Quantifying the role of solar radiative forcing over the 20th century

Shlomi Ziskin, Nir J. Shaviv *

Racah Institute of Physics, Hebrew University of Jerusalem, Giv’at Ram, Jerusalem 91904, Israel

Available online 17 October 2011

Abstract

The 20th century temperature anomaly record is reproduced using an energy balance model, with a diffusive deep ocean. The model takes into account all the standard radiative forcings, and in addition the possibility of a non-thermal solar component. The model is parameterized and then optimized to produce the most likely values for the climate parameters and radiative forcings which reproduce the 20th century global warming. We find that the best fit is obtained with a negligible net feedback. We also show that a non-thermal solar component is necessarily present, indicating that the total solar contribution to the 20th century global warming, of \( \Delta T_{\text{solar}} = 0.27 \pm 0.07 \) °C, is much larger than can be expected from variation in the total solar irradiance alone. However, we also find that the largest contribution to the 20th century warming comes from anthropogenic sources, with \( \Delta T_{\text{man}} = 0.42 \pm 0.11 \) °C.

Keywords: Solar variability impact; Global climate; Climate modeling

1. Introduction

Several open questions still remain inadequately answered due to the complexity of the Earth’s climate system. One of these questions is the possible role of solar variations in climate change, while another is the value of the climate sensitivity. As we shall see below, the two questions are intertwined with each other, when dealing with 20th century warming, and in particular, when trying to answer the question of attribution. That is, it is impossible to find the relative contribution of different components to the observed warming, without knowing the climate sensitivity or the relative role of the sun.

Presently, a considerable amount of evidence points to a clear link between solar activity and climate variations. These correlations, between solar activity and climate proxies can be seen over a very wide range of time scales. Over the solar cycle, for example, sea surface and land temperatures vary by typically 0.1 °C (e.g., White et al., 1997; Shaviv, 2005). On longer time scales, proxies for solar activity and climate appear to correlate as well (e.g., Eddy, 1976; Neff et al., 2001; Bond et al., 2001; Hodell et al., 2001).

Several possibilities exist to explain the large climatic variations observed. One possibility is that the climate system is simply very sensitive to any changes in the radiative forcing, including the small variations in the total solar irradiance. Alternatively, the large non-thermal solar activity variations could be amplified by a mechanism unrelated to the solar irradiance. Examples include hyper-sensitivity of the climate system to UV (Haigh, 1994) and the solar–wind modulated cosmic ray flux (CRF) (Ney, 1959; Svensmark, 1998). An empirical confirmation to the second possibility (i.e., solar activity amplification) can be found when we observe the heat flux that enters the oceans every solar cycle. This rather large heat flux (Shaviv, 2008) suggests that it is the radiative forcing associated with the solar cycle which is large, and not the climate sensitivity.

When trying to understand the 20th century climate change, there exists another uncertain radiative forcing, and that is the total anthropogenic contribution. The main uncertainty in it is the contribution of the indirect aerosol effect (e.g., IPCC, 2007). When empirically fitting the observed warming to solar activity and anthropogenic contributions, one obtains non-negligible contributions.
by both (e.g., Soon et al., 1996). However, without knowing the sensitivity or the absolute radiative forcing of either contributions, not much more can be concluded.

For example, if the sensitivity is high, the observed solar–climate links could arise from just the variations in the total solar irradiance. The residual left when subtracting the solar component from the 20th century temperature anomaly, can then be attributed to anthropogenic causes. The high sensitivity would imply that the net radiative forcing should be low which implies that the indirect aerosol effect introduces a large net cooling that compensates for the effect of the greenhouse gases.

On the other hand, if the sensitivity of the climate system is low, the solar–climate link should be attributed to some amplification mechanism. The residual should again be attributed to anthropogenic causes. The low sensitivity would imply, however, that the cooling indirect aerosol effect cannot be large, such that the net anthropogenic forcing is.

Both options can be avoided if the climate system has large internal variations, which would allow large discrepancies between the radiative forcing and the average climate. We will not consider this possibility but it should be kept in mind.

Our goal in the present work is to model the 20th century temperature anomalies without any prior biases towards particular sensitivities nor the net size of the solar forcing. We do so with an energy balance model which includes enough free parameters to allow for a wide range of possibilities. By fitting to the observed land and ocean temperature anomalies we expect to find a range of climate and radiative forcing parameters for which the observed temperature can be explained with as little residual as possible.

We begin in Section 2 by describing the model we use. In Section 3, we describe the numerical solution, the optimization of the model parameters and the error analysis. In Section 4, we describe the results of the optimization procedure. We then continue with a discussion in Section 5 and a summary in Section 6.

2. The model

We require a climate model if we are to simulate the 20th century temperature anomaly. However, because our goal is to study and optimize the multi-dimensional “phase space” of the climate parameters, we cannot use a full fledged general circulation model. First, the computer time required to run one simulation is huge. This would inhibit the possibility of running multiple runs and optimize the model parameters. Another problem is that many of the GCM (general circulation model) parameters, such as the global climate sensitivity, are inherent to the model and cannot be systematically controlled. The sole exception is the project named climateprediction.net (e.g., Piani et al., 2005), which used the good will of the public to run thousands of GCM simulations, and in which model parameters that indirectly affect the sensitivity were systematically studied. However, even in this case, not all possible parameters were studied, and in particular that of solar amplification in which we are interested here.

Thus, we follow the same line, for example, as Lindzen and Giannitis (1998) and Hegerl et al. (2006), who used an energy balance model (EBM) to study Earth’s climate sensitivity in response to volcanic eruptions, or over the past millennium respectively. That is, we use an EBM that is able to reproduce a temperature anomaly time series given the parameters which describe the climate and the radiative forcing. In order to compare and fit the model’s output to the observed time series, we also need a good optimization program. Last, after the optimization and fit are carried out, we will also need to carry out a statistical analysis to estimate the errors.

