Data-Based Mechanistic Modelling and its Application to the Global Climate System

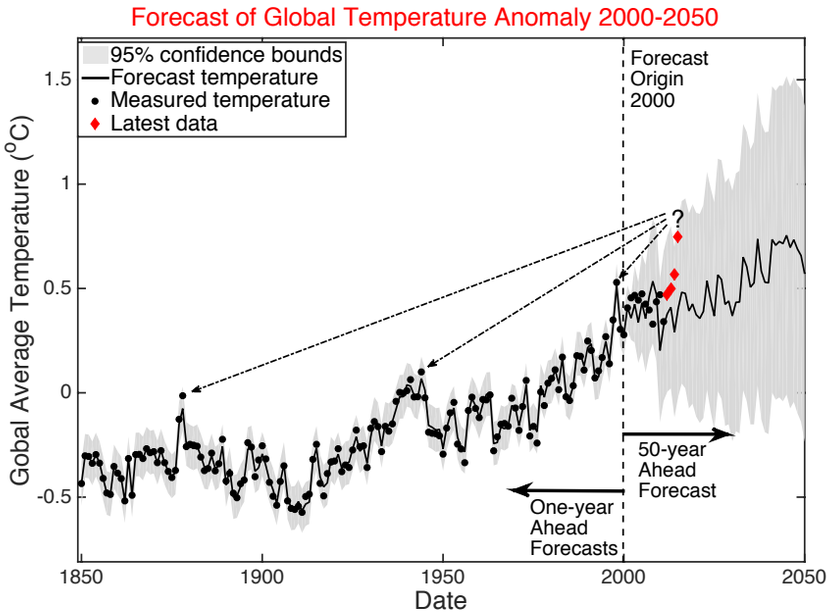
Stained Glass Window, Young Family House, Lancaster, England

Data-Based Mechanistic Modelling and its Application to the Global Climate System

Peter C. Young

Lancaster University, U.K. ; Australian National University, Canberra

SAMO 2016: La Réunion



p.young@lancaster.ac.uk;

Part I

The Background Philosophy and Methodology of Data-Based Mechanistic Modelling

Introduction

- The standard *hypothetico-deductive* approach to scientific research can be contrasted with the *inductive* method, which has a rich history in science and was, indeed, the normal approach to scientific research used during the era of the 'natural philosopher', the term used to describe the scientist prior to the twentieth Century.

- Induction is the opposite of the deduction: it starts by taking as many observations of the natural system as possible, with the aim of inferring from these data how the system works.

- Of course, this differentiation between different approaches to scientific research is too simplistic in the real world. *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 to the scientist and modeller.

- Unfortunately, the *conventional hypothetico-deductive* approach to modelling natural systems often yields very large, nominally complex and normally deterministic simulation models that are seriously over-parameterized in statistical identification terms. As a result, they are difficult, if not impossible, to estimate unambiguously from the normally noisy time-series data.

- In order to reconcile this large simulation model with a low order model that is identifiable from the noisy time series data, therefore, it is necessary to consider some form of *model simplification and/or sensitivity analysis*.

- An appropriate approach to model simplification that serves these requirements well is *model emulation* (sometimes termed *meta-modelling*), where the large model is emulated by a much reduced order model (Young and Ratto, 2009, 2011, and prior references therein).

- Moreover, using this *Data-Based Mechanistic* (DBM) modelling approach to emulation, a *complete statistical uncertainty estimation* can be derived for the surrogate dynamic behaviour, as well as a *comprehensive sensitivity analysis*.

Data-Based Mechanistic (DBM) Modelling

The philosophical ideas behind DBM modelling are very simple:

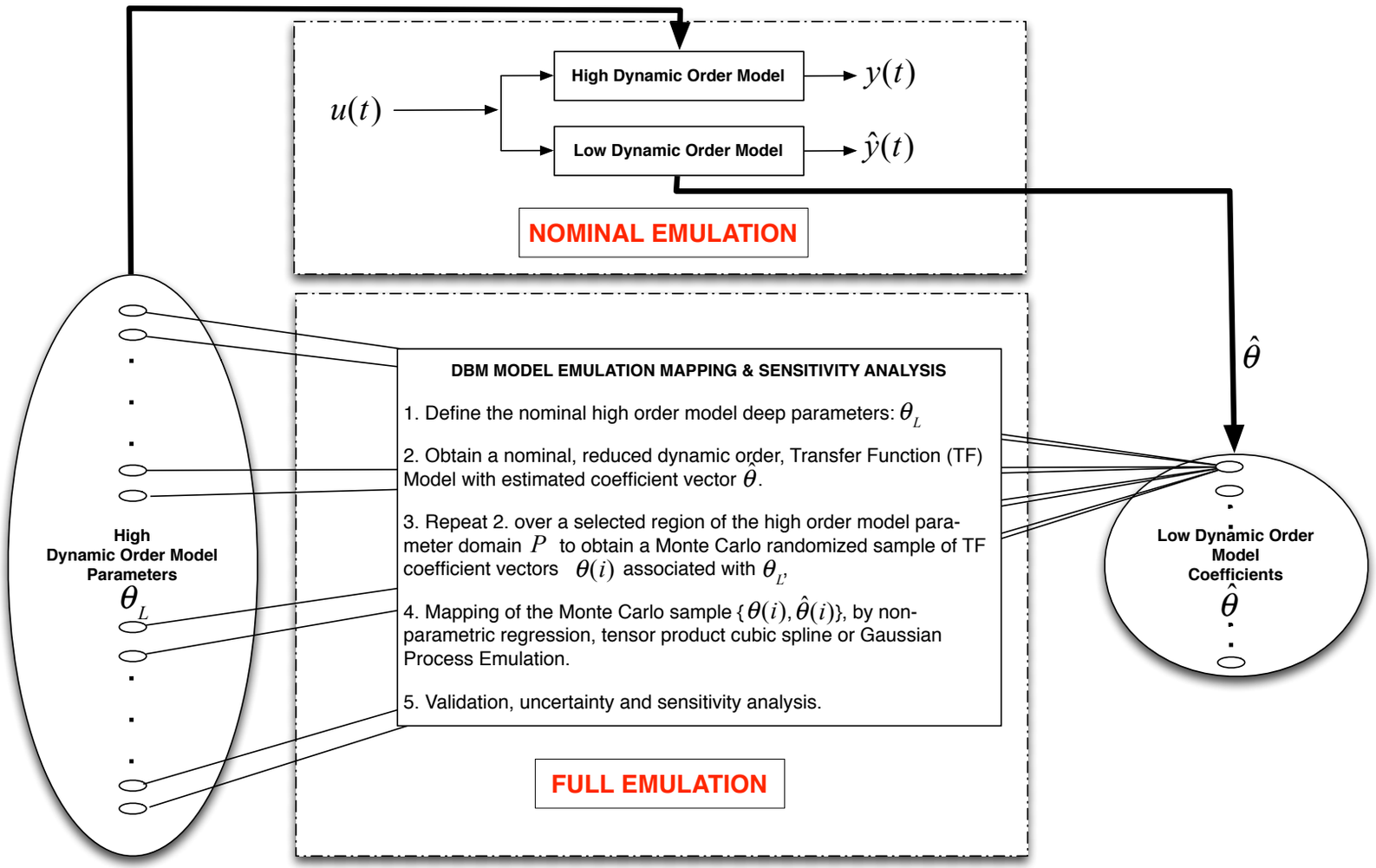
- Data-Based Modelling: where time series data are available, the stochastic-dynamic model structure should be inferred *inductively* from these data to avoid undue prejudice about the nature of the model, as happens quite often in *hypothetico-deductive* conceptual model building, where the model is normally based on a prior hypothesis about the dynamic mechanisms within the system, which is then tested against the data.

- Mechanistic Interpretation: the identified model structure and its parameters *are only deemed credible* if they can be *interpreted in physically meaningful terms*, so that the model can be compared with conceptual models that are more widely understood. *Clearly, continuous-time models are a definite advantage* in this regard when considering scientific and engineering applications, *where continuous-time differential equation models are common*.

DBM Emulation Modelling
The Simplification of High Order Models

There are two phases to DBM emulation modelling:

- a straightforward and quick **nominal emulation** phase which produces an emulation model that behaves like the large simulation model but whose parameters are not directly related to those of the large model. This can also yield *sensitivity information of a different kind*: namely, the relative importance of the dynamic modes of the simulation model, rather than the relative importance of its parameters. These *dominant modes of dynamic behaviour* dictate the nature of the simulated response of the model and can normally be interpreted directly in physically meaningful terms.
- a more complicated **full emulation** phase, where the parameters of the the nominal model are mapped to those of large simulation model, *allowing for the associated sensitivity analysis, if this is required*.
- In this talk, I will consider a global climate modelling and forecasting example, where I use only *nominal emulation*, which is based on a DBM approach to **Dominant Mode Analysis** and associated *Dynamic Model Order Reduction*.


$u(t)$
High Dynamic Order Model
$y(t)$
Low Dynamic Order Model
$\hat{y}(t)$
NOMINAL EMULATION
$\hat{\theta}$
DBM MODEL EMULATION MAPPING & SENSITIVITY ANALYSIS
1. Define the nominal high order model deep parameters: θ_L
2. Obtain a nominal, reduced dynamic order, Transfer Function (TF) Model with estimated coefficient vector $\hat{\theta}$.
3. Repeat 2. over a selected region of the high order model parameter domain P to obtain a Monte Carlo randomized sample of TF coefficient vectors $\theta(i)$ associated with θ_L,
4. Mapping of the Monte Carlo sample $\{\theta(i), \hat{\theta}(i)\}$, by non-parametric regression, tensor product cubic spline or Gaussian Process Emulation.
5. Validation, uncertainty and sensitivity analysis.
FULL EMULATION
High Dynamic Order Model Parameters
θ_L
Low Dynamic Order Model Coefficients
$\hat{\theta}$

Dynamic Model Order Reduction

- It has been known for many years that the response of a high order linear model is dominated by a relatively few 'dominant' modes of dynamic behaviour, with many of the eigenvalues that dictate the dynamic behaviour having little effect on the output response (e.g. Davison, 1966; Liaw, 1986).

- Liaw introduces a computational technique termed *Dispersion Analysis* that is based on the partial fraction expansion of a *Transfer Function* (TF) model and might be termed *Modal Sensitivity Analysis*.

