On the estimation of systematic error in regression-based predictions of climate sensitivity

Benjamin M. Sanderson

© U.S. Government 2013

Abstract An extension of a regression-based methodology for constraining climate forecasts using a multi-thousand member ensemble of perturbed climate models is presented, using the multi-model CMIP-3 ensemble to estimate the systematic model uncertainty in the prediction, with the caveat that systematic biases common to all models are not accounted for. It is shown that previous methodologies for estimating the systematic uncertainty in predictions of climate sensitivity are dependent on arbitrary choices relating to ensemble sampling strategy. Using a constrained regression approach, a multivariate predictor may be derived based upon the mean climatic state of each ensemble member, but components of this predictor are excluded if they cannot be validated within the CMIP-3 ensemble. It is found that the application of the CMIP-3 constraint serves to decrease the upper bound of likelihood for climate sensitivity when compared with previous studies, with 10th and 90th percentiles of probability at 1.5 K and 4.3 K respectively.

1 Introduction

Model-based projections of future climate change are subject to numerous uncertainties, which can be loosely categorized into four groups; the model boundary conditions, the initial variability of the system, model parametrization schemes and so-called 'systematic' uncertainties (Tebaldi and Knutti 2007). In each of the former three cases, experiments have been designed to explore how these unknowns are translated into probabilistic statements concerning future projections. Observational uncertainties are implicitly linked to model uncertainty, as any errors in observations will be manifested in less than optimal tuning of model parameters.

Unknown boundary conditions may be addressed by exploring simulations with different climate forcing scenarios, such as the Special Report on Emissions Scenarios (SRES, Nakicenovic et al. (2000)) which include numerous scenarios for future emissions and concentrations of atmospheric forcing agents. Initial conditions clearly

B. M. Sanderson (✉)
National Center for Atmospheric Research, Boulder, CO, USA
e-mail: bsander@ucar.edu

play a role in short term and seasonal forecasts of climate (Collins and Allen 2002) but whether initial information is relevant on a decadal timescale or longer is still a subject of debate (Hurrell et al. 2009).

On a global scale, the response of the Earth system to changes in boundary conditions can be summarized by a small number of parameters. The 'climate sensitivity' is the equilibrium response of global mean surface temperature to a doubling of atmospheric carbon dioxide concentrations and is inversely proportional to the net global temperature feedbacks in the system. Many studies (Gregory et al. (2002), Annan and Hargreaves (2006), Knutti et al. (2006) to name a few) have sought to constrain the value of the climate sensitivity because many equilibrium changes in climate on both a global and a regional scale are expected to scale with this number Knutti and Hegerl (2008). Equilibrium response is clearly an idealized case, as any real-world climate change will consist of a transient response to slowly changing boundary conditions, which is a function of both the net feedbacks and of the inertia of the system (governed primarily by the rate of ocean heat uptake). However, in this study we will focus on the uncertainties associated with climate sensitivity, which can be easily calculated in a wide variety of atmospheric climate models without the need for a computationally expensive fully dynamical ocean.

Model parameter uncertainty and its impact on climate projections has been studied by various methods; parameter perturbation experiments have been conducted using the Met Office Hadley Centre Model (Murphy et al. (2004), Stainforth et al. (2005)), the Model for Interdisciplinary Research On Climate (Annan et al. (2005)) and the Community Atmosphere Model (Jackson et al. (2008), Sanderson (2011)). There have been various attempts to use these Perturbed Physics Ensembles (PPEs) together with observations to somehow constrain the climate sensitivity or transient response of the real-world system. Murphy et al. (2004) produced a distribution of climate sensitivity, interpolated over the model parameter space and weighted by a measure of model likelihood, but Frame et al. (2005) argued that a Probability Density Function arising from such an approach is inherently dependent on the prior distribution of models in the parameter space. Some more recent Bayesian studies have duly sampled the sensitivity of posterior PDFs to the prior (Sexton et al. 2011), while Sexton and Murphy (2012) argue that if sufficiently strong observational constraint, the result's sensitivity to the prior distribution can be reduced.

Piani et al. (2005) also presented a methodology for relating observable quantities to an unknown response metric, using linear regression to predict climate sensitivity from the amplitude of various independent modes of variability in the model. The modes were derived by rotating a set of Empirical Orthogonal Functions (EOFs) derived from a long control simulation such that they were independent within the PPE. The coefficients of those EOFs in the ensemble can then be related to the climate sensitivity through ordinary least squares regression in the PPE. In order to make an estimate of a PDF for climate sensitivity, observational fields can be projected onto the EOFs and the established regression coefficients can be used to estimate a best guess of climate sensitivity.

Uncertainty due to natural variability in the observations was estimated by considering the variability of the modes projected onto a second, independent long control simulation. The authors did attempt to quantify the systematic uncertainty in their prediction by expressing this control simulation as anomaly about the ensemble mean state. This artificially inflates the uncertainty term by an amount proportional to the projection of the difference between the ensemble mean and the control simulation mean onto the EOFs.