In this section, we describe the details of the model and the various inputs that the model requires. In particular, we describe the parametrization of both the model physics and the model input. We leave the optimization and statistical methods to the subsequent sections, and address the model shortcomings in the discussion afterwards.

2.1. A simple energy balance model

The EBM we use consists of several components and follows the lines of Lindzen and Giannitis (1998, hereafter L&G). First, we have three coupled boxes which represent the climate system, one for the land-covered regions consisting of 30% of the global area, one for the ocean mixed layer, and finally we use a third box to represent the deep ocean. We allow the mixed layer to exchange heat with the two other boxes.

Unlike the land and mixed layer boxes which are assumed to be homogeneous and therefore represented with one temperature, the deep ocean in described with 1D vertical diffusion which serves as a surrogate for all ocean processes acting to carry heat from the mixed layer to the deeper layers. The diffusive layer representing the deep ocean is assumed to have a finite depth. This is done in order to accommodate the effect of a uniform upwelling pattern below 400 meters, that acts to inhibit diffusion of heat to further depths on the multi-decadal time scale.

The second major component of the model is the forcing function that incorporates all the effective radiative forcings that drive the climate system. The forcing function and the temperature response from the various boxes are globally averaged quantities.

Last, we use a single parameter to describe the magnitude of the temperature response of the various boxes to a change in the forcing function. This parameter is the gain parameter and it is closely related to the climate sensitivity as we will see later on. Heuristically, the model is described in Fig. 1. We now proceed to describe it in detail.

Note that there are two main differences between the analysis of L&G and the analysis carried out here. First,
L&G study the climate response to volcanic forcing. Here we study the response to all the main climate drivers over the 20th century, and therefore have to consider more forcings and their uncertainties. Second, we run the model more extensively (i.e., thousands of runs instead of a few), and thus allow all the model parameters to vary systematically, and carry out an unbiased optimization, as opposed to varying only the gain parameter.

In particular, we also allow the ocean diffusivity and land-ocean coupling parameters to change. As a consequence, we can accommodate models with a larger range of possibilities. For example, we can describe models with different long term sensitivities, and different "damping" factors on short time scales. For this reason, L&G obtained a slow decaying response to volcanic forcing when imposing high climate sensitivities, while we can accommodate more possibilities, such as a faster response. In this sense, we can recover both the L&G results and the response obtained with more common GCMs (e.g., Wigley et al., 2005), without a prior bias.

2.1.1. Equations and Parameters

The main equations defining our model are those of energy conservation for the boxes, and heat diffusion for the deep ocean. Heat conservation, per unit area, for the land and for the mixed layer are respectively

\[
C_{\text{land}} \frac{\partial T_{\text{land}}(t)}{\partial t} = \Delta Q(t) + \frac{v}{A_{\text{land}}} [\Delta T_{\text{land}}(t) - \Delta T_{\text{ml}}(t)] - \frac{1}{\lambda} \Delta T_{\text{land}}(t), \tag{1}
\]

\[
C_{\text{ml}} \frac{\partial T_{\text{ml}}(t)}{\partial t} = \Delta Q(t) + \frac{v}{A_{\text{ml}}} [\Delta T_{\text{land}}(t) - \Delta T_{\text{ml}}(t)] - \frac{1}{\lambda} \Delta T_{\text{ml}}(t) + \kappa \frac{\partial^2 T_{\text{deep}}}{\partial z^2} \bigg|_{z=0}. \tag{2}
\]

\[C_{\text{land}}\] and \[C_{\text{ml}}\] are the heat capacities per unit area of the land and mixed layer. \[\Delta Q(t)\] is the prescribed radiative forcing. \[v\] is the land-ocean coupling coefficient, while \[\lambda\] is the climate sensitivity parameter. The respective \[A's\] are the surface area fractions of the land and mixed layer. We discuss the value and meaning of the parameters below.

Heat diffusion in the deep ocean is given by

\[
\frac{\partial T_{\text{deep}}}{\partial t} = \kappa \frac{\partial^2 T_{\text{deep}}}{\partial z^2}, \tag{3}
\]

with the boundary conditions that

\[
\Delta T_{\text{deep}} \bigg|_{z=0} = T_{\text{ml}}, \tag{4}
\]

and

\[
\frac{\partial T_{\text{deep}}}{\partial z} \bigg|_{z=H_{\text{deep}}} = 0. \tag{5}
\]

Here \[\kappa\] is the effective vertical diffusivity of the oceans.

Following the model parameters suggested by L&G, the depth of the deep ocean is taken to be \(H_{\text{deep}} = 400\) m. We have also considered other values, up to 1000m, and obtained virtually indistinguishable results. The depth of the mixed layer is taken to be \(H_{\text{ml}} = 75\) m, which is the global average of the maximum annual mixed layer depth (Shaviv, 2008). Additionally, we take the thermal capacity per unit area of the land to be 30 times smaller than that of...
the mixed layer. The exact value here is unimportant since the small land heat capacity implies that the land reaches thermal equilibrium with its net radiative fluxes in much less than a year, which is the resolution with which the model is compared to the data.

As discussed by L&G, some of the model parameters should be treated as unknowns since they are inadequately constrained by observations, and we should therefore let the optimization procedure determine them. The first of these parameters is the land-sea coupling $v$. The second poorly constrained parameter is the coefficient of eddy heat diffusivity $\kappa$. Direct measurements constrain its value to be between $1.0 \times 10^{-4}$ m$^2$/s and $3 \times 10^{-5}$ m$^2$/s (Law et al., 2003), which we use as general constraints to the parameter which is otherwise left free. This range also includes the diffusivity measured using bomb $^{14}$C (Siegenthaler and Joos, 2003).