- *Dominant Mode Analysis* does not use dispersion analysis at at all; rather it *exploits statistical identification tools*, applied to simulated data from the high order model, to obtain *a reduced order model* that effectively identifies combinations of dynamic modes that dominate the behaviour of the model.

- Before proceeding to the practical example, therefore, it is worth considering the nature of transfer functions and their *Partial Fraction Expansion*.

Continuous-Time TF Models

Most scientific models, are based on ordinary or partial differential equations, such as the following multi-order ordinary differential equation:

$$\frac{d^n x(t)}{dt^n} + a_1 \frac{d^{n-1} x(t)}{dt^{n-1}} + \cdots + a_n x(t) = b_0 \frac{d^m u(t-\tau)}{dt^m} + \cdots + b_m u(t-\tau)$$

τ is a pure time delay to allow for any such delay in the system dynamics; while $a_j,\ j = 1, 2,\ \ldots, n$ and $b_j,\ j = 0, 2,\ \ldots, m$ are normally constant parameters. A model such as this can be transformed straightforwardly into an equivalent state-space form. More importantly in the present context, it can also be written in Single-Input, Single Output (SISO) Transfer Function (TF) form:

$$x(t) = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n} u_i(t - \tau_i) = \frac{B(s)}{A(s)} u(t - \tau)$$

where $s = \frac{d}{dt}$ is the differentiation operator (or Laplace Transform operator) $B(s)/A(s)$ is the transfer function.

Partial Fraction Decomposition of TF Models

Since it is a fraction of two polynomials in s, it can be shown that the transfer function can be expanded as a finite dimensional sum of simple fractions:

$$F(s) = \frac{B(s)}{A(s)} = \frac{r_m}{s - p_m} + ... + \frac{r_2}{s - p_2} + \frac{r_1}{s - p_1} + k(s).$$

This sum is the *Partial Fraction Expansion* of $F(s)$. The values $r_m, ..., r_1$ are the residues, the values $p_m, ..., p_1$ are the poles (both of which can be complex numbers in the case of an oscillatory system), and $k(s)$ is a polynomial in s. For most real problems, $k(s)$ is 0 or a constant which, if present, represents the instantaneous effect of the input on the output.

Conveniently, such a partial fraction expansion can be computed easily using the residue routine in Matlab. A relevant example of this methodology applied within a climate modelling context is discussed on the the next slide.

A Second Order Climate Modelling Example

In this example, *which is considered in the later climate modelling analysis*, the input $u(t)$ is the radiative forcing; the output $x(t)$ is atmospheric temperature,

$$x(t) = \frac{B(s)}{A(s)}u(t) = F(s)u(t)$$

Here, the transfer function $F(s)$ (actually an *emulation* model: *see later*) is identified as:

$$F(s) = \frac{0.153s^2 + 0.0159s + 0.0000524}{s^2 + 0.05106s + 0.000126}u(t)$$

and the partial fraction expansion obtained by calculation or using residue is:

$$F(s) = \frac{0.00781}{s + 0.0485} + \frac{0.000265}{s + 0.00260} + 0.153$$

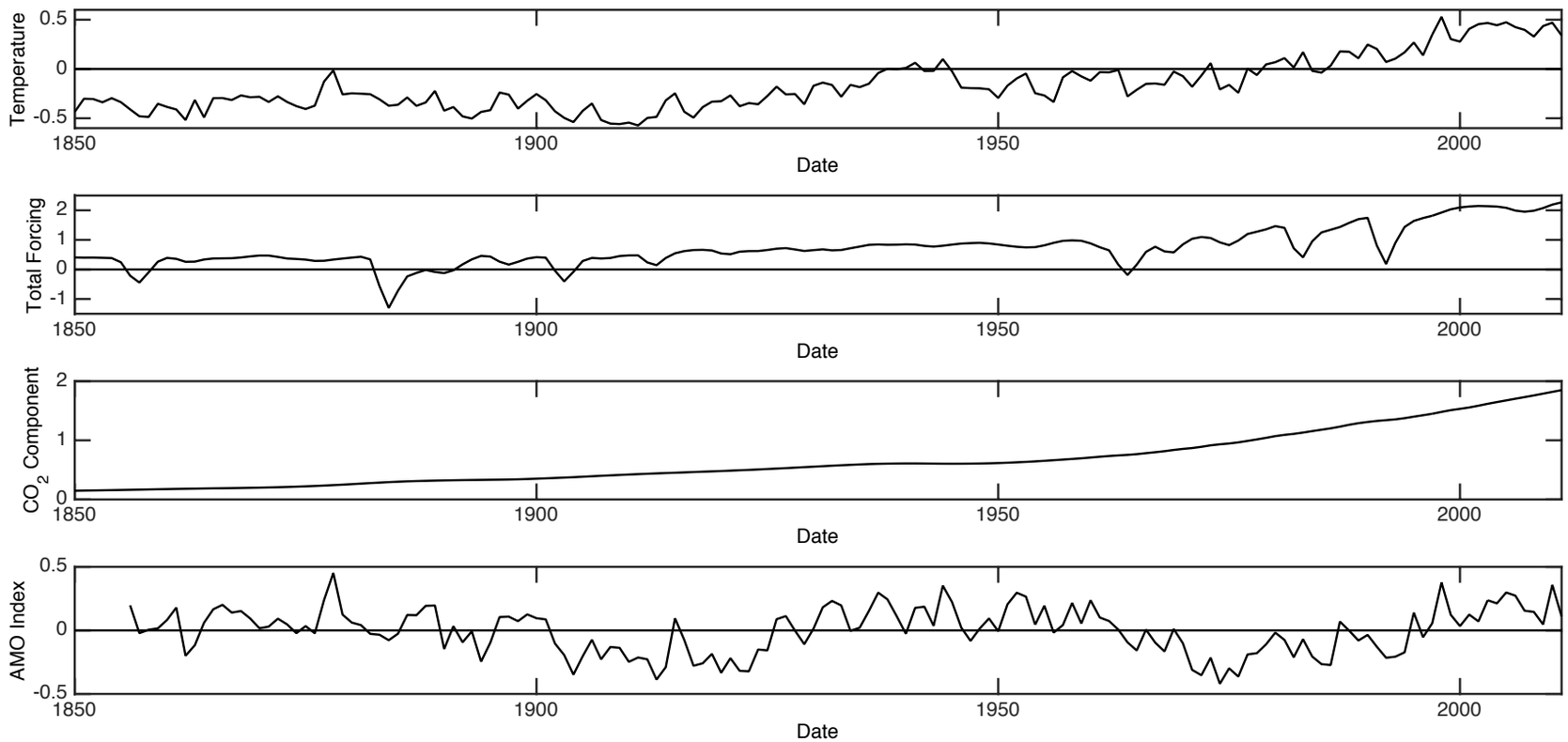
$$= \frac{0.161}{1 + 20.6s} + \frac{0.102}{1 + 385s} + 0.153 = \frac{G_1}{1 + T_1 s} + \frac{G_2}{1 + T_2 s} + G_I$$

As we shall see, this emulation model explains 99.97% of a high order climate model output.

Part II

Data-Based Mechanistic Modelling and Forecasting of the Global Climate System based on Globally Averaged Annual Data

The Globally Averaged Annual Data



Climate data used in the example: the *Total Radiative Forcing* in W/m^2, includes CO_2 and other forcing inputs, such as volcanic activity).

What are Climate Change Models Like?

- Climate models (coupled atmosphere-ocean and carbon cycle models), obtained using a classical, hypothetico-deductive approach, are very large indeed and recent research has been devoted to reduced order emulation models, of which the best known is MAGICC (Meinshausen, 2011).

- But MAGICC is still 80th order and a simple exercise in model reduction, using the optimal *Refined Instrumental Variable* (rivcbj, rivcbjid) routines in the CAPTAIN Toolbox for Matlab^{TM} , shows that it can itself be emulated very well by a *2nd order dominant mode model* estimated in *the same continuous-time, differential equation form as MAGICC and other climate models*.

- Even so, as we see on a later slide, both MAGICC and its low-order emulator do not explain the *measured data* as well as the 1st order DBM model! This simple DBM model is identified from the measured data and estimated in stochastic terms using the same rivcbj and rivcbjid routines.

The CAPTAIN Toolbox for Matlab and My Recent Book

The estimation algorithms used in DBM modelling are fairly complex but fortunately, if you have access to the Matlab software environment, there is no need to program them if you do not wish to. They are available as routines in the CAPTAIN Toolbox for Matlab™ . CAPTAIN can be downloaded *free* via the web site `http://captaintoolbox.co.uk/Captain_Toolbox.html`.

CAPTAIN includes many other routines for time series modelling; as well as other routines for signal processing, forecasting and advanced control system design.

The algorithms are all developed and there use is illustrated by many examples in my recent book *Recursive Estimation and Time Series Analysis: an Introduction for the Student and Practitioner*: see