In other words, if the ensemble mean shows a radically different base climate to the long control simulation, there will be a large inflation of error due to assumed systematic uncertainty. The problem is that the ensemble mean state is dependent upon the arbitrary sampling strategy of the PPE and it is not clear that this results in a robust estimate of the systematic uncertainty in the methodology. The estimate remained untested because the predictors were not applied to any other independent climate models.

Knutti et al. (2006) use a non-linear regression to relate regional seasonal cycles of temperature to Climate Sensitivity, though in principal, the methodology may be applied to any measure of model response. The study was conducted using a PPE, of which a subset was used to train a neural network which acts as a transfer function between an observable quantity and an unknown response, while the prediction error was estimated by calculating the error in the prediction of Climate Sensitivity for the remaining models in the PPE. This relationship was applied to observations of the climate system to produce a 'best-guess' estimate of the true value of the response metric. The width of the resulting PDF is a function of natural variability in the observables, plus uncertainty in the observed values together with the ensemble-derived prediction error.

The final source of uncertainty in regression-based predictions of climate response is the systematic or irreducible component which is due to the difference in underlying formulation between different models and the true climate system. Knutti et al. (2006) did show some success in using their predictors to predict the unseen climate sensitivities of other models in the CMIP-3 archive, but did not include this information directly into their PDFs (they did, however, show the sensitivity of their PDFs to an increase in the assumed systematic error).

Rougier (2007) and Sexton et al. (2011) address the problem differently to the transfer function approach, by explicitly evaluating model likelihood as a function of the parameter space of the model. The systematic error is represented by a 'discrepancy' term, which cannot be reduced by model parameter adjustment, although they use different methodologies to define the term itself. Rougier (2007) advocates the use of additional 'hyperparameters' to define the model discrepancy on a (pre-defined) regional scale, where a sufficient number of degrees of freedom are created to allow the observations and ensemble to coexist in the parameter space. Sexton et al. (2011) reject this approach owing to concerns of double-counting observational information for both the likelihood estimate and the discrepancy term. Instead, they derive a discrepancy term by treating different members of the CMIP-3 ensemble as truth, and calculating PDFs for the future response using the PPE. The resulting errors in that prediction inform on what additional discrepancy term must be introduced to account for intra-ensemble systematic errors.

However, these approaches to model discrepancy are not directly applicable to a transfer function approach like Piani et al. (2005) or Knutti et al. (2006). In a PPE, the models are by definition not tuned to their optimal state and some models are very poor representations of the present day climate (Sanderson et al. 2008). Potential systematic errors in Piani et al. (2005) and Knutti et al. (2006) could arise if relationships between observable quantities and unknown response metrics were present within the PPE, but could not be generalized to other climate models or to reality.

One approach to this problem would be simply to evaluate the performance of the predictors when applied to a MME (Piani et al. (2007), Sanderson et al. (2008)). However, given that different models may share components and many models have similar resolutions and inherent assumptions, this may lead to overconfidence in the accuracy of the predictors when applied to the real world. In addition, if the

systematic error proves to be very large when estimated using the MME, a methodology for reducing this error is desirable.

In this study, we propose a regression-based constraint on Climate Sensitivity which excludes those correlations which cannot be validated in a wide range of GCMs. Inherent in this approach is a methodology for estimating the systematic error in the resulting PDF, using the assumption that the MME provides a sufficient sample of different systematic model behavior, and that the MME as a whole is not significantly biased.

2 Methodology

2.1 Ensemble principal component analysis

Our goal is to find multivariate predictors of climate sensitivity within a perturbed physics ensemble, with predictions arising from the variables listed in Table 1 in the additional material section 7.1. The variables are chosen to provide an overall evaluation of model climate, although the list is clearly not exhaustive. The variables were chosen both for their availability in all the ensembles considered, together with the availability of observational values. The overall conclusions of the study were not found to depend strongly upon the variables chosen.

Multiple gridded seasonal variables for precipitation, surface temperature, atmospheric humidities and radiative fluxes are normalized and concatenated to form a long state vector for each model's climatology. This process which is repeated for each model in the PPE. The state vectors in the PPE are then subjected to a Principal Component Analysis, such that the base climatology in each model in the ensemble may be described by a short, truncated vector with independent components. This process is explained in detail in the additional material section 7.1.

Table 1 Observable quantities used to describe the climate state. The middle column shows the domain over which the fields are taken, lat/long represents latitude and longitude, linearly interpolated to a 73 by 96 (2.5° by 3.75°) grid (if necessary). lat/pres represents zonal mean fields on pressure levels, interpolated to a 73 by 17 grid consistent with HadAM3 output. The right-hand column indicates the source of observation or reanalysis data used to estimate real world values. In all cases, values from DJF and JJA seasons are concatenated to form a single vector. NCEP data is described in Kalnay et al. (1996), CERES-2 data in Wielicki et al. (1996) and AIRS data in Susskind et al. (2003)