Although the parameter $\kappa$ is presented here as a coefficient of eddy diffusivity, its role in the model is to represent all sorts of mixing and heat transfer processes that carry heat into the deep ocean. As we shall see below, this parameter can be constrained once the ocean heat content data is included, to be within its experimental limits. Note also that $\Gamma$ is related to the diffusivity $\kappa$ through $\Gamma = \kappa C_v \leq 126$ W/(m K), where $C_v$ is the volumetric heat capacity of water taken as $4.2 \times 10^6$ J/m$^3$ K.

For convenience, Table 1 proves a nomenclature and list of abbreviations used in this paper.

### 2.1.2. Radiative forcings

The term $\Delta Q(t)$ in the equations above represents the forcing function that includes all the radiative forcings we put into the model. These include among other, well mixed greenhouse gases (GHGs), stratospheric aerosols, reflective tropospheric aerosols and aerosol indirect effect (AIE). These forcings terms are taken from Hansen et al. (2005).

Among the standard radiative forcing terms, by far, the least known is that of the Aerosol Indirect Effect (AIE). This term is negative and describes a cooling effect, but its size is unconstrained. We therefore describe it with a parameter to be determined in the fitting, and of which only the sign is predetermined.

In addition to the standard forcings, we introduce a few more. The first one describes an indirect solar/climate effect (ISE), which may exist in addition to changes in the total solar irradiance (TSI) which we take as well, and which we assume to be proportional to the AA index. We use this geomagnetic index because of several reasons. First, it is an indirect index which can be used to describe the non-thermal activity of the sun. Second, it has a long enough record to cover our simulation time span. And last, we require a proxy which does not have a short time or long time distortion. For example, it would have been better for us to use a cosmogenic isotope, such as $^{10}$Be, however, because of the finite precipitation times and diffusion in ice-cores, the ratio between the 11-year solar cycle and the secular variations is distorted and therefore cannot be used in the present analysis. On the other hand, direct measurements of the cosmic ray flux span only half of the period we simulate.

One should note that the indirect solar effect can be one of several different mechanisms (or a combination of them), such as hypersensitivity of the climate system to UV (Haigh, 1994, Haigh et al., 2010), or climate sensitivity to the atmospheric ionization, governed by solar induced cosmic ray modulation (e.g., Svensmark, 1998; Tinsley, 1990). The present model cannot distinguish between them. It can only address whether such a mechanism, which depends on the non-thermal component of the solar activity, exists.

Table 1

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIE</td>
<td>Aerosol Indirect Effect</td>
</tr>
<tr>
<td>EBM</td>
<td>Energy balance model</td>
</tr>
<tr>
<td>IPCC</td>
<td>Inter governmental panel on climate change</td>
</tr>
<tr>
<td>GHG</td>
<td>Greenhouse gas</td>
</tr>
<tr>
<td>ISE</td>
<td>Indirect solar effect</td>
</tr>
<tr>
<td>LST</td>
<td>Land Surface Temperature</td>
</tr>
<tr>
<td>NAO</td>
<td>North Atlantic Oscillation</td>
</tr>
<tr>
<td>OHC</td>
<td>Ocean heat content</td>
</tr>
<tr>
<td>PDF</td>
<td>Parameter distribution function</td>
</tr>
<tr>
<td>PDO</td>
<td>Pacific Decadal Oscillation</td>
</tr>
<tr>
<td>SOI</td>
<td>Southern Oscillation Index</td>
</tr>
<tr>
<td>SST</td>
<td>Sea Surface Temperature</td>
</tr>
<tr>
<td>TSI</td>
<td>Total solar irradiance</td>
</tr>
<tr>
<td>$A_{\text{land},\text{ml}}$</td>
<td>Surface area of the land/mixed-layer</td>
</tr>
<tr>
<td>$C_{\text{land},\text{ml}}$</td>
<td>Total land/mixed-layer heat capacity per unit area</td>
</tr>
<tr>
<td>$C_v$</td>
<td>Volumetric heat capacity of water</td>
</tr>
<tr>
<td>$H_{\text{mol}}/H_{\text{deep}}$</td>
<td>Depth of the mixed layer/deep ocean (Thermocline)</td>
</tr>
<tr>
<td>$\Delta T_i(t)$</td>
<td>Temperature anomaly of component $i$</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Heat diffusion constant in the oceans</td>
</tr>
<tr>
<td>$k$</td>
<td>Climate sensitivity</td>
</tr>
</tbody>
</table>

Fig. 2. A comparison between the indirect solar effect (ISE) forcing with $x_{\text{AA}} = 1$, the total solar irradiance (TSI) forcing (long dashed) and the Aerosol Indirect Effect (AIE) forcing, assuming $x_{\text{AIE}} = 1$ (dotted).
for (ISE), is proportional to the AA index. In particular, we assume that it has the same phase (see Fig. 2).

We also add internal climate forcings – proportional to the Pacific Decadal Oscillation (PDO) index and the Southern Oscillation Index (SOI). In principle, such internal oscillations should arise from an ideal climate model. However, even the standard GCMs do not predict them. We therefore prescribe them “by hand”. This is done because they improve the fit to the observations, which in turn helps to better constrain the other model parameters.

In many previous analyses, the SOI and occasionally the PDO components were removed from the observed SST through different statistical procedures. We chose an alternative, which is to use our own optimization tool, to obtain the SOI and PDO components in the oceanic and land temperature data. Both procedures effectively remove the SOI and PDO components from the observed data set and should produce the same results. Our approach has the disadvantage that it conceptually treats the PDO and SOI as if they are external forcings affecting the climate, clearly this is not the case. The advantage is that it is much simpler to implement. In one run, we have also incorporated the North Atlantic Oscillation (NAO) index, but as explained in Section 4.3.1, we have decided to have only the SOI/PDO in the nominal runs.