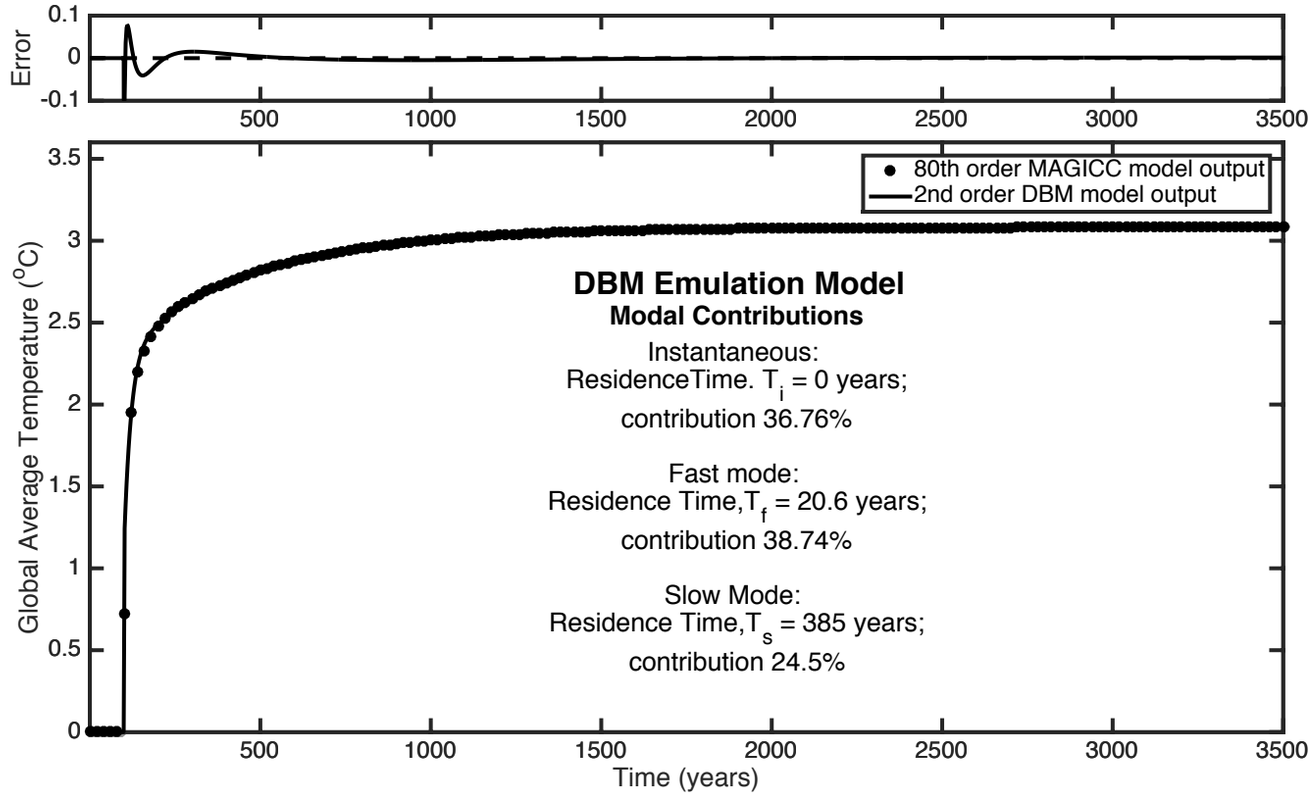
`http://www.springer.com/engineering/control/book/978-3-642-21980-1`

Data-Based Mechanistic Emulation of the MAGICC Model

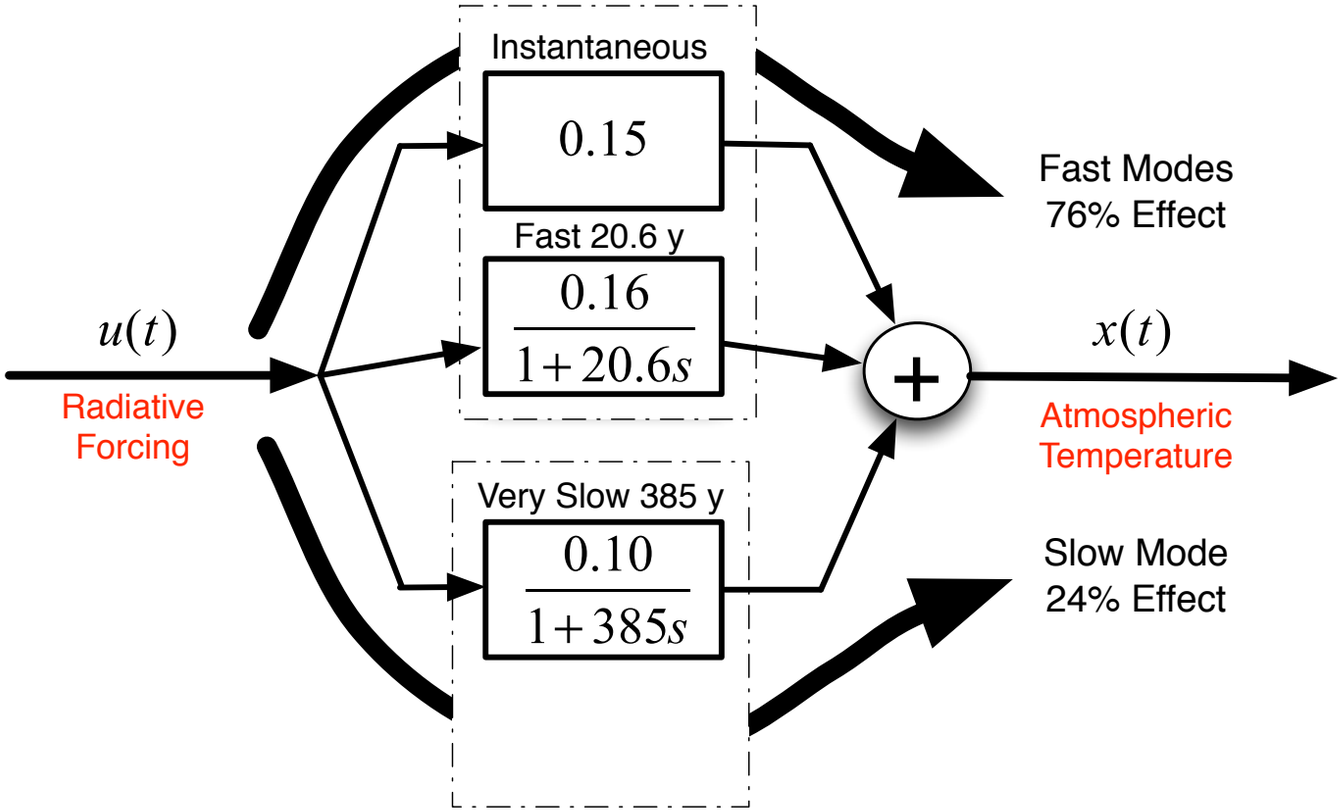
Before proceeding to consider DBM modelling from real data, we consider the much reduced order emulation of the MAGICC model so that we can see later if this low order emulation model can be reconciled with the DBM model.

Second order DBM emulation of the 80th order MAGICC simulation model:

The model explains 99.7% of the simulated model output variance



Decomposition of the 2nd Order DBM emulation model (see earlier example)



Data-Based Mechanistic Modelling from Real Data

Having considered conventional climate models and their emulation, we can now proceed to the main objective of this example: modelling, forecasting and control analysis based on the real, globally averaged data.

DBM Model Estimated from Real Data

$$\text{System Model} \quad x(t) = \frac{0.070s + 0.032}{s + 0.044}u(t) - \frac{0.40s + 0.006}{s + 0.009}u_{IC}(t)$$

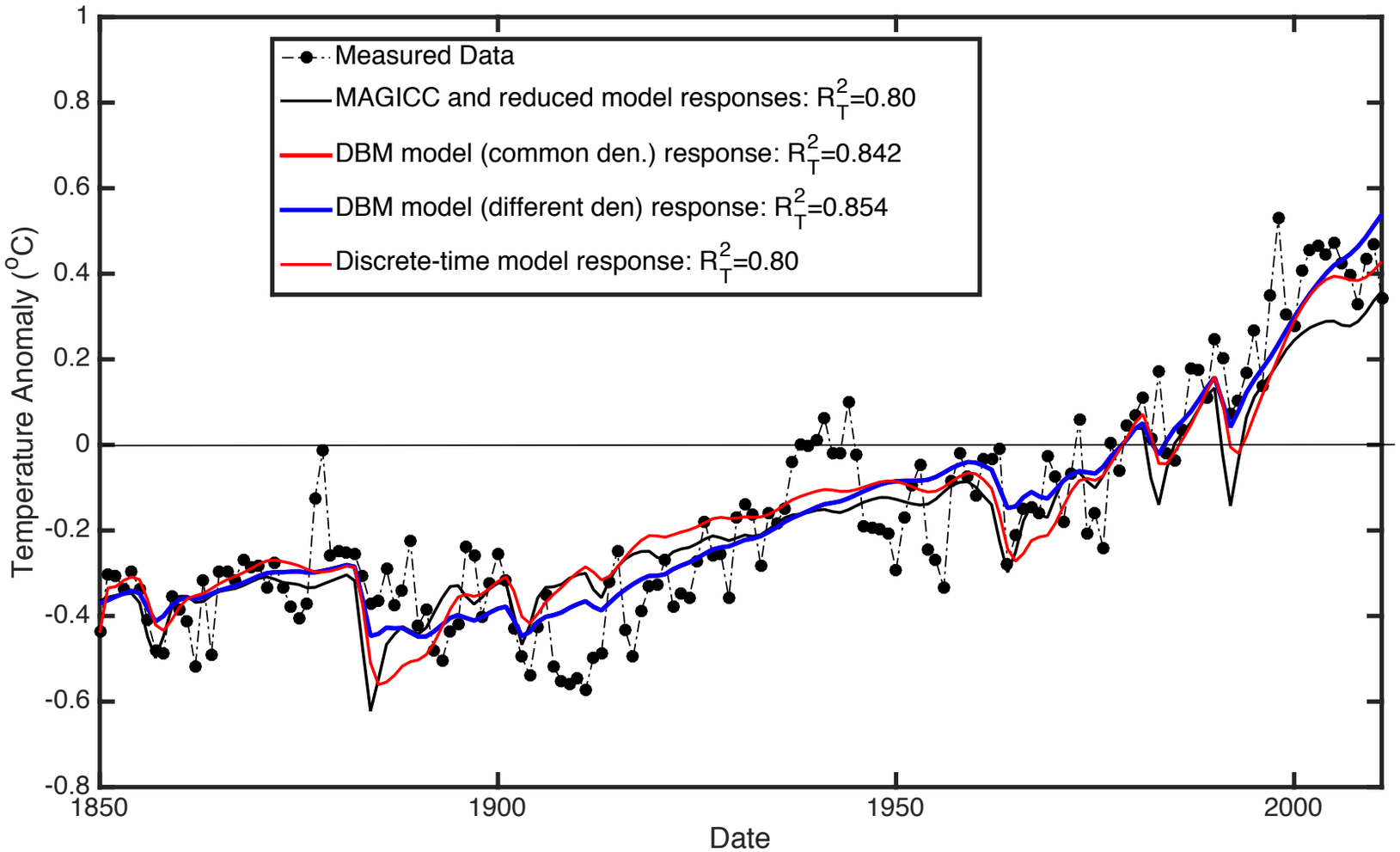
$$\text{AR Noise Model} \quad \xi(k) = \frac{1}{1 - 0.507z^{-1}}e(k); \quad \epsilon(t_k) = \mathcal{N}(0, 0.0075)$$

$$\text{Output Equation} \quad y(k) = x(k) + \xi(k)$$

where z^{-1} is the backward shift operator (L, B depending on background) and the *second TF, with constant input $u_{IC}(t) = 1.0 \; \forall \; t$, follows from Laplace Transform theory and estimates initial conditions*. Written informally as a 'snapshot' in single equation terms, *excluding the initial condition transfer function*:

$$y(k) = \frac{0.070(\pm 0.01)s + 0.032(\pm 0.01)}{s + 0.044(\pm 0.02)}u(k) + \frac{1}{1 - 0.507z^{-1}}e(k)$$