Diagnostic	Domain	Obs./reanalysis
Surface temperature (K)	lat/long/seas	NCEP
Total precipitation rate (ms^{-1})	lat/long/seas	NCEP
Clear sky LW TOA flux	lat/long/seas	CERES-2
Clear sky SW TOA flux	lat/long/seas	CERES-2
All sky LW TOA flux	lat/long/seas	CERES-2
All sky SW TOA flux	lat/long/seas	CERES-2
Air temperature (K)	lat/pres/seas	AIRS
Relative humidity (%)	lat/pres/seas	AIRS
Zonal wind (ms^{-1})	lat/pres/seas	NCEP
Meridional wind (ms^{-1})	lat/pres/seas	NCEP

3 Unconstrained regression

3.1 Methodology

We begin by demonstrating a methodology similar to Piani et al. (2005), which we demonstrate produces similar results. Piani et al. (2005) used predictors formed from eigenvectors of natural variability, orthogonally rotated such as to be independent within the ensemble. In this work, we omit the rotation step—simply using eigenvectors derived by applying a PCA to the ensemble itself. This has practically little effect on results but allows ensemble variability to be expressed exactly, rather than forcing the eigenvectors to be expressed as a linear combination of an incomplete basis set. Furthermore, eigenvectors derived directly from the ensemble are more interpretable as relating to physical parametrizations within the ensemble (Sanderson et al. (2008), Sexton et al. (2011)), rather than forming an arbitrary set of basis vectors.

The mathematical details of the regression process are detailed in the additional material section 7.2, but the technique may be summarized as follows: The sensitivity of each model in the PPE is related to the principal components derived in Section 2.1 by means of a least squares regression. The observations are projected onto the set of EOFs derived from the PPE, and together with the regression coefficients, yield a best-guess value for climate sensitivity.

The methodology then considers three sources of error for this projection. The first, due to natural variability is derived by taking a long control simulation, dividing it into sections, subtracting the mean, projecting onto the EOFs and using the regression coefficients to calculate a distribution of values. The variance of this distribution gives the uncertainty due to natural variability.

The second source of error in the unconstrained regression arises from the systematic bias estimate and is described in detail in the additional material section 7.2.1. For this, we follow Piani et al. (2005) by representing the long control as an anomaly about the climate*prediction*.net ensemble mean. This has the effect of inflating the derived variance, but is a largely arbitrary decision which is itself dependent on the distribution of models in the climate*prediction*.net ensemble. The final source of error is due to intra-ensemble prediction, using the climate*prediction*.net ensemble itself as a transfer function. This calculation is described in additional material section 7.2.2.

3.2 Application to a PPE

In Fig. 1a, we demonstrate the above methodology using the climate*prediction*.net ensemble of perturbed climate models (Stainforth et al. 2005). As in Piani et al. (2005), we take a subset of models from the ensemble (in this case $n_{ens}=1696$, after removal of models with a drifting control climate). We apply the methodology detailed in the previous section to the ensemble, estimating the systematic bias of the model by expressing the 500 year control simulation variability about the climate*prediction*.net ensemble mean. Although in keeping with the methodology of Piani et al. (2005), this bias clearly does not represent any difference between models and observations, rather as a representative inter-model bias.

Despite a slightly different methodology (the use of different observable fields and direct PCA of the ensemble, rather than using rotated natural EOFs), we obtain results very similar to those found by Piani et al. (2005), with a most likely climate sensitivity of 2.8 K and 95th percentiles at 1.3 K and 6.0 K. For comparison, Piani et al. (2005) found 5th and 95th percentiles of 2.2 and 6.8 K.
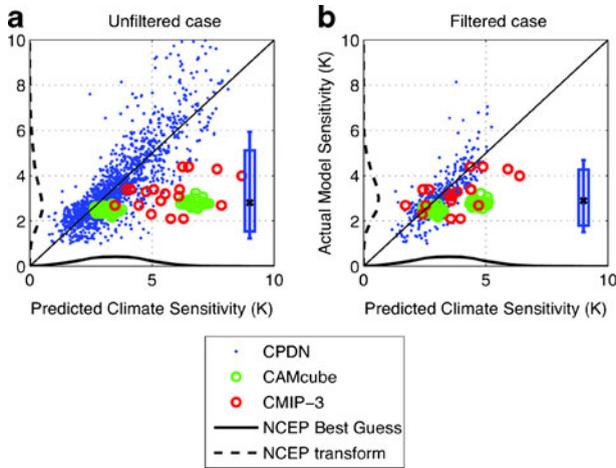
Fig. 1 a A plot in the style of Piani et al. (2005) showing true model climate sensitivity expressed as a function of the value predicted from independent modes of control state variability within the climate*prediction*.net ensemble. Each *blue point* represents a single model in the climate*prediction*.net ensemble. *Red points* show the same predictors applied to CMIP-3 control simulations, where the 'true' sensitivities of each model are the slab model climate sensitivities of each respective models. *Green points* show the predicted and true climate sensitivities of slab models in the NCAR CAMcube ensemble (Sanderson 2011). The *solid curve* on the horizontal axis is the distribution for likely climate sensitivity using the observations detailed in 1, and adding uncertainty due to natural variability and systematic error using the methodology in Section 3 (this curve does not include any estimate of error due to the prediction itself). The dashed curve on the vertical axis is the final PDF using the climate*prediction*.net ensemble as a transfer function to estimate the intra-ensemble prediction error. The box and whisker plot shows the median, along with the 90th and 95th percentiles of this PDF. **b** As for (**a**) except for a pre-filtering of climate*prediction*.net models to exclude those with greater than 8 Wm^{-2} global mean top of atmosphere energy imbalance