To sum up, our model accepts the various radiative forcings as knowns, with exception of the new ones (ISE, SOI and PDO indices) and the Aerosol Indirect Effect (AIE), which we take as a cooling only effect.

2.1.3. The gain parameter and climate sensitivity

One of the more important parameters we discuss is the equilibrium climate sensitivity, defined as \( \lambda \equiv \Delta T_{\text{equil}} / \Delta Q \). This sensitivity can also be expressed as \( \lambda \equiv \lambda_{\text{BB}} \times \text{gain} \). Here \( \lambda_{\text{BB}} \approx 0.3 \, ^\circ C/(W \, m^{-2}) \) is the so called black body sensitivity, which is the sensitivity of a feedback-less Earth.\(^1\) The gain parameter refers to the amplification of the equilibrium temperature response relative to the ideal sensitivity.

2.2. Observed temperature anomaly records

The time series generated by the model should be compared with real observations of 20th century variations in the land and ocean temperatures. We use the observed Land Surface Temperature (LST) and the Sea Surface Temperature (SST) from the National Climatic Data Center (NCDC). The NCDC uses new and improved land and sea data sets that are taken from Smith and Reynolds (2005).

Note that we do not know a priori the absolute offset of either the land or ocean data. That is, the zero point radiative forcing corresponds to some equilibrium temperatures for the land and ocean, but since neither were ever in a long equilibrium with the radiative forcing, the value of these zero point temperatures anomalies are unknown, and are therefore left as a free parameter. Technically, we define

\[
T_{\text{land}}^{\text{new}}(t) = T_{\text{land}}^{\text{old}}(t) + T_{\text{land}}^0, \\
T_{\text{ml}}^{\text{new}}(t) = T_{\text{ml}}^{\text{old}}(t) + T_{\text{ml}}^0,
\]

where \( T_{\text{land}}^0 \) and \( T_{\text{ml}}^0 \) are the parameters that determine the necessary shift. Note that the shift need not be the same for the land and mixed layer boxes. The deep ocean box is coupled to the mixed layer box in such way that the radiative forcing does not influence it directly, only through the mixed layer box.

The last free parameters are \( \Delta T_{\text{land}}^{\text{initial}} \) and \( \Delta T_{\text{ml}}^{\text{initial}} \), which determine the initial temperature anomalies at the land and mixed layer boxes. Without better assumptions, we assume that initially the deep ocean anomaly is the same as the mixed layer.

2.3. Ocean heat content data

In several runs, we also include the annually averaged ocean heat content data (Levitus et al., 2009) between 1955 and 2003, to check whether it can better constrain the model parameters.

In these runs, the model calculates each time step the total heat content in each layer of the ocean and vertically integrates over. This can then be compared to the observed value.

As we shall see in Section 4.5, the only parameter that is better constrained using this procedure is the effective ocean diffusivity \( \kappa \).

3. Numerical methods

The backbone in our analysis is the computation of a 20th century climate for given climate model parameters (i.e., a single run, in Section 3.1). The next steps are to carry out many model integrations and optimize the climate and forcing parameters such that the model predictions best agree with observations. The last step is to carry out a statistical analysis, in order to estimate the predicted errors in the model parameters.

3.1. Solving for a single run

In essence, our model consists of ordinary differential equations (Eqs. (1) and (2)), and a parabolic differential equation (Eq. (3)), which are coupled through the boundary conditions given by Eqs. (4) and (5).

To solve the parabolic equation, we discretize the Laplacian operator and divide the deep ocean into \( N = 16 \) layers of 25 m each, totaling in 400 m of depth. This transforms the parabolic partial differential problem into \( N \) ordinary

---

\(^1\) Actually the black body sensitivity is a common misnomer. It should be called the gray body sensitivity, having a fixed albedo and emissivity which are not 0 and 1 respectively.
differential equations. Together with the mixed layer and land boxes, we have \(N + 2\) ODEs.

Note that because we are solving the parabolic equation explicitly (that is, the spatial derivative is calculated in the old time step), the time steps must be small enough to satisfy the Courant–Lewy condition, namely that \(4\kappa\Delta t/\Delta x^2 < 1\).

The set of ordinary equations are then solved using the 4th order Runge–Kutta with a constant time step. The time integration starts at 1880 and ends at 2003. The time step we choose is typically 0.01 yr, which was found to satisfy the Courant–Lewy condition for any diffusivity solved for, and conserve energy at the \(10^{-3}\) level, after a century.

### 3.1.1. Applying the radiative forcing

For the standard forcings we use a tabulated form of the radiative forcings described in Hansen et al. (2005). To this table we add the ISE which is proportional to the AA Index (Mayaud, 1972), the SOI (Trenberth, 1984) and the PDO Index (Zhang et al., 1997; Mantua et al., 1997). The data is annual such that we obtain 124 data points representing yearly averaged forcings from 1880 to 2003. The data is then interpolated to the 0.01 yr time steps. The initial tabulation of high resolution data improves the speed considerably, which is why the system is integrated with a fixed time step.

As mentioned above, some of the forcings are known well enough and therefore can be directly used in the model. The other forcings are treated as unknowns. Thus, before being used in the model, each unknown forcing source term is multiplied by a coefficient that represents the magnitude of that forcing, which is to be determined in the optimization. We have 4 such parameters, for the ISE, the PDO index, the SOI, and the AIE. These parameters are then optimized together with the climate parameters of the model. These parameters are summarized in Table 2.