MAGICC and DBM Results with Real Data



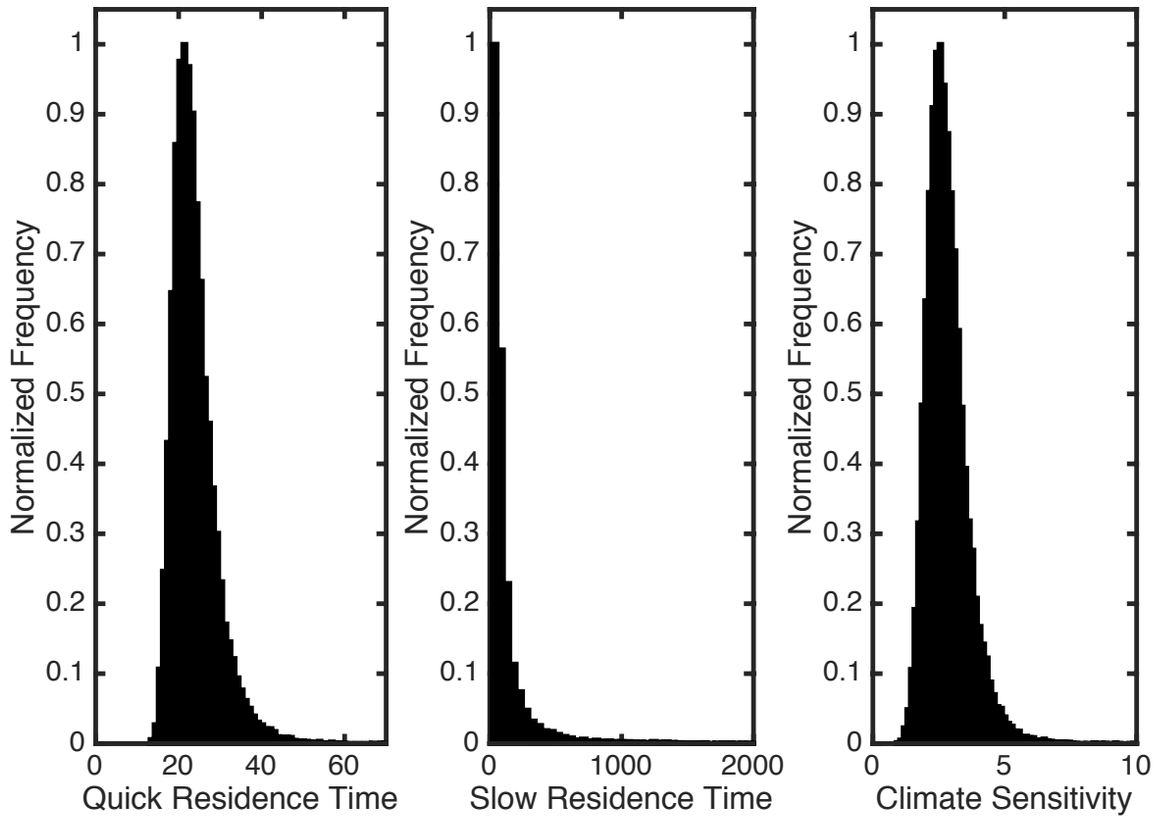
Although nominally 'black-box', this model satisifies DBM modelling requirements by explaining the data rather better than the climate models and being credible in physical terms:

- The *continuous-time model* is a typical differential equation for a 'mixing process' or 'mixing box' *of the kind used in climate modelling*.

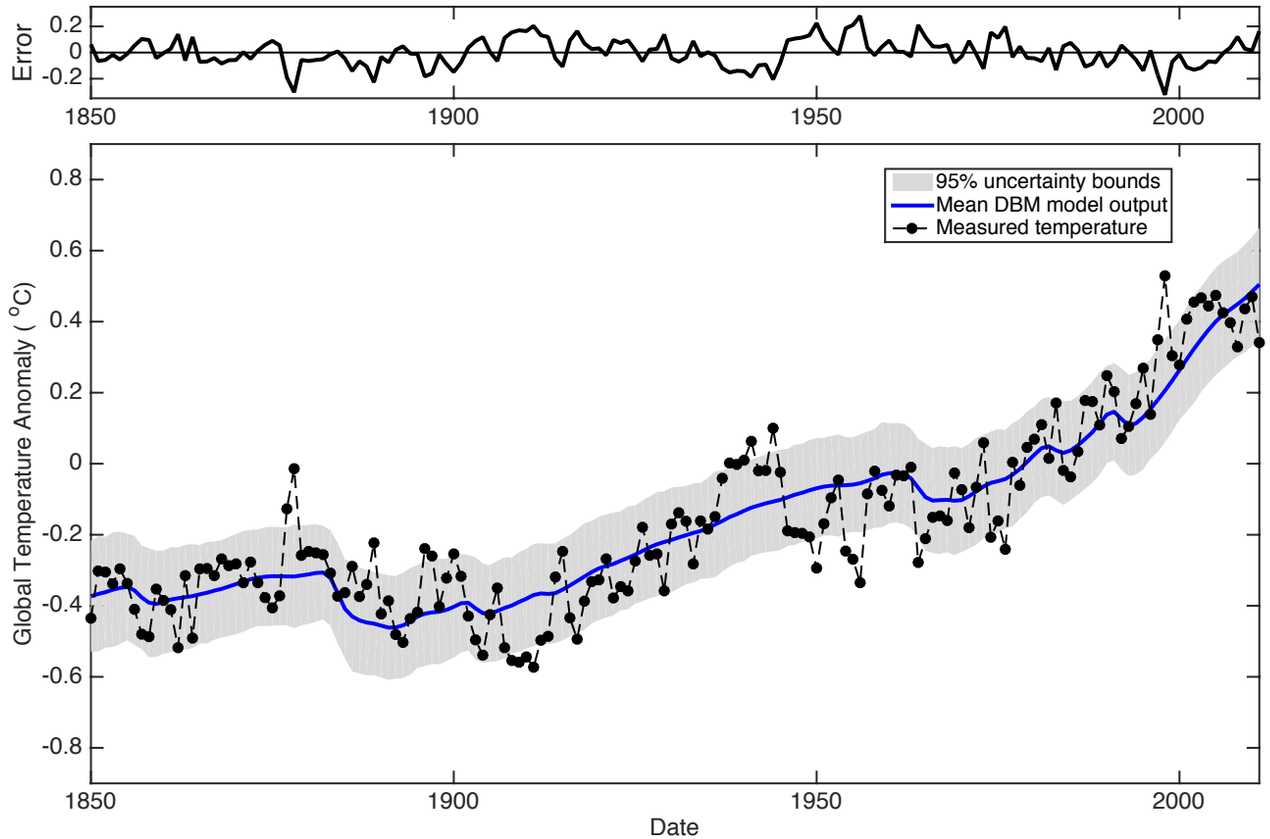
- The 'climate sensitivity' CS $= 2.8^o$C (bounds: 1.78 to 4.25) of this model compares with MAGICC's 4.4^{o}C but *is in the middle of the usually accepted range* of 1.5 to 4.5 ^{o}C (Cubasch et al, 2001) and quite close to that of 2.14^{o}C obtained by Mills (2009).

- The instantaneous response and residence time of the DBM model mixing process: 0.07^oC (bounds: 0.062 to 0.078) and 23.8 years (bounds: 17.3 to 33.6) compare with MAGICC fast mode values (0.068^oC and 20.6 years), which account for 76% of the MAGICC response. *The long period initial condition response suggests that there is some evidence of a poorly defined, long term mode with slow residence time* > 100 *years*.