One way we can verify the assumptions which go into the estimation of systematic error with this method is to apply the predictors to different climate models where both a pre-industrial simulation and a double CO_2 sensitivity experiment are available. We evaluate the predictor performance on two other ensembles. Firstly, the predictors may be applied to pre-industrial control simulations from the Coupled Model Inter-comparison Project (CMIP-3), and the predicted sensitivities can be compared to the known climate sensitivities of each model. Figure 1a shows that the climate*prediction*.net derived predictors consistently overestimate the sensitivities of each model in the CMIP-3 ensemble. The mean predicted CMIP-3 sensitivity is 5.4 K, compared to the true CMIP-3 mean sensitivity of 3.2 K. Similarly, if we apply the climate*prediction*.net derived predictors to control simulations from the NCAR CAMcube ensemble (Sanderson 2011), a bimodal distribution of sensitivity is predicted; 54 CAM models in this ensemble have convective mass transport parametrization switched on, and these models have sensitivities which are accurately predicted by the climate*prediction*.net derived predictors (a predicted mean sensitivity of 2.9 K and a true mean sensitivity of 2.6 K). However, in the remaining 27 CAMcube models, the convective mass transport is switched off and the predicted sensitivities are strongly overestimated (a predicted mean of 6.3 K as compared to the true mean sensitivity of 2.8 K). Both these verifications with other ensembles suggest that the estimated systematic error using the above technique may be an underestimate.

A possible criticism of calculations of the type shown in Fig. 1a is that a large number of models in the PPE can often be dismissed as unphysical by evaluation of their base climatology.

Although model weighting according to some pre-defined likelihood function is difficult to achieve and often dependent on arbitrary choices, we can illustrate that pre-filtering an ensemble for model quality can potentially change the result of a regression-based prediction of climate sensitivity. In Fig. 1b, we perform the regression using only a subset of the climate*prediction*.net ensemble where models have a Top of Atmosphere energy imbalance of less than 8 Wm^{-2} (leaving 411 viable models). Although this is clearly not a comprehensive evaluation of model climate, it serves to generally eliminate those models which have instantly dismissible base state climatologies. We find that even with this simple restriction, that the results of the regression-based predictor exhibit fewer biases when applied to the CMIP-3 ensemble, with a mean predicted sensitivity of 3.8 K, much closer to the actual value of 3.2 K. This suggests that unrealistic models in the PPE can potentially dominate the terms in a regression prediction of climate sensitivity.

We can also explore the sensitivity of the technique to the choice of systematic model bias. In Piani et al. (2005), the systematic bias was estimated by adding a bias to the noise estimate of natural variability about the observed mean, where the bias was taken to be the difference between the climate*prediction*.net ensemble mean control state and the mean of a HadCM3 500 year control simulation. We can test the sensitivity of the technique to the choice of systematic bias by taking the bias to be the difference between the climate*prediction*.net mean and various models in the IPCC ensemble. The results of this sensitivity study are shown in Fig. 2.

The results of this sensitivity study show that the upper bound of likelihood for climate sensitivity is especially sensitive to the choice of systematic bias estimate. The upper bound (95th percentile) ranges from 4.4 K (using the Canadian Climate Centre Model 3.1 at T63 resolution) to 8.0 K (using Miroc 3.2 'medium' resolution model). So, clearly the upper bound of the PDF is highly sensitive to the choice of model used to evaluate the systematic bias component.
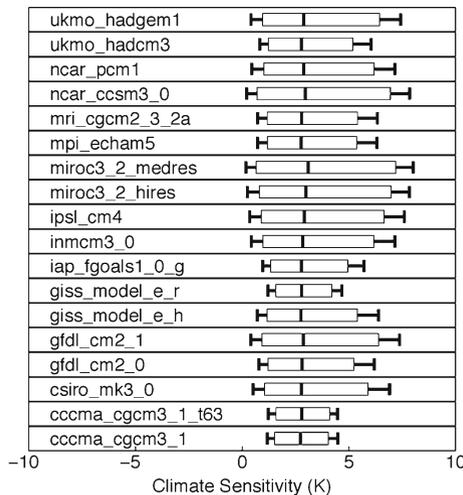

Fig. 2 Repeating the methodology to produce the PDF in Fig. 1, but using different models in the CMIP-3 ensemble to estimate systematic model bias. To produce each PDF, the bias added to the noise estimate as described in the additional material, section 7.2.1 is taken to be the difference between the climate*prediction*.net model mean and each of 18 pre-industrial simulations from the IPCC CMIP-3 ensemble. Each box and whisker plot shows the 5th, 10th, 50th, 90th and 95th percentiles of the resulting PDF using each one of the CMIP-3 models to estimate systematic bias from the climate*prediction*.net mean

Although we have shown that the results of a regression-based prediction of climate sensitivity can potentially be made less biased by considering a subset of models close to observed climatology, this is not an ideal solution for two reasons. Firstly, it has been shown by Rougier (2007) and others that rigorous inter-model weighting within a PPE is not a trivial task, and is potentially sensitive to the choice of metrics considered (Gleckler et al. 2008). Secondly, the methodology as presented above using the constrained subset of climate*prediction*.net still requires a somewhat arbitrary computation of a systematic bias term. Therefore, in the following section we propose a modified regression-based predictor with two goals of producing a somewhat quantitative systematic error term and decreasing inter-model prediction bias by making the process less sensitive to potentially unrealistic models in the PPE.

