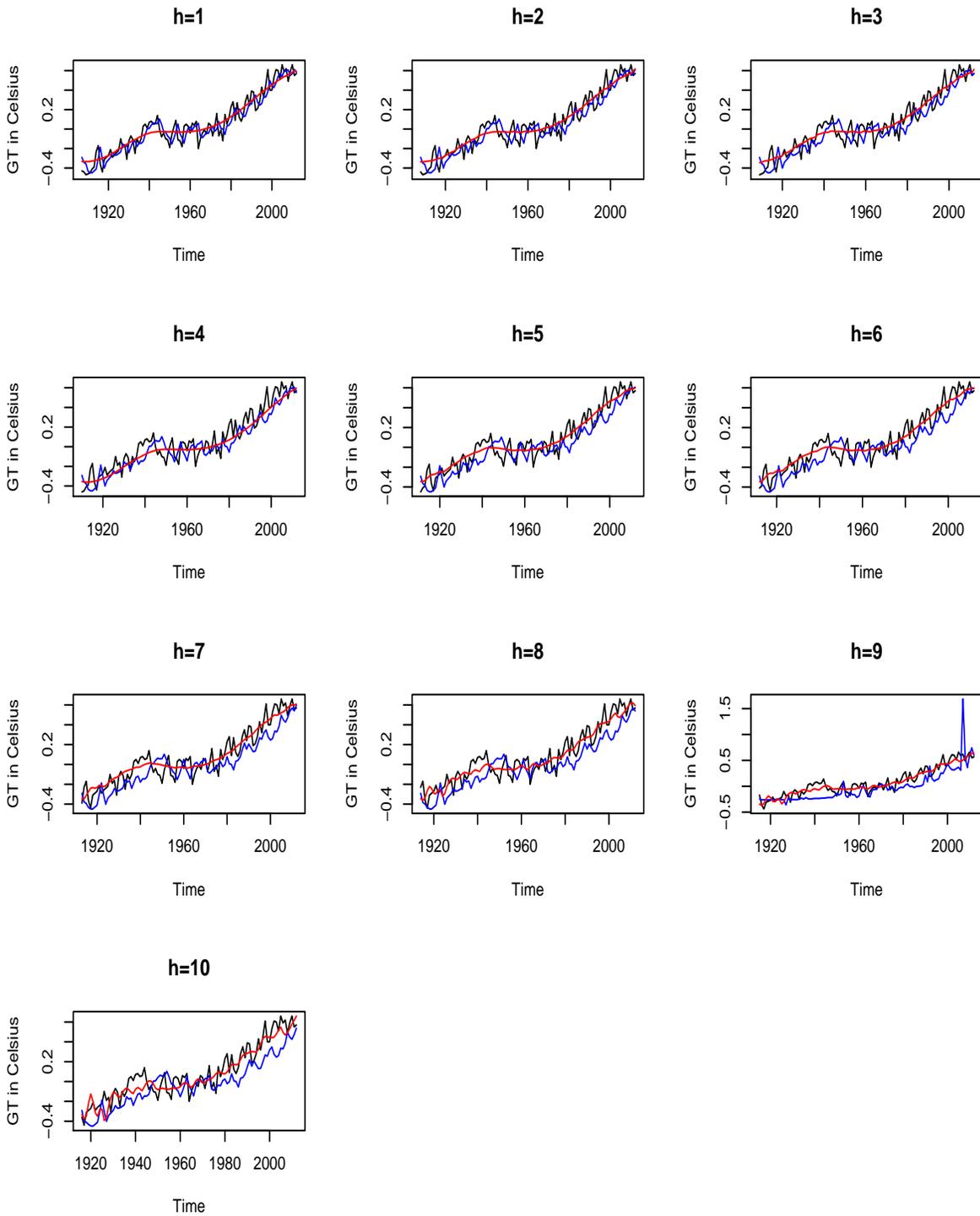


Figure 4: Best DC predictions: GT (Black), univariate model (Blue), MSSA model (Red).