MCS Uncertainty Analysis: Derived, Physically Meaningful Model Parameters

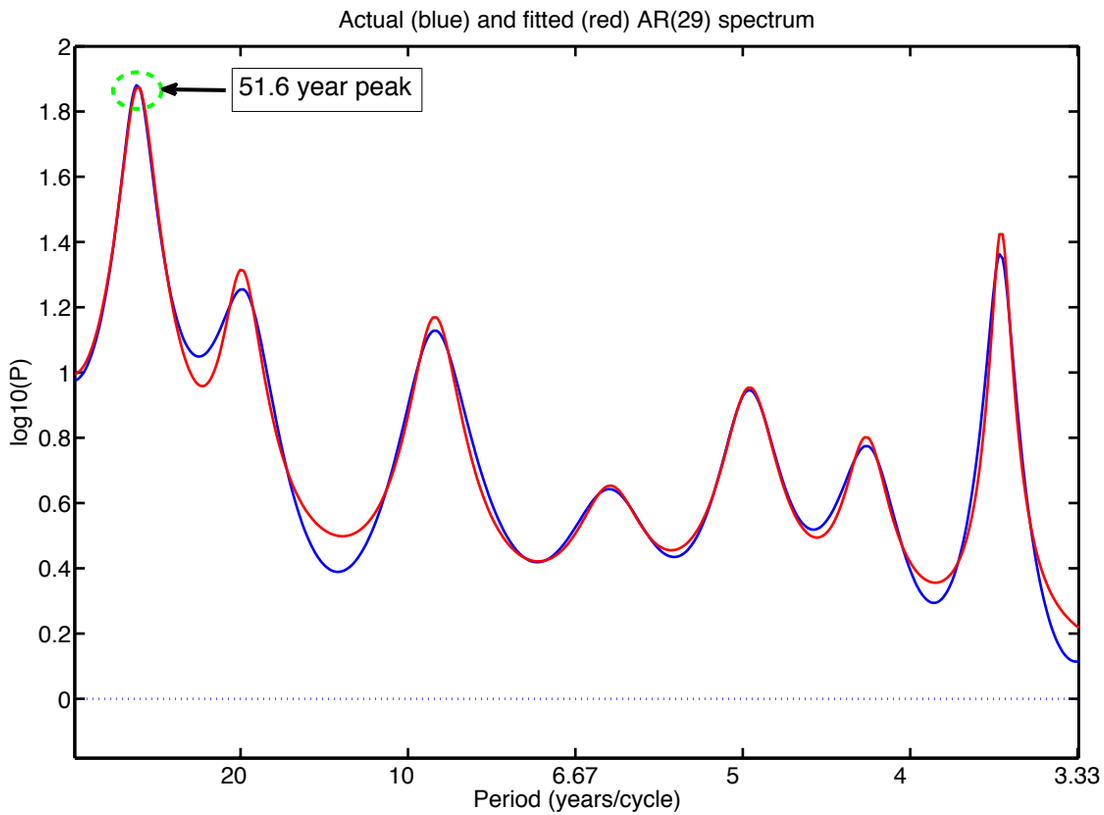


MCS Uncertainty Analysis: Model Response

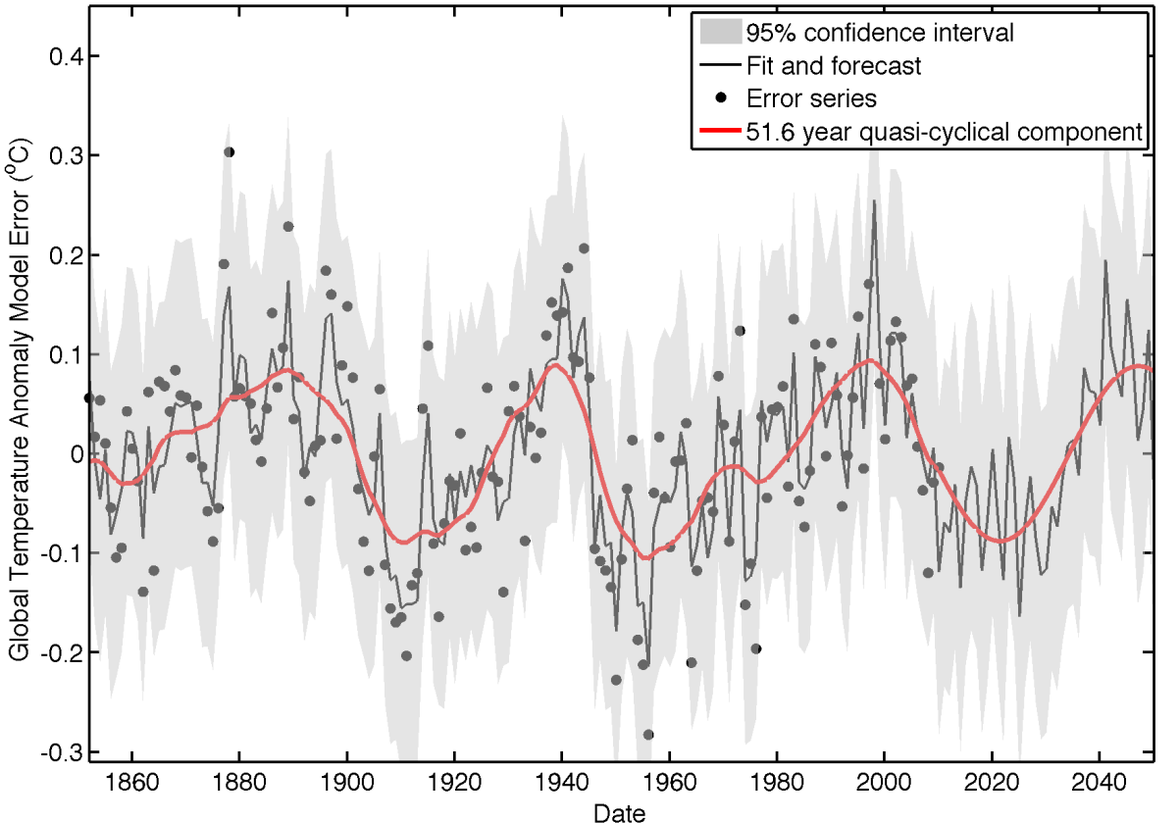


But is this a good enough explanation of the data? Is there a possible cycle in error series?

DHR Analysis suggests 50 year quasi-cycle



DHR optimization based on AR(29) spectrum.

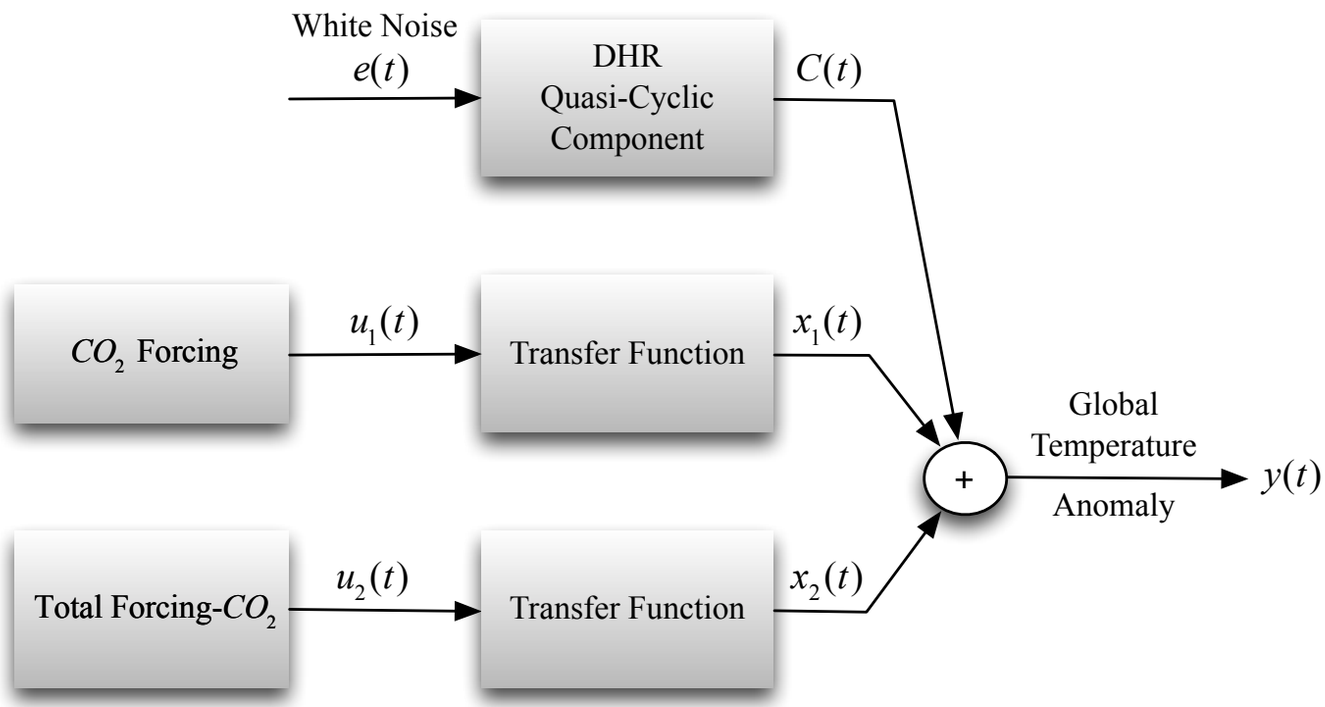


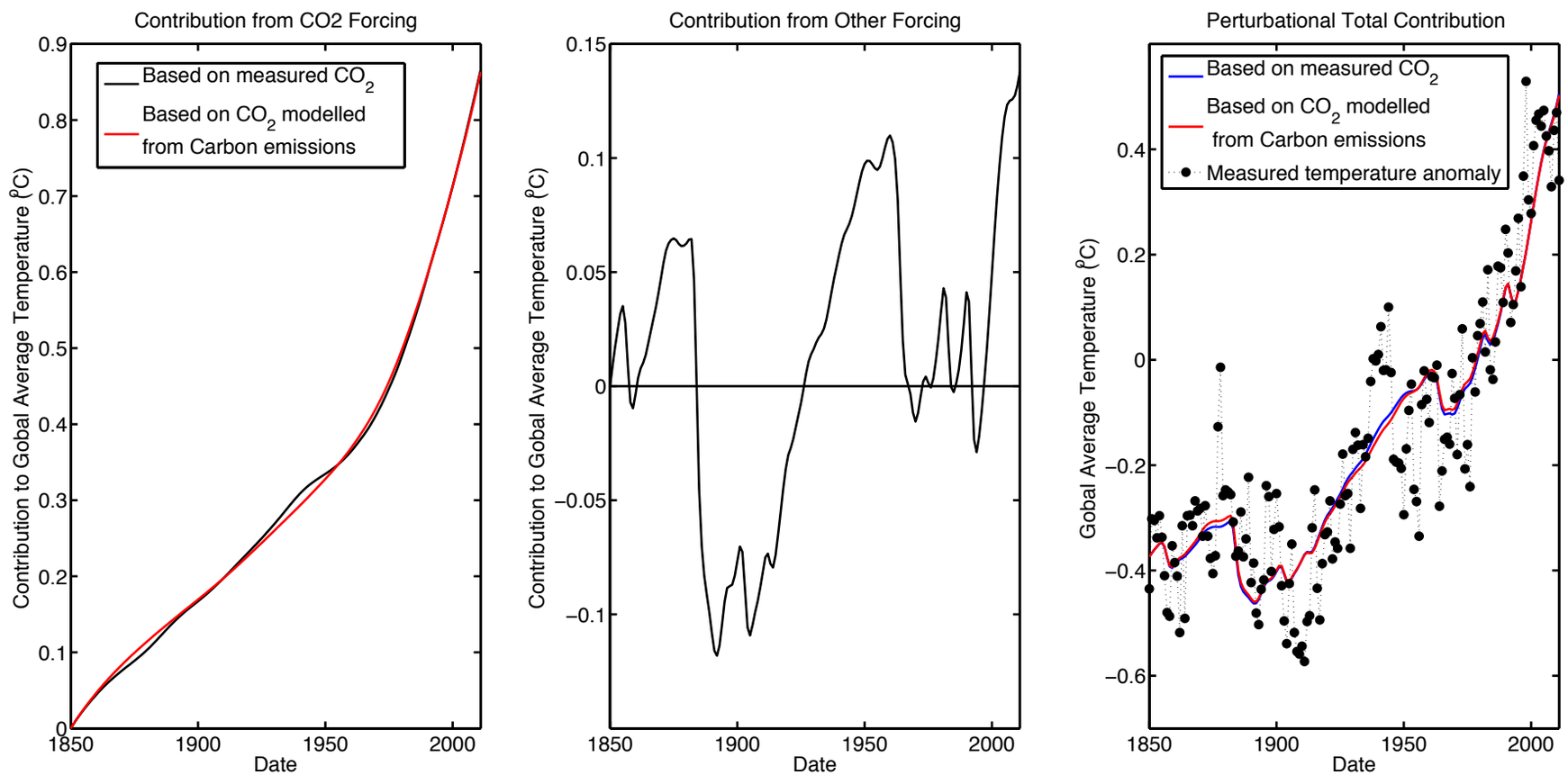
Dynamic Harmonic Regression (DHR) model based on AIC identified AR(29) spectrum reveals a major 51.6 year period component in the model error series (see red curve).