We can to some degree test our methodology with an independent PPE, CAMcube, which is derived using the Community Atmosphere Model. If model sensitivity within this validation ensemble can be predicted within the derived error margins, we can have more confidence in our methodology correctly estimating the systematic error in the prediction. We choose to use CAMcube as validation only because the range of climate sensitivity simulated in this ensemble is significantly smaller than that of climate*prediction*.net, and thus cannot be used to significantly constrain relationships between sensitivity and base climate state.

4 Constrained regression

Figure 1 shows that the climate*prediction*.net derived predictors of sensitivity are strongly biased towards high predicted values when applied to the CMIP-3 dataset, suggesting that the observation-based predicted sensitivity value may also be biased towards high values. In this section, we seek to use the available simulations from the CMIP-3 ensemble to constrain the regression process itself.

The methodology is presented in detail in the additional material section 8, but can be summarized as follows: we seek to eliminate those correlations which cannot be verified in the CMIP-3 ensemble. To do this, we invoke a constrained regression algorithm which finds the least squares relationship between the base vectors defined in the previous section and climate sensitivity *subject to the constraint that* the sensitivities in the CMIP-3 ensemble are predicted correctly within an assumed error E_{max}. In choosing the constraint E_{max}, we are effectively establishing an upper limit for the allowed error in the prediction of sensitivity for any model in the CMIP-3 archive and making an *a priori* assumption about the inherent systematic error in applying the predictor to another climate model. The predictors are constrained to predict valid sensitivities for the CMIP-3 models, within the permitted systematic error margins. Intra ensemble prediction error is then calculated as in the additional material Section 8.0.3, by using the climate*prediction*.net ensemble as a transfer function to produce the final PDF for climate sensitivity.

Systematic error in the constrained calculation is taken to imply any error which might arise from the assumption that relationships which are valid in the climate*prediction*.net ensemble are also valid in other models, or in the real world. We estimate this term by measuring the mean error in the prediction of sensitivities for the CMIP-3 archive. If this is greater than the combined uncertainty due to natural variability and intra-ensemble prediction error, the systematic error term is deemed to be non-zero and the variance in the predicted result is inflated to account for the CMIP-3 prediction error. The technique is described in detail the additional material section 8.0.3.

The three plots in Fig. 3 show three different *a priori* assumptions for E_{max}, with the weakest constraint requiring CMIP-3 model Sensitivities to be predicted within 0.5 K, 1.5 K and 3.0 K of the true value. The final PDFs for climate sensitivity exhibit their 95th percentiles at 6.0 K, 4.5 K and 5.1 K respectively.

Figure 4a shows the different errors as a continuous function of E_{max}. When the allowed error E_{max} is less than 0.4 K—there is no solution to the constrained regression algorithm, implying that there are not sufficient degrees of freedom in the regression equation to predict the CMIP-3 sensitivities.

When E_{max}=0.4, the regression is entirely based upon correlations in the CMIP-3 models, the algorithm cannot optimize the regression coefficients to better fit the climate*prediction*.net ensemble. In this case, the intra ensemble prediction error is at its maximum and the systematic error term E_{sys} is zero.

As the allowed CMIP-3 error E_{max} is increased, the accuracy of predictions within the climate*prediction*.net ensemble is improved and E_{pred} becomes smaller, but the skill in predicting models in the CMIP-3 archive is reduced until the point that the systematic error term E_{sys} becomes non-zero.

When E_{max} is greater than 7 K, the problem is effectively unconstrained and the result is identical to the ordinary least square regression approach in Section 7.2.2. Figure 4b shows the PDFs as a continuous function of E_{max}, and it can be shown that the width of the PDF exhibits a
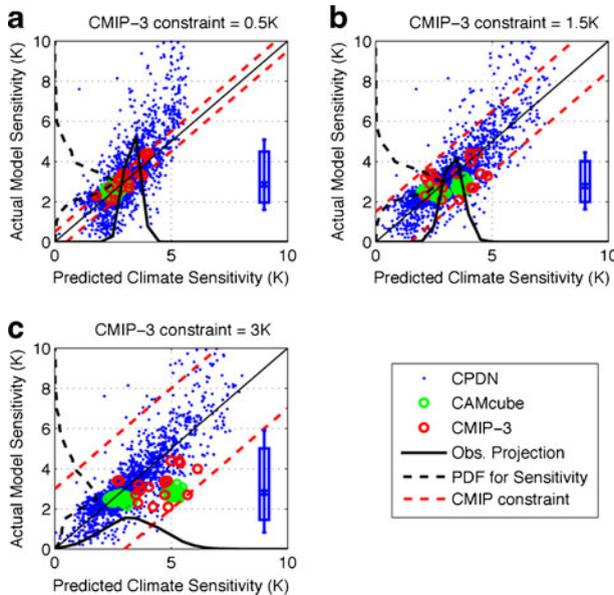

