INSIGHTS FROM LARGE ENSEMBLES WITH PERTURBED PHYSICS

Heiko Paeth

With 9 figures and 1 table
Received 09 October 2014 · Accepted 25 June 2015

Summary: State-of-the-art climate models are characterized by substantial deficiencies and their projections for the 21st century differ considerably depending on their specific initial conditions and physical parameterizations. Perturbed physics ensembles represent a promising tool to delimit the range of model uncertainty in a probabilistic sense. Yet large ensembles with complex climate models are still constrained by available computer resources. Here, an extended energy balance model with a stochastic weather term, feedback, and natural and anthropogenic forcings is used to investigate the potential of large perturbed physics ensembles. Up to 100,000 simulations for the 20th and 21st centuries are realized, assembling appropriate random numbers for 9 different model parameters. Bayesian model averaging is applied to filter the simulations in the face of observed temperatures and measurement error. The resulting ensemble mean has learned noticeably from the data and is very close to the observed time series of global-mean temperature. It differs markedly from the CMIP3 and CMIP5 multi-model ensemble means, especially in terms of more realistic decadal temperature variations due to the learning effect by means of the Bayesian approach. The sensitivity of the model and the Bayesian filter is assessed with respect to the model design and the estimated observational error, respectively.

Zusammenfassung: Aktuelle Klimamodelle unterliegen zahlreichen Unsicherheiten und ihre Projektionen für das 21. Jahrhundert unterscheiden sich erheblich aufgrund unterschiedlicher Anfangsbedingungen und physikalischer Parametrisierungen. Modellensembles mit gestörter Physik stellen ein viel versprechendes Instrument zur Eingrenzung des Wertebereichs der Modellunsicherheit im Sinne einer probabilistischen Vorhersage dar. Die gegenwärtig verfügbaren Rechenressourcen schränken die Anzahl an möglichen Ensemblesimulationen jedoch immer noch stark ein. Im vorliegenden Beitrag wird ein Energiebilanzmodell verwendet, das um einige Komponenten wie ein stochastischer Term der Wettervariabilität, ein Rückkopplungsmechanismus sowie natürliche und anthropogene Antriebsfaktoren erweitert wurde und sich für die Untersuchung sehr großer Modellensembles mit gestörter Modellphysik besonders gut eignet. Bis zu 100.000 Simulationen werden für das 20. und 21. Jahrhundert durchgeführt, wobei jeweils unterschiedliche Zufallszahlen für 9 Modellparameter eingesetzt werden. Anschließend werden die Modellsimulationen mit Hilfe des Bayes-Theorems anhand von Beobachtungsdaten und deren Messfehlern bewertet. Das resultierende kalibrierte Ensemblemittel offenbart einen deutlichen Lerneffekt durch die Beobachtungsdaten und stimmt mit der beobachteten Zeitreihe der global gemittelten bodennahen Temperatur sehr gut überein. Es unterscheidet sich hingegen deutlich von den Ensemblemitteln der CMIP3- und CMIP5-Multimodellensembles, vor allem in Hinblick auf eine realistischere Simulation der beobachteten dekadischen Variabilität. Dies ist auf den bayesischen Lerneffekt zurückzuführen. Im Beitrag wird ferner die Sensitivität des Energiebilanzmodells und des bayesischen Filteransatzes gegenüber dem Modelldesign und dem geschätzten Messfehlers dargelegt.

Keywords: Climate change, energy balance model, perturbed physics ensemble, global mean temperature, Bayesian approach

1 Introduction

Climate modeling is an indispensable tool to assess the future pathway of the Earth's climate under ongoing anthropogenic forcing. In contrast to purely statistical models which may be useful for shorter-term climate predictions (Lean and Rind 2009; Schönwiese et al. 2010), physically based climate models account for most of the processes and their non-linear interactions needed for the simulation of longer-term climate change (Murphy et al. 2004). However, climate model projections are uncertain by nature. Major sources of uncertainty are given by the unknown initial conditions of a simulation, missing processes and feedbacks and, in particular, the various physical parameterizations representing the effects of unresolved subgrid-scale processes (Palmer and Anderson 1994; Palmer and Williams 2008).

DOI: 10.3112/erdkunde.2015.03.01 ISSN 0014-0015 http://www.erdkunde.uni-bonn.de

As a consequence, climate models exhibit more or less striking biases with respect to observed climate features and their projections differ considerably. The first problem can be partly solved by statistical calibration in order to make climate model data suitable for impact models and past climate assessment (Hansen and Emanuel 2003; Doblas-Reyes et al. 2005; Paeth 2011; Paeth and Diederich 2011). The latter problem is more complex since the reliability of a future climate simulation cannot be determined immediately. Therefore, it is expedient to have a variety of model projections that span a certain range of conceivable future evolutions under reasonable boundary conditions and forcings (IPCC 2007). This leads to the concept of probabilistic climate prediction that aims at quantifying – though not limiting – uncertainty in climate projections (Räisänen and Palmer 2001; Wigley and Raper 2001; Forest et al. 2002; Collins et al. 2006; Paeth et al. 2013).

