International Journal of Forecasting 34 (2018) 314–335

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Data-based mechanistic modelling and forecasting globally averaged surface temperature

Peter C. Young *

Lancaster Environment Centre, Lancaster University, UK
Integrated Catchment Assessment and Management Centre, Australian National University College of Medicine, Biology & Environment Canberra, ACT, Australia

ARTICLE INFO

Keywords:
Global Temperature Anomaly
Data-Based mechanistic modelling
Differential equation model
Quasi-cyclic variations
Adaptive forecasting

ABSTRACT

The main objective of this paper it to model the dynamic relationship between global averaged measures of *Total Radiative Forcing* (RTF) and surface temperature, measured by the *Global Temperature Anomaly* (GTA), and then use this model to forecast the GTA. The analysis utilizes the *Data-Based Mechanistic* (DBM) approach to the modelling and forecasting where, in this application, the unobserved component model includes a novel hybrid Box-Jenkins stochastic model in which the relationship between RTF and GTA is based on a continuous time transfer function (differential equation) model. This model then provides the basis for short term, inter-annual to decadal, forecasting of the GTA, using a transfer function form of the Kalman Filter, which produces a good prediction of the 'pause' or 'levelling' in the temperature rise over the period 2000 to 2011. This derives in part from the effects of a quasi-periodic component that is modelled and forecast by a *Dynamic Harmonic Regression* (DHR) relationship and is shown to be correlated with the *Atlantic Multidecadal Oscillation* (AMO) index.

1. Introduction

Most climatic modelling research conforms to the 'scientific method', following what Popper (1959) has termed a 'hypothetico-deductive' approach. But Popper's philosophical stance is based mainly on his consideration of laboratory-based science: one has only to search for the word 'experiment' in his famous book to realise how important experimentation, and particularly planned experimentation, is in his view of the scientific method. Such planned experimentation can help, of course, to minimize noise and remove any ambiguity on measured data, so clarifying understanding of the potentially complex mechanisms that underlie this observed behaviour. Unfortunately, it is not possible to conduct well planned experiments on large, naturally occurring systems, such as the global climate, and so modelling has to be based on data collected during the 'normal operation' of the system.

This hypothetico-deductive approach to scientific research can be contrasted with the *inductive* method, which has a rich history in science (Young, 2012). Induction is the opposite of deduction: it starts by taking as many observations of the natural system as possible, with the aim of inferring from these data alone how the system works, without the introduction of any hypotheses that may be prejudicial and so distort the modelling process. In the real world, of course, this differentiation between these two approaches to scientific research is too simplistic: inductive and hypothetico–deductive modelling are synergistic activities, the relative contributions of which will depend upon the system being modelled and the information of *all* types, not only time-series data, that are available.

* Correspondence to: Lancaster Environment Centre, Lancaster University, UK.
E-mail address: p.young@lancaster.ac.uk.

https://doi.org/10.1016/j.ijforecast.2017.10.002

The present paper applies the *Data-Based Mechanistic* (DBM) method of modelling, forecasting and control (see chapter 12 in Young, 2011 and the prior references therein) to globally averaged surface temperature data, as represented here and most other publications on this topic by the Global Temperature Anomaly (GTA) series (see later, Section 5). This approach is primarily based on induction but it attempts to produce models that can be interpreted in an understandable mechanistic manner; i.e. in manner which relates to the mechanisms that are normally used to describe the nature of the system being analysed, as well as the concepts and assumptions that underlie the way these mechanisms are described in mathematical terms. This DBM model normally relates input signals of some kind to the outputs that are most likely to be influenced by these inputs.