Disaggregated TF block diagram of DBM model.

$$y(t) = \frac{0.070s + 0.032}{s + 0.044} u_1(t) + \frac{0.070s + 0.032}{s + 0.044} u_2(t) + C(t) + e(t)$$

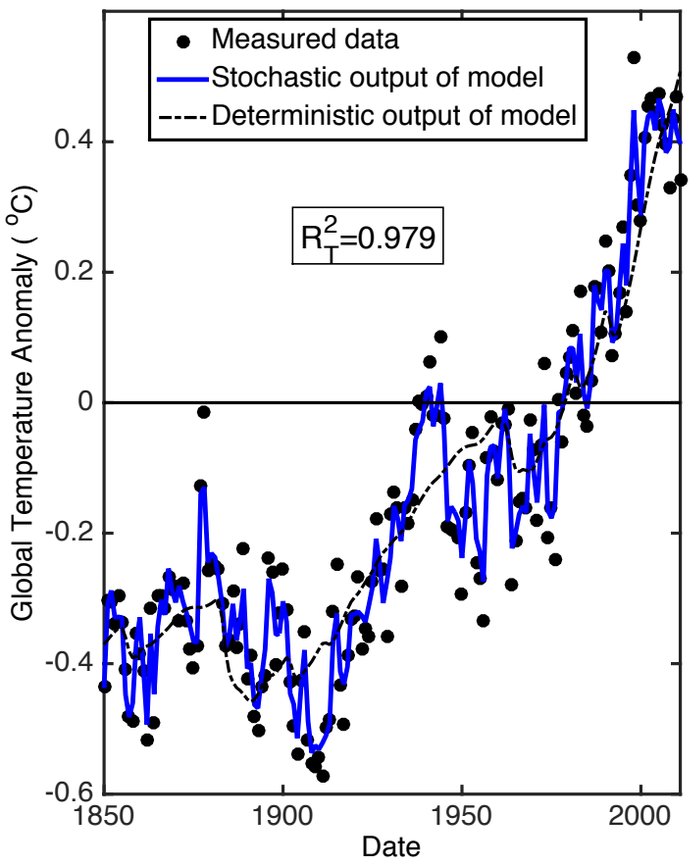
$$u(t) = \text{Total Forcing} \;\; ; \;\; u_1(t) = CO_2 \;\; ; \;\; u_2(t) = u(t) - u_1(t)$$

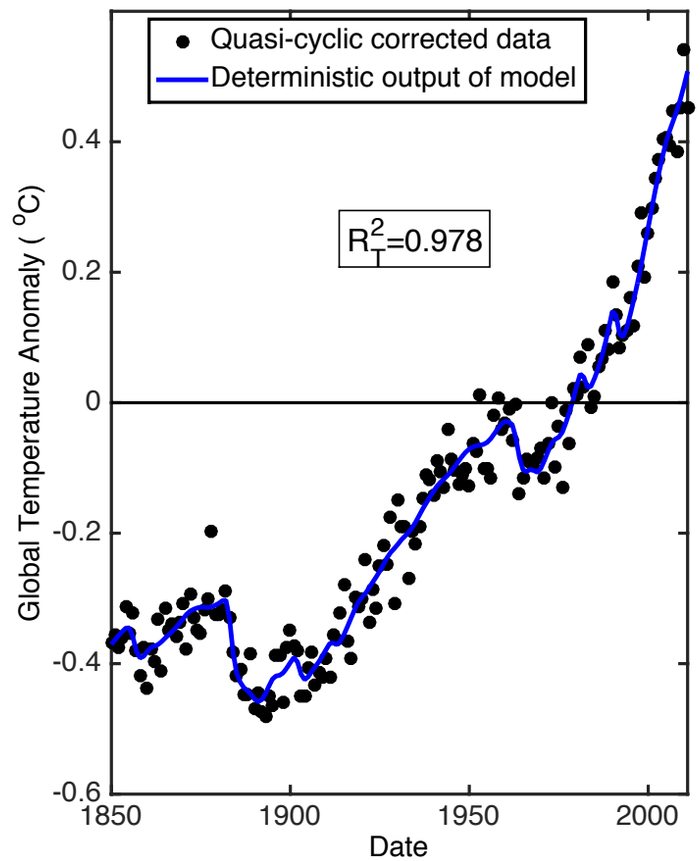




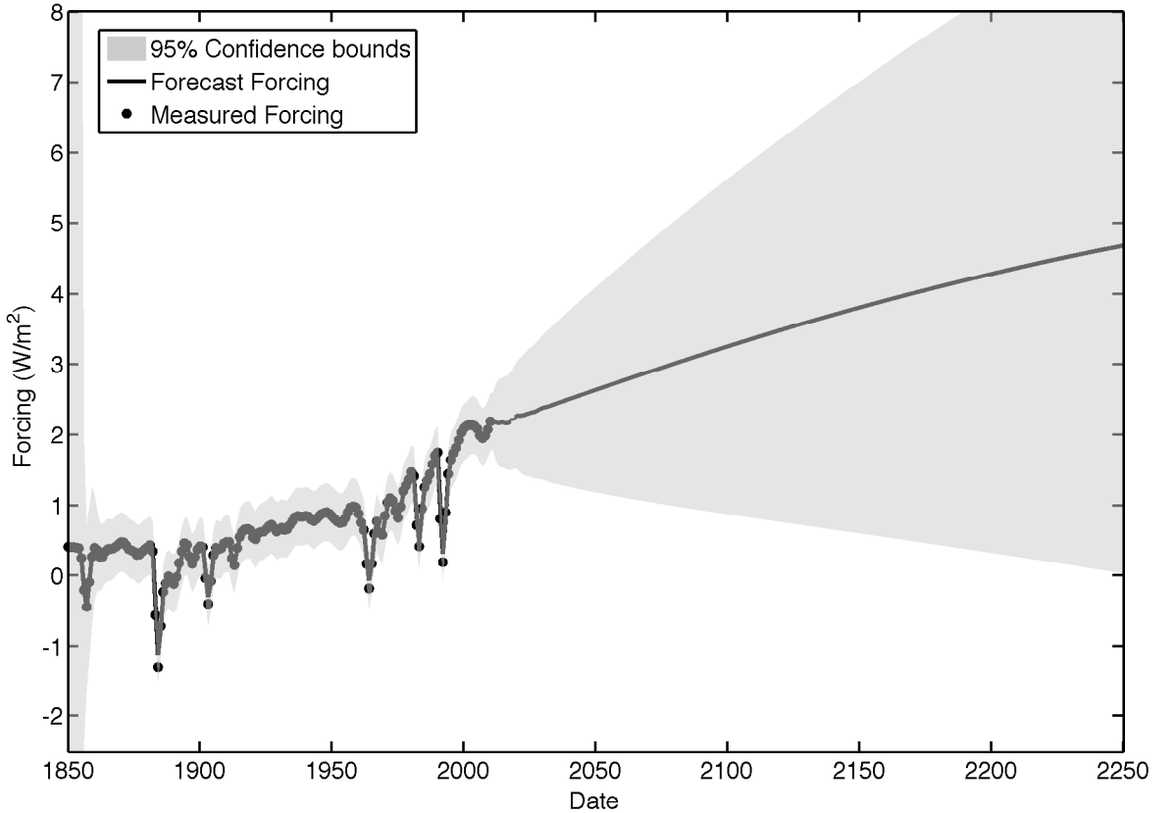
DBM model: contributions to the global temperature anomaly.