### 3.1.2. The fit function

After constructing the radiative forcing table and interpolating it, we require a fit function to evaluate the goodness of the fit between the predicted temperature evolution and the observed time series. We use the standard \(\chi^2\) test, as follows:

\[
\chi^2 = \sum_i \left( \frac{O_{\text{sat},i} - P_{\text{ml},i}}{\sigma_{\text{sat}}} \right)^2 + \sum_i \left( \frac{O_{\text{sl},i} - P_{\text{land},i}}{\sigma_{\text{sl}}} \right)^2,
\]

where the index \(i\) runs over the range of years used, between 1880 and 2003. For the weight coefficient, we take the values of \(\sigma_{\text{sl}},\sigma_{\text{sat}}\) to be 0.143 and 0.058 °C for the LST and SST respectively. These coefficients are the standard deviations obtained from a simple pairwise comparison:

\[
2\sigma^2 = (O_{i+1} - O_i)^2.
\]

This analysis measures the annual variability of the dataset, which arises from both internal variance (i.e., “climatic noise”) and measurement errors. Evidently, the LST is almost 2.5 times “noisier” than the SST, which arises from the damping effect of the high heat capacity of oceans.

Note that the observed time-series comes in a form of 124 data points representing years. However, the model resolution is much higher. Thus, we interpolate the observed time series to the higher resolution needed for the model.

### 3.2. Genetic optimization

Once a goodness-of-fit is obtained, the next step is to optimize the model and radiative forcing parameters to minimize the goodness-of-fit, i.e., the \(\chi^2\). We use the PIKAIYa package, which is a genetic algorithm (GA) based optimization package (Charbonneau, 1995).

GA is a class of search techniques inspired from the biological process of evolution by means of natural selection. The main advantage of numerical optimization using GA is the robustness in problems having a multi-dimension search space. That is, GA is usually very effective at finding global minima. Their main disadvantage is a slower speed, since more points in the multi-dimensional phase space should be sampled (e.g., \(10^6\) integrations here), than in other techniques, such as the steepest gradient method. However, the noisy nature of the observations implies that the fitness function has many local minima, and there is no option but to sacrifice speed for robustness.

### 3.3. The bootstrap error analysis

Critical to our analysis is the error estimation. However, it is extremely difficult to assess all the errors coming into

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta T_{\text{land}/\text{ml},i})</td>
<td>Initial land/mixed layer temperature (in 1880)</td>
<td>(-2^\circ) to (+2^\circ)</td>
<td>(-2^\circ) to (+2^\circ)</td>
</tr>
<tr>
<td>(\Delta T_{\text{land}/\text{ml},0})</td>
<td>Land/mixed layer temperature at 0 radiative forcing</td>
<td>(-2^\circ) to (+2^\circ)</td>
<td>(-2^\circ) to (+2^\circ)</td>
</tr>
<tr>
<td>(\kappa)</td>
<td>Coefficient of eddy heat diffusivity</td>
<td>(1.0 - 30.0 \times 10^{-6}) m²/s</td>
<td>(1.0 - 30.0 \times 10^{-6}) m²/s</td>
</tr>
<tr>
<td>(v)</td>
<td>Land–ocean coupling coefficient</td>
<td>0.05–20 W/m °C</td>
<td>0.05–20 W/m °C</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>Coefficient of climate sensitivity</td>
<td>0.1–4.0</td>
<td>0.1–4.0</td>
</tr>
<tr>
<td>(s_{\text{PDO}})</td>
<td>Coefficient that multiplies the PDO</td>
<td>(-4) to (+4)</td>
<td>(-4) to (+4)</td>
</tr>
<tr>
<td>(s_{\text{SOI}})</td>
<td>Coefficient that multiplies the SOI</td>
<td>(-4) to (+4)</td>
<td>(-4) to (+4)</td>
</tr>
<tr>
<td>(s_{\text{AIE}})</td>
<td>Coefficient that multiplies the AIE</td>
<td>0.01–4</td>
<td>0.01–4</td>
</tr>
<tr>
<td>(s_{\text{AA}})</td>
<td>Coefficient that multiplies the AA index</td>
<td>(-4) to (+8)</td>
<td>(-4) to (+8)</td>
</tr>
</tbody>
</table>
the model and its integration, as some of the errors arise from measurement while some from intrinsic climate noise. In order to avoid this problem altogether, we use the bootstrap method to carry out a reliable error estimation (e.g., Press et al., 1986). This method is often used when there is insufficient information about the underlying statistical processes or the nature of the measurement errors.

In the bootstrap method, the same analysis is carried out many times over but with a degraded data set. Each analysis is carried out with a data set that randomly disregards \(1/e \approx 37\%\) of the original data points, by taking other points instead. Statistically, it can be shown that the distribution of the set of optimization parameters obtained under the different bootstrap realizations follows the probability distribution function for the parameters as should be obtained for the best fit.

The average and variance for the obtained distribution can then be calculated, though one cannot be assured that the distribution should follow the normal distribution.

The method also addresses another issue. Although the GA is generally a robust method to find global minima, we cannot be assured that the minimum found is indeed the global one. By “degrading” the data, we are forcing the algorithm to sample a larger phase space. If there are several local minima, then the algorithm will sample them and provide a wider range of parameters. In this sense, we obtain a conservative error estimate.

4. Results

We now proceed with the analysis of the results. We begin with the model’s ability to reproduce the 20th century temperature anomaly record. We will then present the optimal values for the various parameters of the model and their error estimation. This will allow us to ascertain the size of the different contributions towards the observed 20th century global warming. Last, we will discuss the correlation in the parameter uncertainties, compare to previous analyses, and the effect of different assumption on the results.