Many authors have evaluated climate change signals against the background of varying initial conditions and model spread and concluded that the human dimension of climate change often stands back from these sources of uncertainty (e.g., Paeth and Hense 2002; Tebaldi et al. 2005; IPCC 2007; Paeth and Pollinger 2010; Paeth et al. 2011a), which leaves us in a somehow embarrassed situation: On the one hand, scientists and decision makers on the climate impact side require a quite accurate insight into the future behaviour of climate, pleading towards a reduction of model spread. On the other hand, we still risk underestimating the range of possible future climate pathways since some feedbacks in the system or more extreme emissions scenarios are not yet accounted for in state-of-the-art climate models (Stainforth et al. 2005). In addition, climate models are quite close to each other in terms of some parameterizations and other characteristics which, themselves, are subject to large uncertainty (cf. IPCC 2013).

The best option is to vary all more or less uncertain model parameters simultaneously, according to appropriate probability distributions, and to evaluate the resulting ensemble in a probabilistic sense. This goes back to the idea of stochastic climate modeling (Hasselmann 1976). Indeed, Palmer and Williams (2008) postulate the upcoming era when climate models become stochastic with respect to their subgrid-scale processes. However, the challenge is enormous: Jackson et al. (2004) estimated that to properly assess a multi-dimensional probability distribution, which is based on no more than 10 model parameters, between 10^4 and 10^6 runs with a general circulation model (GCM) are required – a task that they judged to be impractical. Meanwhile, some benchmarks have been achieved in the form of so-called perturbed physics ensembles (PPEs). Some model parameters are randomly disturbed in PPEs before the simulation starts, and are then held constant over the entire simulation. In contrast, the original idea of stochastic modeling was to disturb the model parameters continuously during the simulation (Hasselmann 1976). It could be shown very recently that from a mathematical perspective such an approach is principally feasible on the basis of primitive equations that represent the dynamical core of climate models (Andreas Hense, personal communication).

In terms of previous studies with PPEs, Murphy et al. (2004) realized a 53-member ensemble with the atmospheric version of the Hadley Centre climate model, perturbing certain model parameters according to expert appraisal. Collins et al. (2006) made a similar analysis based on 17 ensemble members from the coupled version of the same model. Stainforth et al. (2005) even carried out 2,578 equilibrium experiments with perturbed physics for three 15-year time slices, which lead to a future warming rate that ranged between +2 °C and +11 °C; Piani et al. (2005) evaluated the confidence intervals for these.

As these extensive modeling efforts still do not meet the criterion according to Jackson et al. (2004), other studies relied on climate models with lower complexity in order to investigate the potential and requirements of climate modeling with perturbed parameters (Allen et al. 2000; Tomassini et al. 2007; Paeth et al. 2011b). Energy balance models (EBMs) play an important role in these climate models. In contrast to GCMs, EBMs are not three-dimensional but one- or zero-dimensional. Being based on the Earth's radiation budget, EBMs are closely tied to the global mean temperature. Although they do not account for atmospheric circulation and may neglect a number of processes and feedback mechanisms in the climate system, they have two prominent advantages: (1) ensemble sizes of several thousands of simulations can be realized, and (2) they represent an excellent tool towards sensitivity studies and process understanding.

Given the constraints in classic GCMs and the need for climate model ensembles with perturbed physics, the goal of the present study is to assess the potential of very large ensembles of climate model simulations with perturbed physics. For this purpose, a sophisticated EBM with feedback and a stochastic weather term was developed that simulates

transient variations and changes of the global-mean near-surface temperature under natural and anthropogenic forcings over the 20th and 21st centuries. The model comprises 9 parameters that are perturbed by imposing appropriate random numbers. It does not account for spatial variability and reduces the climate change problem to only one variable. Nevertheless, global-mean temperature is a crucial indicator of climate change for policy makers and planners (Stott and Kettleborough 2002; Lean and Rind 2009; Schönwiese et al. 2010) and it can be expected that EBMs have some skill in terms of the observed temperature time series, if boundary conditions and model parameters are realistically assigned (North et al. 1981). However, the major advantage is that the EBM approach can produce ensemble sizes in the order of magnitude required by Jackson et al. (2004) and, hence, allows for assessing the potential of large PPEs under various constraints.

Some of these constraints are the observational data themselves and their uncertainty. The idea of assigning weights to individual climate model simulations according to their ability to reproduce specific observed climate characteristics has some tradition (Allen et al. 2000). Model ability is determined by means of statistical methods such as regression models (Krishnamurti et al. 1999; Hansen and Emanuel 2003), eigen-value approaches (Doblas-Reyes et al. 2005) or, in particular, the Bayesian decision theorem (Tebaldi et al. 2005; Tomassini et al. 2007). The latter has its points because it accounts for the undeniable uncertainty of observational data and does not make assumptions on linearity (Jackson et al. 2004). The Bayesian method used in this study is called Bayesian model averaging (BMA) (Min and Hense 2006; Paeth et al. 2011b), although we do not use different but rather the same model with different parameter settings, or Bayesian updating according to Rougier (2007). It can be described as a filter of model simulations based on observational data combined with prior assumptions on the simulation and/or model (Min and Hense 2006; Paeth et al. 2011b). Related applications of the Bayesian approach refer to the climate change detection and attribution issue (Leroy 1998; Paeth et al. 2008). Craig et al. (2001) even suggest that all model simulations of complex non-linear systems should pass through a Bayesian filter.