Fig. 3 Sensitivity distributions assuming three different values of E_{max}, which defines the maximum allowed error in the prediction of model sensitivity in the CMIP-3 archive. *Blue dots* represent individual models in the climate*prediction*.net archive, with predicted sensitivity on the horizontal axis and actual climate sensitivity on the vertical axis. *Green dots* show models in the CAMcube ensemble, and *red dots* show models in the CMIP-3 archive. The *dashed red line* shows the CMIP-3 constraint. The curves on the bottom axes show the distribution of observational predictions, centered on the most-likely value with a distribution width due to both simulated natural variability and a pre-defined systematic uncertainty distribution of standard deviation E_{max}. The curve on the vertical axis used the climate*prediction*.net ensemble as a transfer function to estimate the final PDF for climate sensitivity. The *vertical boxes* and whiskers represent the 5th, 10th, 90th and 95th percentiles of the final PDF

minimum at a value $E_{max} = 1.5\ K$. The PDF is wider at larger values E_{max} because the CMIP-3 prediction error is larger, and at small values of E_{max} because the climate*prediction*.net prediction error is increased. The resulting PDF for the optimal E_{max} value of 1.5 K is shown in Fig. 3b.

The process of constraining the regression restricts correlations to those which can be validated within the CMIP-3 ensemble, and we can use the independent CAMcube ensemble to test these climate*prediction*.net derived constrained predictors in a different environment (the CAMcube ensemble is not itself used to constrain the regression). In each case, the constrained regressions are considerably more accurate in predicting sensitivities in the CAMcube ensemble than the Ordinary Least Squares regression approach. The regressions with the smallest CMIP-3 constraint produce the smallest errors in the prediction of CAMcube sensitivity (Fig. 4).
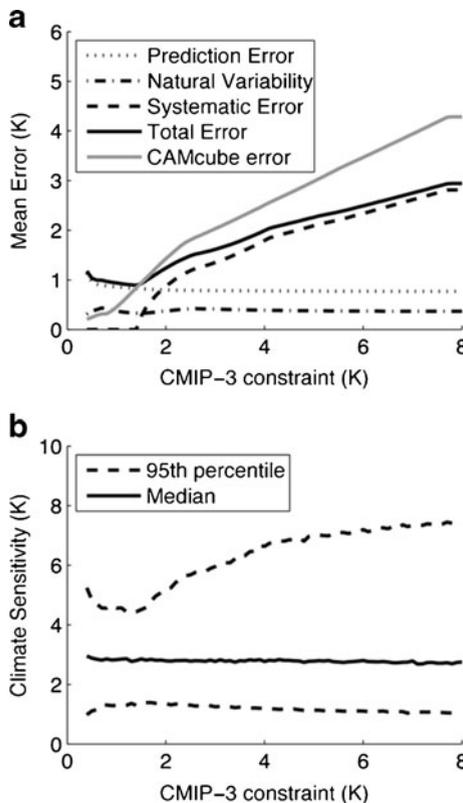

Fig. 4 **a** The standard deviation of the three forms of error leading to uncertainty in the PDF for climate sensitivity, as a function of the *a priori* CMIP-3 constraint. E_{max}. The 'prediction error' is the standard deviation of the residuals in the prediction of models within the climate*prediction*.net ensemble. The 'natural variability' is the standard deviation of the predictor applied to 64×15 year segments from a 500 year control simulation. The 'systematic error' is the additional error which must be combined with the other terms required to describe the standard deviation of CMIP-3 residuals. The 'Total error' is the root sum square combination of these separate sources of error. Finally, 'CAMcube error' is the standard deviation of the residuals in the prediction of models within the CAMcube ensemble. **b** 90 and 95th percentiles of the final distribution for climate sensitivity shown as a function of of the constraint on CMIP-3 model sensitivity. The PDFs are calculated as for the curves on the vertical axis in Fig. 3

All of the metologies presented in this study require an decision on the appropriate truncation length for the EOF basis set. In Piani et al. (2005), this was determined by an f-test, but we show in Fig. 5 in the additional material that this decision becomes less critical in the constrained regression case. Whereas the unconstrained regression shows a large dependency of the resulting PDF for climate sensitivity on the truncation length, the constrained case shows little dependency for truncation lengths above 10. Interestingly, the unconstrained result using the pre-filtered ensemble using only models in approximate energy balance is also not highly dependent on truncation length, with the results of the latter two calculations largely in agreement for truncation lengths of 10 or more.

The distribution for sensitivity with an optimal E_{max} value of 1.5 K and a truncation length of 20 (or greater) indicates a most likely value of 2.9 K with 5th and 95th percentiles at 1.7 K and 4.6 K. There are some caveats to this result. Firstly the assumption of a linear transfer function between the climate state variables and sensitivity could omit important potentially nonlinear relationships. Secondly, we assume that the errors due to natural variability in each predictor are normally distributed, which may be an oversimplification.