4.1. 20th century reconstructed

Figs. 3 and 4 represent the best fit for the land temperature anomaly and the mixed layer anomaly respectively. Fig. 5 is the model’s best fit to the global temperature record, which consists of 30% of the land and 70% of the mixed layer. Fig. 6 depicts the 3-year running average of the observed and model fit global temperature. Evidently, the residual for the mixed layer is smaller than it is for the land. The reason for this is that the LST data set is much noisier (presumably because of higher climatic fluctuations). However, because the relative size of the fluctuations was taken into consideration when performing the fit (see Eq. (7)), the size of the error relative to the fluctuations is similar in both cases.
The total effective number of degrees of freedom for the fit is \(2 \times 124 - 11 = 237\), where \(2 \times 124\) is the number of annual data points (for the ocean and land data), while 11 is the number of model parameters. The \(\chi^2\) obtained per effective degree of freedom is 2.42. The large size implies that besides the annual fluctuations (due to measurement noise and year to year climate variability), there is also multi-year variability which is not accounted for by the model, of about \(\leq 0.1\) °C (with the maximum annual residual being 0.2 °C).

It is hard to compare EBM results to the results of GCM fits, because the latter include random climate noise (like the real system does), which implies that the residuals from a perfect GCM would be a factor of \(\sqrt{2}\) larger than the residuals of a perfect EBM fit. Nevertheless, fits by GCMs (e.g., Stott et al., 2000) typically produce multi-year residuals of \(\leq 0.2\) °C (and maximum annual residuals of up to 0.4 °C). In other words the present EBM can be considered successful.

### 4.2. Main climate drivers

Fig. 7 summarizes the PDFs for the main parameters in the model, as obtained with the bootstrap method. From the PDFs we can learn that some of the climate and radiative forcing parameters are better determined, while others cannot be significantly constrained.

For example, the AIE is poorly constrained. This is because there is the well known degeneracy between the positive contribution of the GHGs and the negative AIE, both of which monotonically increase in absolute value over the 20th century. As a consequence, there is a degeneracy that the optimization algorithm cannot resolve.

On the other hand, because there is no degeneracy between the ISE forcing and other contributions, this term is much better constrained. Similarly, because the climate sensitivity (i.e., the gain) affects different forcings which are non-degenerate with other forcings (such as the stratospheric aerosols), changing the sensitivity causes changes which cannot be compensated for with other parameters. The result is that the sensitivity can too be significantly constrained.

Perhaps the most interesting question we can attempt to answer with the present analysis is the relative role of anthropogenic and total solar forcings over the 20th century. Fig. 8 depicts the optimal forcings of the two categories.

The anthropogenic contributions consist of primarily the GHG’s and aerosols while the solar contributions are those of the irradiance (for which the standard value of the forcing is assumed) and the ISE. The stratospheric aerosols are also important but they fall into neither category.

To obtain the net effect of each type of contribution, we take the optimized total solar forcings and total anthropogenic forcings (see Fig. 8) and run them through the model again, this time separately. The result, which is depicted in Fig. 9, is the separate contribution of each component to the global temperature increase through the entire 20th century. The PDF of the average temperature increase over the 20th century caused by each of contribution types is given in Fig. 10.

Quantitatively, the total solar forcing and temperature contributions to the 20th century that we obtain are

\[
\Delta F_{\text{solar}} = 0.8 \pm 0.4 \text{ W m}^{-2},
\]

\[
\Delta T_{\text{solar}} = 0.27 \pm 0.07 \text{ °C}.
\]

The corresponding anthropogenic contributions are:

\[
\Delta F_{\text{man}} = 2.0 \pm 0.3 \text{ W m}^{-2},
\]

\[
\Delta T_{\text{man}} = 0.42 \pm 0.11 \text{ °C}
\]

Last, we obtain that the optimal gain (i.e., relative to a feedback-less Earth) is:

\[
gain = 0.85 \pm 0.3,
\]

which yields a climate sensitivity of

\[
\lambda = 0.25 \pm 0.09 \text{ °C/(W m}^{-2}).
\]

### 4.3. Role of internal oscillations

Another interesting aspect is the role that the PDO and SOI play in the global climate. They were added to the model in order to increase the fitness, under the notion that they may explain some of the variability in the temperature record. Indeed, we find in our analysis that the introduction of the two signals produces a better fitness to our model.

Without the signals inserted, the \(\chi^2\) per degree of freedom obtained was 2.85, while it decreased to 2.42 with the signals included, i.e., a reduction of 15% in the residual annual variance. Fig. 11 depicts these signals after the model determined their best estimated magnitudes. Apparently, the signals contribute a non-negligible radiative forcing, but their net long term contribution is small, as can be seen in Fig. 9.
The best fit for the SOI is with a negative SOI/forcing factor. That is, El Niño events (with a negative SOI) cause global warming as expected. We also see that there is a smaller global forcing associated with the PDO variations. 

Quantitatively, the optimal fits and bootstrap analysis give the following radiative contributions from the PDO and SOI signals:

\[
\Delta F_{\text{PDO}}(t) = (0.17 \pm 0.17 \text{ W m}^{-2}) \times \text{PDO}_{\text{index}}(t),
\]

\[
\Delta F_{\text{SOI}}(t) = (-0.56 \pm 0.13 \text{ W m}^{-2}) \times \text{SOI}_{\text{index}}(t).
\]

The best fit for the SOI is with a negative SOI/forcing factor. That is, El Niño events (with a negative SOI) cause global warming as expected. We also see that there is a smaller global forcing associated with the PDO variations.

Quantitatively, the optimal fits and bootstrap analysis give the following radiative contributions from the PDO and SOI signals:

\[
\Delta F_{\text{PDO}}(t) = (0.17 \pm 0.17 \text{ W m}^{-2}) \times \text{PDO}_{\text{index}}(t),
\]

\[
\Delta F_{\text{SOI}}(t) = (-0.56 \pm 0.13 \text{ W m}^{-2}) \times \text{SOI}_{\text{index}}(t).
\]

The NAO

We can try and improve the fit by adding additional “internal oscillations”. The next oscillation expected in relative importance after the SOI and PDO, is the North Atlantic Oscillation (NAO, Hurrel, 1995). We therefore carry out a run which includes an additional forcing, \( \Delta F_{\text{NAO}}(t) = x_{\text{NAO}} \times \text{NAO}_{\text{index}}(t) \), and therefore an additional free parameter.
In addition, the ocean–land coupling becomes undetermined once the NAO is included, as it obtains a very wide PDF. In any case, because there is virtually no improvement in the $\chi^2$, we do not include the NAO in the additional analyses.