Final DBM Model Results Including Quasi-Cycle

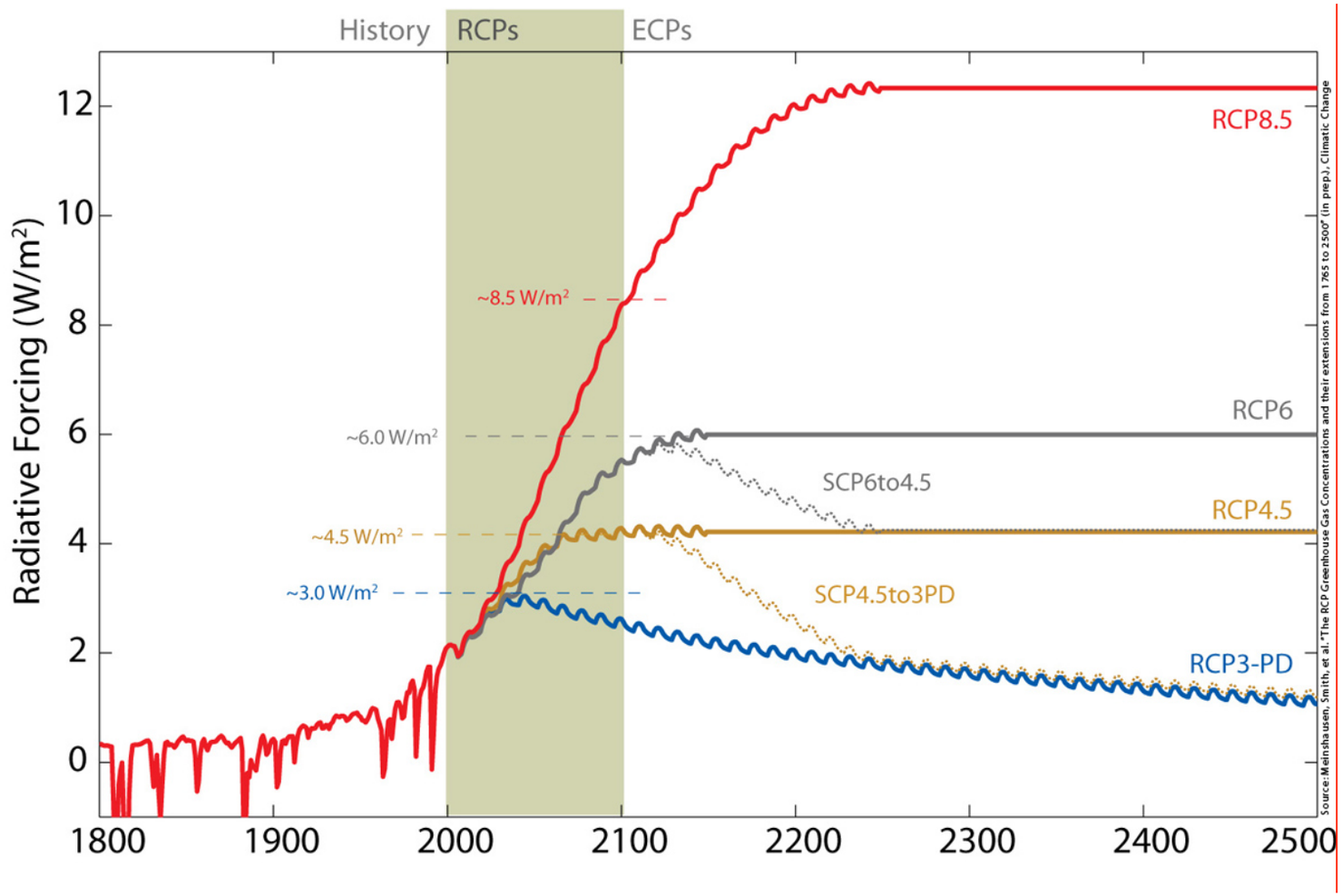




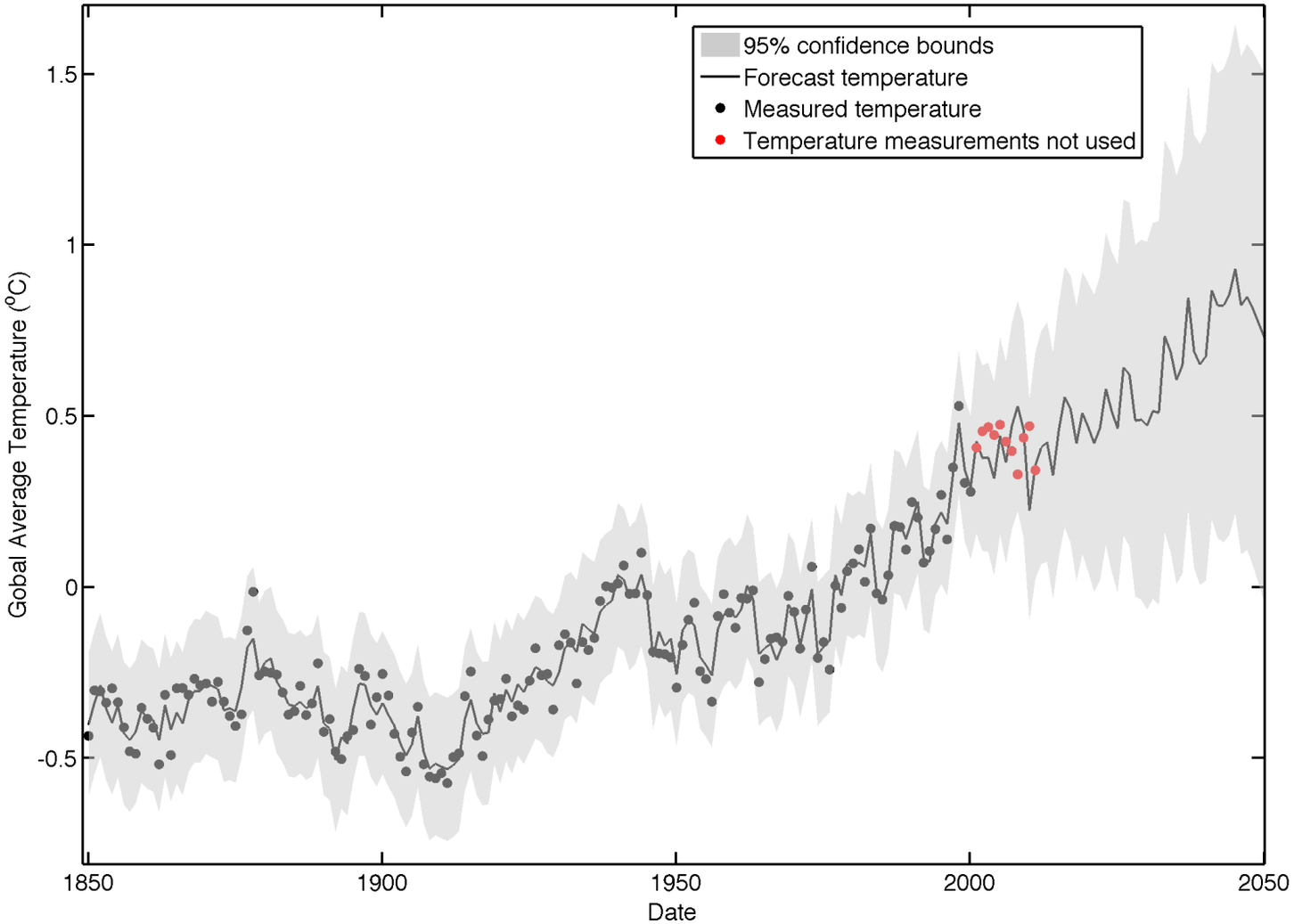
Model-Based Forecasting



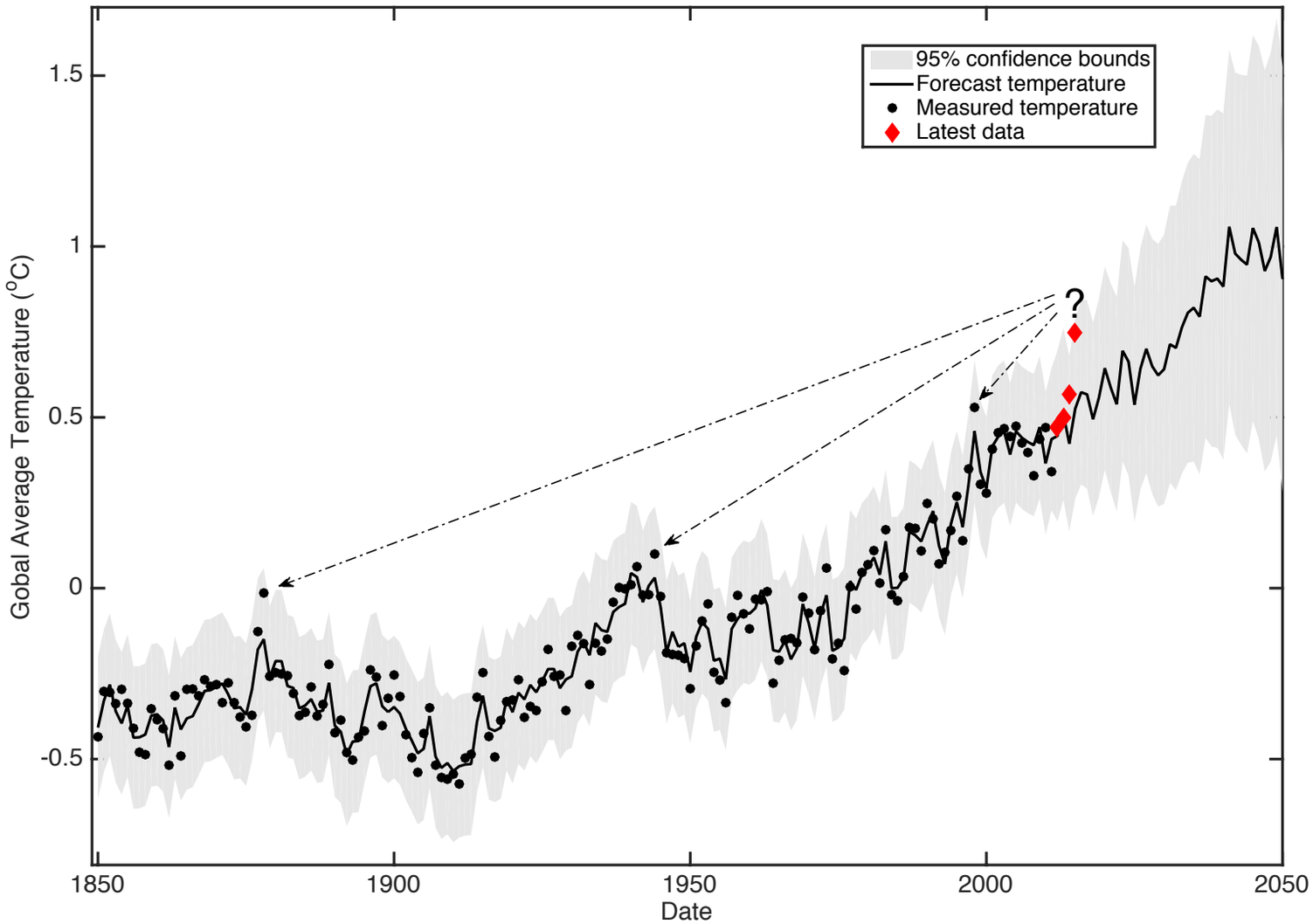
Dynamic Harmonic Regression (DHR) Forecasting of the Forcing Signal $u(k)$.



The DHR Forcing signal forecasts can be compared to the above climate community scenarios.