5 Discussion

The future response of the climate to changing boundary conditions caused by an increase in anthropogenic greenhouse gas concentrations is highly dependent upon the strength of the net feedbacks in the climate system. Perturbed physics ensembles (PPEs) exhibit a wide range of climate sensitivity to increasing greenhouse gases, and are thus an invaluable tool for studying and constraining likely real-world response. Past approaches have relied on finding correlations between observable quantities and climate sensitivity within a PPE. However, we have shown that such approaches may not be robust when PPE-derived predictors of climate sensitivity are applied to entirely separate models, such as those found in a Multi Model Ensemble
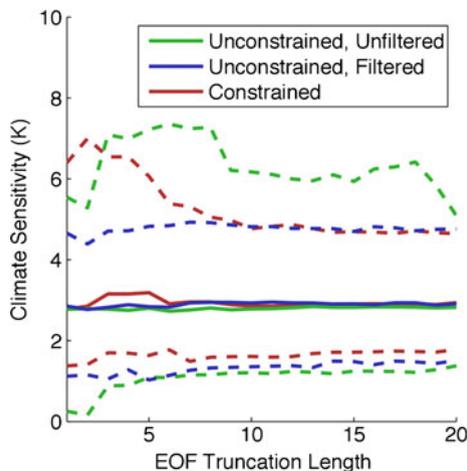

Fig. 5 Median (*solid lines*) and 5,95th percentiles (*dashed lines*) of the final distribution for climate sensitivity shown as a function of the truncation length for the climate state EOFs used as predictors of sensitivity. Plots in *green*, *blue* and *red* show conventional OLS regression, pre-filtered regression as in Fig. 1b and optimally constrained CMIP-3 regression as in Fig. 3b

(MME) such as CMIP-3. We show that prediction biases can be significantly reduced by performing regressions on a subset of plausible models within the PPE, but this requires at least two arbitrary decisions to be made: the choice of metrics to be used for model weighting and the choice of how to represent an unknown systematic error term.

We have developed an alternative approach in which we estimate uncertainty in a PPE regression-based prediction of climate sensitivity using a MME which we assume contains models which are indistinguishable from reality (i.e. the models are drawn from a distribution of which the real world is a potential member). Our methodology differs from that used in previous work primarily in our treatment of the systematic uncertainty which arises when correlations within a particular climate model structure are used to predict attributes of a completely different model or the real world.

Past approaches have considered the uncertainty in the observed state as a minimum estimate for the systematic uncertainty term, and then examined the sensitivity of the PDF to an arbitrary inflation of this term (Knutti et al. 2006). Other approaches have inflated the natural variability term to represent the systematic uncertainty—but in a fashion which is arbitrarily dependent on the fashion in which the ensemble parameter space has been sampled (Piani et al. 2005). We have shown that the methodology in Piani et al. (2005) is highly influenced by the presence of unphysical models within the PPE, and by performing a simple pre-filtering to exclude those models significantly out of energy balance, we can significantly reduce the uncertainties in the projection of a PDF for climate sensitivity. Such an approach in undesirable, however, because an arbitrary decision is required to define suitable metrics for determining whether a model is acceptable or not.

We instead propose an approach which is not directly dependent on the choice of sampled parameters or included models, instead making the assumption that the CMIP-3 multi-model ensemble represents a range of model environments. The methodology requires an *a priori* decision on the degree to which the CMIP-3 models should be constrained, but an optimal constraint can be determined by sampling values to minimize the uncertainty in the final PDF. Once an optimal value is determined, the regression of model base-state predictors onto model response is performed over the PPE. An additional systematic error term is introduced to inflate the error due to natural variability to account for the error in predicting the CMIP-3 models. A final PDF for climate sensitivity is obtained by using the climate*prediction*.net ensemble as a transfer function to describe any intra-PPE prediction error.

If we consider the extreme cases: as the systematic constraint tends to zero in the case that there are fewer independent predictors than CMIP-3 models, there is no solution to the problem. As the constraint is increased to the minimum value with a solution, the regression is made over the CMIP-3 ensemble only. However, this over-constraint becomes self-evident when calculating the internal prediction error in the PPE, the predictors show little skill on that dataset and the internal prediction error term becomes very large.

Considering the other extreme case, where the CMIP-3 constraint becomes infinite, the regression over the PPE is unconstrained and the systematic error term is estimated from the spread of CMIP-3 residuals. However, this spread may be very large because the predictors may be dominated by correlations within the PPE which arise from highly perturbed models and processes which may only been significant within the PPE itself. When these predictors are applied to the CMIP-3 model to estimate the systematic error term, large prediction errors are observed which inflate the systematic uncertainty component of the total error.

Clearly, both of these extreme cases results in large uncertainties in the final PDF for climate sensitivity. However, we optimize the prediction by minimizing the width of the PDF as a function of the systematic constraint. This 'optimal' constraint ensures that regression predictors derived from the PPE are not used if they are not validated within

the CMIP-3 ensemble. This approach assumes, of course that the CMIP-3 models themselves are to some degree physical. This is largely justified by the fact that PPEs contain many models which easily be disregarded on physical grounds (Sanderson et al. 2008), whilst it is often difficult to produce consistent rankings for models in the multi-model case (Gleckler et al. 2008). The use of the multi-model ensemble as a testbed for PPE derived predictors is thus imperfect, but is likely the best strategy with current available simulations.