4.4. Correlation between climate parameters

Because the parameter phase space is more than one dimensional, one can study higher order probability distribution functions, containing two parameters or more. In Figs. 13 and 14, we plot the distribution of the bootstrap results as projected onto various 2-parameter planes in the high-dimensional parameter space. This allows us to see correlations in the model results.

For example, we see that if the climate has a higher sensitivity than our best estimate, the model would prefer a smaller solar contribution and a larger anthropogenic contribution. But our choice of a specific sensitivity will not change the best estimate for the size of the indirect aerosol effect, which as mentioned before, is not really constrained by the model.

4.5. Effects of the ocean heat content data

An important data set is that of the ocean heat content (OHC). In principle, it provides an independent constraint on the model, and we should consider it as well. The problem, however, is that the OHC suffers from significant systematic uncertainties. This can be seen by comparing the OHC derived by different groups. For example, Ishii et al. (2006) and Levitus et al. (2005) found that the OHC increased by about $5 - 6 \times 10^{22}$ J between 1975 and 2000. On the other hand, Levitus et al. (2009) and Domingues et al. (2008) found an increase of $13 - 14 \times 10^{22}$ J over the same period. Clearly, with such large discrepancies it is impossible to trust any conclusions based on the OHC data.

Nevertheless, we ran a version of our model which includes the comparison to the ocean heat content as a further constraint. In this model, we add to the fit function (Eq. (7)) a sum over the 49 OHC data points, which are normalized by the pairwise calculated standard deviation (Eq. (8)). We carry out this analysis to provide us with a flavor of what to expect once the OHC data will be reliable.

Fig. 15 shows the obtained PDF for the model diffusivity, which is $\kappa = 2.3 \pm 0.6 \times 10^{-5}$ m$^2$/s. With the OHC, both the SOI and the ocean diffusivity are better constrained. Nevertheless, one should note that the effective vertical diffusivity that the model finds is not necessarily the vertical eddy diffusivity, as it may include other mixing processes. However, the fact that the best fit value is within the observed range suggests that other such processes may at most be as important as the eddy diffusivity itself.

Other model parameters, such as the ISE and the gain are somewhat modified from the analysis which does not consider the OHC. Specifically, the best estimates of both
the ISE and gain decrease. For example, the gain becomes $0.75 \pm 0.25$. Although it is clear that the OHC can be useful in better constraining the climate parameters (in particular, those directly related to the oceans), we do not adopt these results because of the inherently large systematic errors that the OHC introduces.

4.6. Comparison with previous analyses

To understand why our results differ from previous analyses, we shall consider more cases which correspond to different assumptions that were either explicitly or implicitly assumed in previous models.
4.6.1. Effects of Constraining the Aerosol Indirect Effect – AIE

In Section 4.4, we showed that a specific choice of sensitivity will not change the best estimate for the size of the indirect aerosol effect, because it is not constrained by the model. But, would a specific choice for the AIE change the climate sensitivity? In order to properly answer this question, we first consider only the TSI variations as a solar forcing (i.e., while neglecting a possible ISE) and introduce upper boundaries to the AIE (since it is negative, upper boundaries are lower boundaries for the absolute cooling effect). These are set at $-0.5 \text{ W m}^{-2}$ and $-1.0 \text{ W m}^{-2}$ in the two cases we consider.

The results of this analysis is shown in Fig. 16. As can be seen, the gain increases to 1.19 for the $-0.5 \text{ W m}^{-2}$ case and to 1.45 for the $-1.0 \text{ W m}^{-2}$ case. In terms of climate sensitivity, these gains translate to $\Delta T_{2x}$ of 1.5 °C and 1.8 °C respectively. A possible explanation for this result is that by increasing the cooling effect (AIE), the model compensates for the larger cooling by enhancing the warming effect of GHG’s and other warming forcings, in order to reach the best fit solution to observations. It does so by increasing the climate sensitivity parameter (gain). This is consistent with the results obtained by Andronova and Schlesinger (2001), who, using an EBM, find high climate sensitivities when assuming large aerosol forcing (AIE). The median sensitivity obtained by their model is 5.9 °C for CO$_2$ doubling. If they add the solar forcing (TSI variations only), they obtain a median of 3.2 °C. Interestingly enough, they also find with very high probability, that the median climate sensitivity is 1.4 °C for the case when solar forcing (TSI) is introduced but without the AIE. The reason that large sensitivities are ruled out when the solar forcing is introduced is because a large sensitivity will give rise to 11-year temperature variations larger than those seen in the temperature data. This is confirmed with the next analysis when decadally averaging the data.

4.6.2. Using Decadally Averaged Data

When using decadally averaged temperature data, the model produces higher sensitivity values. This can be seen in Fig. 16, where decadally averaged data gives a mean gain of 1.4, i.e., a sensitivity of 1.8 °C. This value is very close to the gain obtained when we place an upper limit of $-1.0 \text{ W m}^{-2}$ on the aerosol effect. A possible explanation could be that such averaging of the data erases information about the TSI variations and thus, higher sensitivities are permissible as the fit function is insensitive to the mismatches.
over the 11-year solar cycle. This can explain the results of Forest et al. (2002) who decadally averaged the data, and found a sensitivity of 2.5 °C.

When decadally averaging the data and placing an upper limit of −1.0 W m⁻² to the AIE, we obtain an even higher sensitivity, of 2.75 °C. These results are summarized in Fig. 16.