Climate modellers tend to use a standard hypothetico-deductive approach but the nature of the system does not allow them to exploit planned experimentation. This often results in complex, high order models based on deterministic, ordinary or partial differential equations. As a result of the high order and the associated large number of model parameters, standard optimization based on simple cost functions is not possible and resort needs to be made to other procedures. An interesting discussion on the difficulties of optimizing these models and some ways of handling these difficulties is given in Neelin, Bracco, Luo, McWilliams, and Meyerson (2010). They point out that the underlying nature of the system has an enormous impact on which of the strategies that are available for the evaluation of sensitivity and optimization are most suitable. The model parameters and the associated range of possible values are normally chosen on a scientific basis based on their physical interpretation and function within the model. Model evaluation normally involves a critical examination of the simulated behaviour in response to internal or external inputs. Sometimes this is accompanied by the application of parametric sensitivity analysis, which indicates which parameters are most affecting the simulated response and so are deserving of particular attention and optimization.

The main inputs that are thought to most affect the GTA are the various globally averaged 'radiative forcing' signals, of which the best known is the greenhouse gas, atmospheric carbon dioxide. It is appropriate, therefore, that DBM models of global surface temperature behaviour, based on the available globally averaged radiative forcing and surface temperature data, should be formulated in terms of similar differential equation models to those used in simpler global climate 'emulation' models (see later Section 3). In contrast to most climate models, however, the DBM models considered in the present paper are inherently stochastic and of a normally low dynamic order that reflects the few 'modes' of dynamic behaviour that, as we shall see, dominate both the simulated output of the large climatic models and the real globally averaged data.

This DBM approach to modelling and forecasting can be contrasted also with the methodology used most often by the forecasting profession, where the standard inductive approach is usually based on discrete-time, 'black box' statistical models with various degrees of complexity. Continuous-time models are used in some applications, such as stock prices, but these are normally in the form of purely stochastic differential diffusion equations based on the application of complex Itô calculus, rather than, as here, ordinary differential equations represented by simple transfer functions in the differential operator.

It is not clear why there has been a reticence to use such continuous-time transfer function models for forecasting but it may be that the methods of statistical identification and estimation for such models are not well known because they were developed originally within the control and systems community (see chapter 8 in Young, 2011 and the prior references therein, going back to the first optimal solution proposed in 1980 Young and Jakeman, 1979). However, there are number of reasons why it can be advantageous to use continuous-time models (Garnier & Young, 2014), in addition to the obvious advantage that such models are better suited for the analysis of rapidly sampled time series, where discrete-time models can encounter estimation difficulties if the eigenvalues of the model are too close to the unit circle in the complex domain. It is felt, therefore, that the forecasting community might consider continuous-time models when the system being modelled is affected dynamically by inputs entering through transfer function models and/or the system description needs to be interpreted in a mechanistic manner. In simple terms, this means that the equivalent and familiar discrete-time equation containing lagged outputs (dependent or target variables) and inputs (independent, explanatory or exogenous variables) is replaced by a related differential equation, with the discrete-time lag operators replaced by suitable differential operators (see Section 4.1).

Bearing in mind the above remarks, the main objective of the present paper is to synthesize a continuous time, differential equation model, considered in a convenient transfer function form, for the changes in globally averaged surface temperature arising from the measured changes in total radiative forcing. In particular, this model needs to be well suited for relatively short term, inter-annual to decadal, prediction of the temperature where, as pointed out by an anonymous reviewer, techniques employing explicit time-stepped numerical simulation models currently run into difficulties caused by 'initialization and cold-start' problems (see e.g. http://blogs.nature.com/climatefeedback/2007/06/predictions_of_climate.html; and Meehl et al., 2007). As such, this differential equation model needs to represent efficiently the dominant modal behaviour of the global climate system; and it should be a stochastic model, obtained using optimal methods of statistical inference, so that it can quantify the uncertainty in the predictions over the forecasting interval.

Finally, note that this kind of stochastic-dynamic modelling is as important as the forecasting in the present study. Indeed, if forecasting alone was the objective, then a simpler approach based, for example, on univariate time series modelling and forecasting, can provide somewhat inferior but still acceptable short-term forecasts of the GTA, as shown in Section 6.3.2 of the paper. But such a univariate approach is purely black box in form and so does not conform with the aims and philosophy of DBM approach to modelling and forecasting.