One legitimate concern with this approach is whether systematic difference between models in the CMIP-3 ensemble is comparable to that between any given model and the real-world. Certainly, the models have many approximations in common, such as limits in resolution which may cause a failure to resolve atmospheric features such as blocking (Palmer et al. 2008) and omitted components of the earth system such as an interactive carbon-nitrogen cycle which might significantly affect the real-world response to greenhouse gas forcing. Our analysis therefore produces a lower bound on uncertainty, on the assumption that the CMIP-3 models are a sample of possible 'worlds'.

What differs from previous PPE derived estimates of sensitivity is the exclusion of correlations which cannot be validated in other GCMs. The benefits of this are two-fold: firstly, if models are present in the PPE which are highly perturbed and give very poor representations of the climate, they are no longer allowed to dominate the prediction. Secondly, the dependency of the prediction on EOF truncation length is eliminated because predictions based upon the higher, noisy predictors are suppressed by the constraint. The methodology established here may be extended to provide multivariate, or regional predictions of unknown climate parameters based upon data from both a PPE and a MME used to validate and constrain the prediction.

References

Annan J, Hargreaves JC (2006) Using multiple observationally-based constraints to estimate climate sensitivity. Geophys Res Lett 33(6):L06,704

Annan JD, Hargreaves JC, Ohgaito R, Abe-Ouchi A, Emori S (2005) Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations. SOLA 1:181–184

Collins M, Allen M (2002) Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. J Climate 15(21):3104–3109

Frame DJ, Booth BBB, Kettleborough JA, Stainforth DA, Gregory JM, Collins M, Allen MR (2005) Constraining climate forecasts: the role of prior assumptions. Geophys Res Lett 32:L09,702

Gleckler P, Taylor K, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06,104

Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA (2002) An observationally based estimate of the climate sensitivity. J Climate 15:3117–3121

Hurrell J, Delworth T, Danabasoglu G, Drange H, Griffies S, Holbrook N, Kirtman B, Keenlyside N, Latif M, Marotzke J, et al. (2009) Decadal climate prediction: opportunities and challenges. Ocean Obs 9

Jackson C, Huerta G, Sen M, Deng Y, Bowman K (2008) Error reduction and convergence in climate prediction. J Climate p 35

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J et al (1996) The ncep/ncar 40-year reanalysis project. Bull Am Meteorol Soc 77(3):437–471

Knutti R, Hegerl G (2008) The equilibrium sensitivity of the earth's temperature to radiation changes. Nat Geosci 1(11):735–743

Knutti R, Meehl GA, Allen MR, Stainforth DA (2006) Constraining climate sensitivity from the seasonal cycle in surface temperature. J Climate 19(17):4224–4233

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772

Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S, Gregory K, Grubler A, Jung T, Kram T et al (2000) Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change. PNNL-SA-39650. Cambridge University Press, New York

Palmer T, Doblas-Reyes F, Weisheimer A, Rodwell M (2008) Toward seamless prediction: calibration of climate change projections using seasonal forecasts. Bull Am Meteorol Soc 89(4):459–470

Piani C, Frame DJ, Stainforth DA, Allen MR (2005) Constraints on climate change from a multi-thousand member ensemble of simulations. Geophys Res Lett 32(23)

Piani C, Sanderson B, Giorgi F, Frame D, Christensen C, Allen M (2007) Regional probabilistic climate forecasts from a multithousand, multimodel ensemble of simulations. J Geophys Res-Atmos 112(D24):D24,108

Rougier J (2007) Probabilistic inference for future climate using an ensemble of climate model evaluations. Clim Chang 81(3):247–264

Sanderson BM (2011) A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. J Clim 24:1362–1377

Sanderson B, Piani C, Ingram W, Stone D, Allen M (2008) Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. Clim Dyn 30(2):175–190

Sexton D, Murphy J (2012) Multivariate probabilistic projections using imperfect climate models. part ii: robustness of methodological choices and consequences for climate sensitivity. Clim Dyn 1–16

Sexton D, Murphy J, Collins M, Webb M (2011) Multivariate prediction using imperfect climate models part i: outline of methodology. Clim Dynam 1–30

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight SHE, Martin A, Murphy JM, Piani C, Sexton DMH, Smith L, Spicer R, Thorpe A, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433(7024):403–406

Susskind J, Barnet C, Blaisdell J (2003) Retrieval of atmospheric and surface parameters from airs/amsu/hsb data in the presence of clouds. IEEE Trans Geosci Remote Sens 41(2):390–409

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos T R Soc A 365(1857):2053–2075

Wielicki B, Barkstrom B, Harrison E, Lee R, Smith G, Cooper J (1996) Clouds and the earth's radiant energy system (ceres): an earth observing system experiment. Bull Am Meteorol Soc 77(5):853–868