5. Discussion

The climate model employed in this analysis is relatively simple. It is an EBM which includes several “boxes” and a diffusive ocean. Nevertheless, it is rich enough to describe many of the basic aspects of the climate, and parameterize them. This allows for an extensive parameter study, something which is nearly impossible with a full global circulation model.

With the model, we have shown that the observed land and ocean anomalies can to a large extent be described as a response to the theoretical radiative forcings. Statistically, the fit improves considerably if we include a solar forcing driver (ISE) other than the variations in the total solar irradiance (TSI). In fact, we can rule out the no-nonthermal component assumption at the 2% level.

The fit further improves by introducing the PDO and SOI signals. These describe internal oscillations of the climate, but they introduce an interesting radiative forcing feedback, of which the several year average can be as large as 0.5 W m⁻².

With all the additional radiative forcings, the residual left after fitting the observed parameters is typically ≤0.1 °C. This is twice the expected value from the year to year variability, implying that there are more internal climate variations which are unaccounted for, but they could simply be internal fluctuations. On the other hand, the residual we obtain is typically half of the residual obtained in GCMs. One reason is the fact that GCMs have less calibration freedom than the energy balance model that we use. For example, the climate sensitivity or ocean diffusivity are inherent characteristics of a given GCM and cannot be readily changed. Moreover, GCMs typically have a higher sensitivity than we find, therefore they tend to have exaggerated responses to different radiative forcing variations, in particular, to volcanic eruptions (e.g., see Lindzen and Giannitsis, 1998). In fact, the optimal climate sensitivity we find is somewhat lower than can be expected for a black body Earth. For comparison, the canonical IPCC sensitivity range is about 1.3 to 4 times the black body value. The third reason is that GCMs have internal variability which EBMs lack. This internal variability increases the residual.

The total 20th century solar forcing that we find is 0.8 ± 0.4 W m⁻². This is much higher than the estimated contribution of the TSI variations alone (of 0.1–0.2 W m⁻²). Thus, we conclude like Shaviv (2008), that an additional solar activity amplification mechanism should exist.

It is interesting to note that the constraint we derive on the climate sensitivity is not primarily due to the correlation between solar activity and temperature. This is because the response to the solar activity depends on both the unknown sensitivity and unknown solar amplification. Thus, the sensitivity is derived from the overall fit between the model and observations, and in particular, the response to volcanic eruptions which is a unique and relatively known radiative forcing. Grossly speaking, the correlation between the solar activity and climate therefore leads to the constraints on the ISE.

Because the secular changes in different solar activity related parameters is typically between 50% and 100% of the solar cycle related variations (e.g., in the sunspot number, the amplitude of the CRF variations, or the changes in the AA index), we should expect the long term solar-activity related change in the radiative forcing to be between half and the full variation over the solar cycle. This is consistent with our observations and the fact that Shaviv (2008) found that the solar-cycle related variations in the radiative forcing is 1.3 ± 0.4 W m⁻².

As to a future solar contribution towards climate change, it is impossible to predict the solar activity beyond one solar cycle (Yeates et al., 2008). Nevertheless, because over the last several decades the sun has been as active as it can be, we can expect a similar or a decreased activity. This can therefore result in a cooling of up to a few 0.1 °C if the activity will indeed decline.

5.1. Caveats

Just like any analysis, it is important to understand the limitations of our results and conclusions.
First, the model presented here is limited since it is not designed to represent in detail the relevant physics but rather, through parametrization, to demonstrate the physical significance of the feedback processes involved. In essence, it allows us to form a relatively simple qualitative and quantitative picture of the phenomena associated with climate change. As a consequence, it is not able to capture complex behavior which may have systematic effects on the system as a whole, and we cannot rule out corrections to our fitted parameters, such as the model gain.

For example, there is no geographic resolution other than the land/ocean identification. Complex phenomena such as that of ocean currents, tropical upwelling and polar downflows are altogether missing since they cannot be described through 1D diffusion, but they can affect the nature of the heat flow into the oceanic heat reservoir.

Another limitation of our analysis has to do with the fact that over the 20th century, several of the forcings have a monotonic or almost monotonic behavior. In our analysis, this produces hard to resolve degeneracies. The primary example is well known in climate studies. While the GHGs have monotonically increased over the 20th century, the indirect aerosols have monotonically decreased (becoming more negative). As a consequence, any uncertainty in the absolute normalization of the indirect aerosol effect can be counteracted with a different overall climate sensitivity. The almost monotonic increase in the average solar activity would have made it fall into the same trap, however, because solar activity includes solar cycle oscillations, its cooling. J. Geophys. Res. 103, 5929–5942, 1998.

We should also point out that the analysis implicitly assumes that the radiative budget response to changes in the temperature are the same on any time scale between a year and a century. This implies that the analysis neglects the effects of feedbacks which operate on longer scales, such as the ice-albedo feedback.

Last, we should emphasize that the present analysis cannot be used to point to any particular mechanism for an indirect solar effect, but rather that such an effect should be present. This is because there is a large degeneracy between the TSI and the ISE. Namely, we cannot distinguish between hypersensitivity to the TSI (or some component of it), and an indirect effect, such as sensitivity to cosmic ray flux variations.

6. Summary

To summarize, the following are the main conclusions of our work.

- The sun has a much more significant role on Earth’s climate than is commonly thought. Its estimated 20th century forcing on the climate is 0.8 ± 0.4 W m⁻².
- Earth’s climate sensitivity is very close to that of a “black body”. Thus, the various feedbacks cancel each other out.

- Nominally, we can account for 40% of the 20th century global warming by the sun alone while 60% should be attributed to anthropogenic activity.

Furthermore, we show in the present work that a simple energy balance model can shed significant light on the understanding and quantification of the climate system, and in particular, that such models can improve our understanding of the solar–climate link.

References