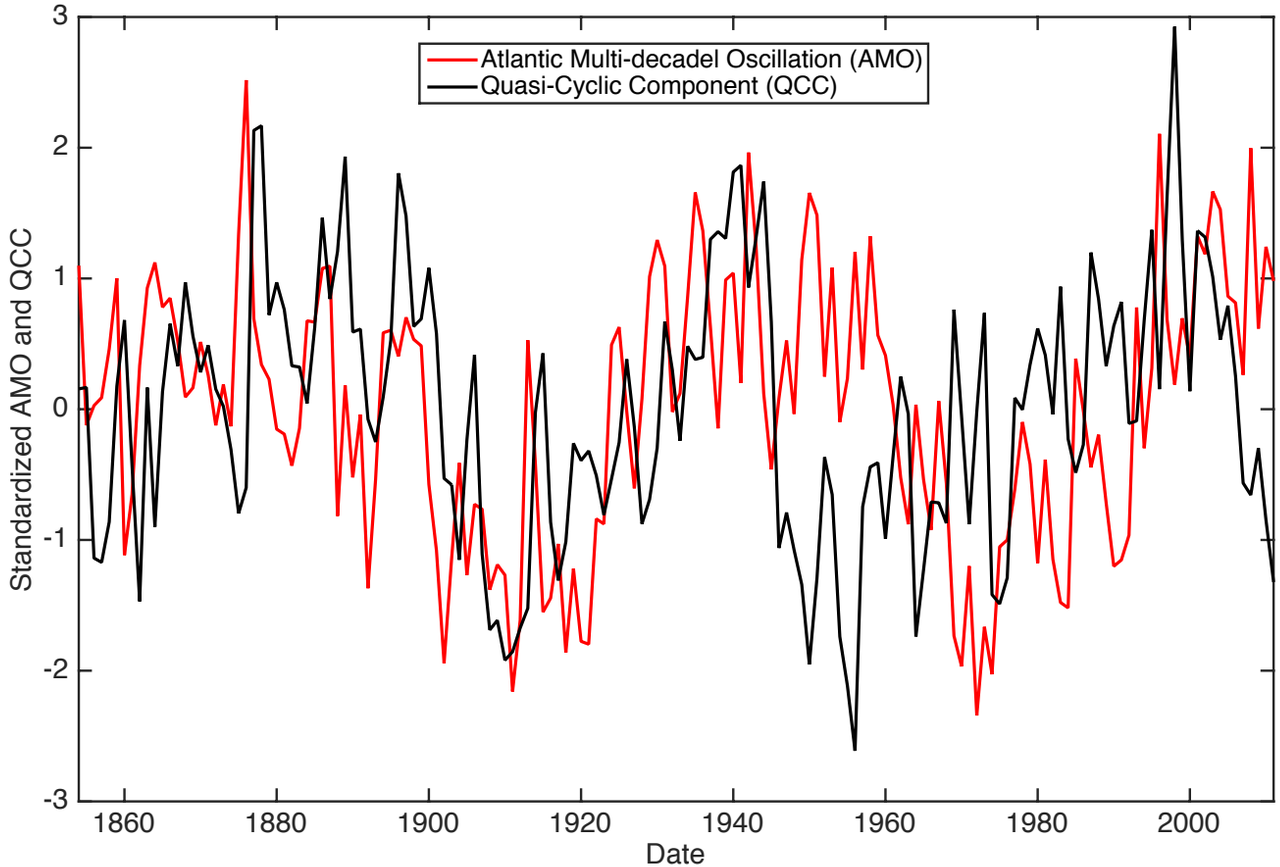


GAT forecast, including quasi-cyclic error: short term, from 2000.



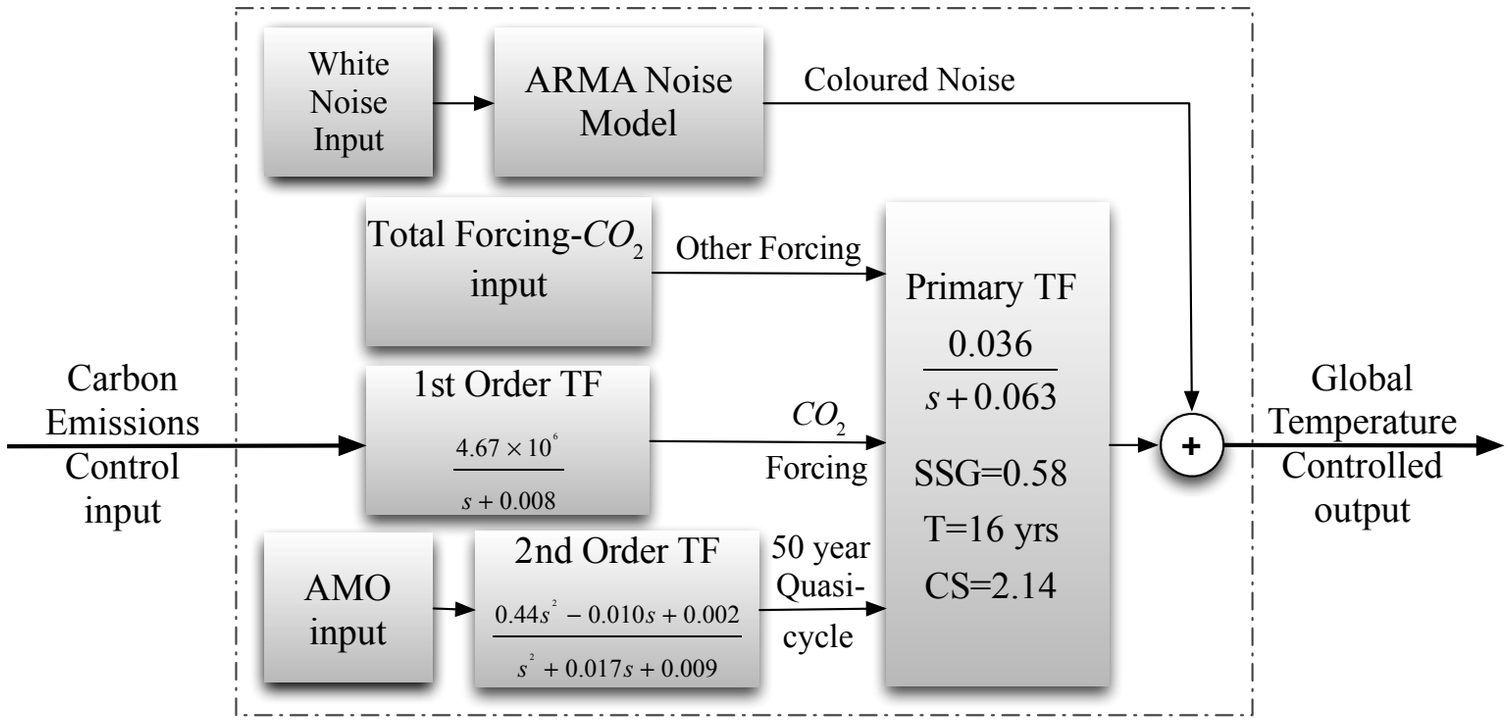
GTA forecast, including quasi-cyclic error: short term, from 2011.

HI-DBM Modelling



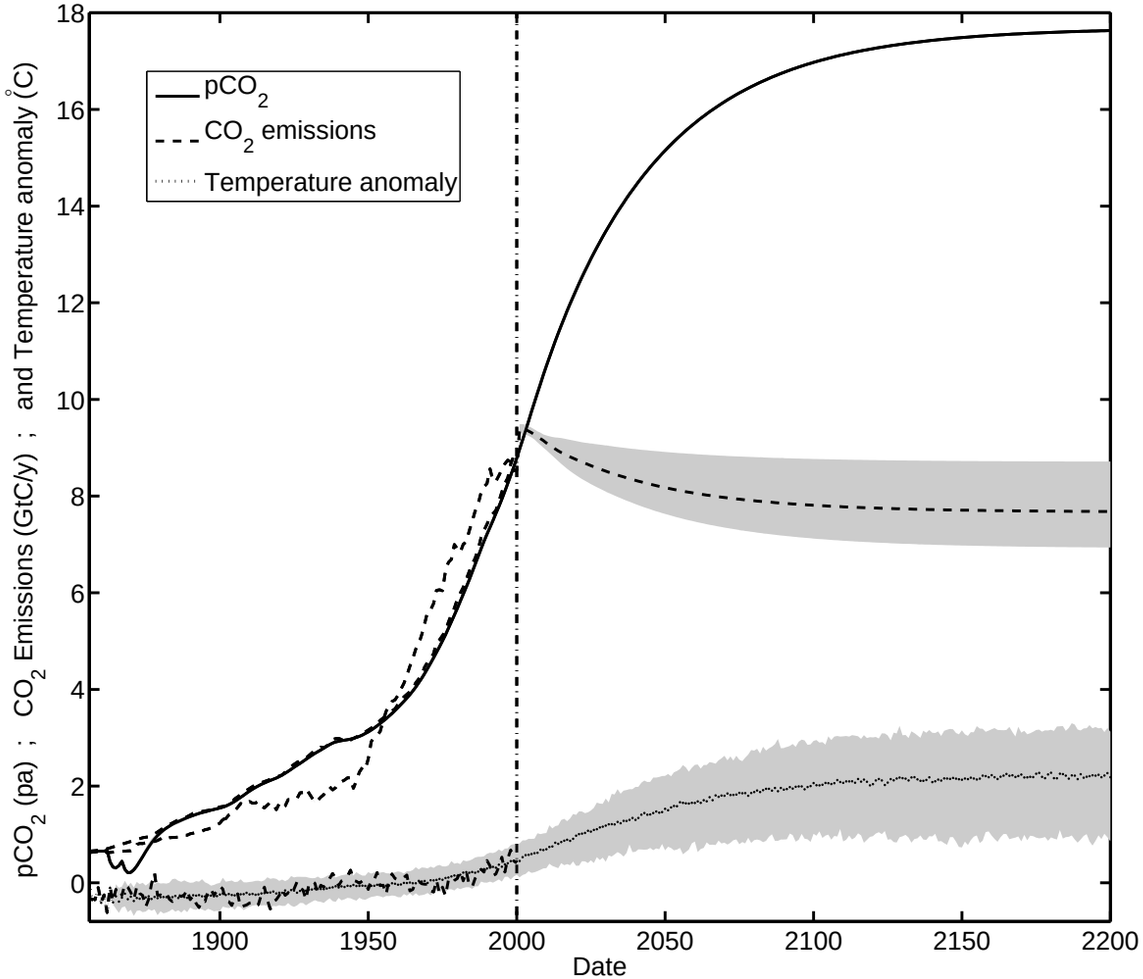
Atlantic Multi-decadel Oscillation (AMO) and Quasi-Cyclic Component (QCC)

Expanded HI-DBM Model for Control Studies



Download the full details of this two input model, *including the introduction of the AMO input*, from report *HIDBM-Climate* at: `http://captaintoolbox.co.uk/Captain_Toolbox.html/Publication_Downloads.html`

Atmospheric CO_2 Control Simulation



The DBM Global Temperature Anomaly Model:

- is fully stochastic and suggests a high degree of uncertainty in the *future forcing* that *appears to match the climate scientists' perceived uncertainty*;
- *accepts the causation assumed by the climate scientists* and, not surprisingly, suggests increased global warming into the future but with a short levelling off period before this (but there is a high level of uncertainty in the forecasts, *as well as* the DBM model coefficients and the parameters derived from these);
- is credible in its structure compared with the MAGICC and reduced order MAGICC models; and provides an independent assessment of the uncertainty and ambiguity in the data used by the climate scientists;
- represents a relatively quick and non-exhaustive analysis. **But shouldn't a more comprehensive, inductive and prejudice-free analysis of this same general type, that emphasizes the uncertainty, be part of the the global warming studies?**