The methodical approach is described in section 2, including the setup of the EBM and of the Monte Carlo experiments and the design of the Bayesian filter. Results are presented in section 3 and discussed in section 4 with respect to the climate change issue in general and the required PPE design for climate models of higher complexity. In section 5, the main conclusions from this study are drawn.

2 Methods

The approach basically consists of four steps that are illustrated in figure 1. First, a sophisticated EBM is developed, which represent the physical process model of this study. Second, model parameters are perturbed statistically and a large ensemble of Monte Carlo simulations with perturbed physics is realized. Third, a Bayesian filter is applied based on the observed time series of global mean temperature, leading to a calibrated ensemble of climate simulations.

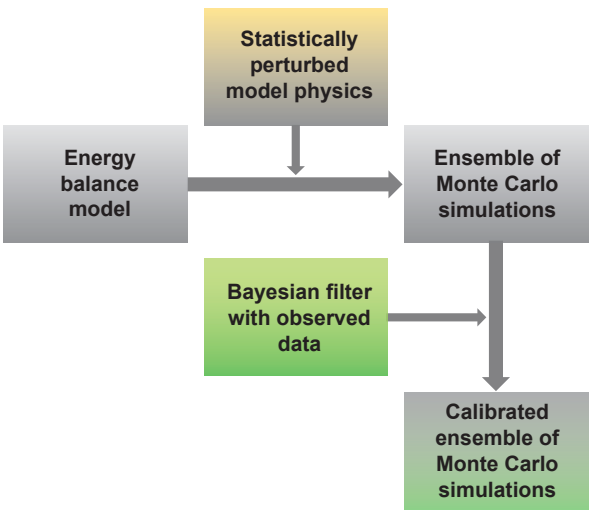

Fig. 1: Conceptual and methodical framework of the present study

2.1 Process model

For specific climatological problems in the field of process understanding and sensitivity studies, EBMs represent a useful option (North et al. 1981; Kettleborough et al. 2007). In contrast to Tomassini et al. (2007), the EBM used here is zero-dimensional, requiring some cautious consideration of the remaining degrees of freedom (see below). It simulates the temporal evolution of the global-mean near-surface temperature T as a function of the Earth's surface energy balance. As a positive feedback surface albedo is dependent on temperature and a stochastic term is introduced to excite tem-

perature variability at various time scales according to STORCH et al. (1999). In addition, solar variability and variations in greenhouse gas (GHG) and aerosol concentrations are taken into account. The discretized model equation for T can be written as

$$T(t) = T(t-1) + \frac{\Delta t}{c_w} \cdot \left[\frac{I_k(t)}{4} - \left(\alpha\left(T(t-1)\right) + \kappa(t) \right) \cdot \frac{I_k(t)}{4} - E \cdot \sigma \cdot T(t-1)^4 \cdot \pi(t) \right] \quad (1)$$

where Δt denotes a time step of 10 days according to the typical sequence of weather types in the middle latitudes, c_w is the oceanic heat storage capacity, which governs the response time of temperature to any kind of external forcing or internal perturbation. The default for c_w is set to $20 \cdot 10^7 \frac{\mathrm{J}}{\mathrm{m^2\ K}}$, representing an aqua-planet with an oceanic mixed-layer depth of 50 m. Note this one and some other parameter settings are adopted from STORCH et al. (1999). The default values for every model parameter are given in the third column of table 1. E accounts for the different spectral emissivity of the Earth's surface and is set to 0.95 by default, $\sigma = 5.67 \cdot 10^{-8} \frac{\mathrm{W}}{\mathrm{m^2\ K^4}}$ is the Stefan-Boltzmann constant.

Total solar irradiance I_k is time-dependent following the major quasi-periodic cycles of solar variability. For the simulation period 1900–2100 a future prediction of I_k is required. A simple repetition of solar cycle 23 as in LEAN and RIND (2009) is not convenient for a long-term prediction until 2100. In addition, past and future solar irradiance is subject to uncertainty and, hence, should be part of the perturbed physics approach. Therefore, a harmonic analysis has been applied (see Fig. 2) to the observed time series of solar irradiance since 1900 (LEAN 2000; WANG et al. 2005), which leads to the following equation for I_k depending on time t

$$I_k = \overline{I_k} + \sum_{k=1}^{\frac{n}{2}} \left[A_k \cdot cos\left(\frac{2\pi}{P} \cdot k \cdot t \right) + B_k \cdot sin\left(\frac{2\pi}{P} \cdot k \cdot t \right) \right] \quad (2)$$

with

$$A_k = \frac{2}{n} \sum_{t=1}^{n} y(t) \cdot cos\left(\frac{2\pi}{P} \cdot k \cdot 2 \right) \quad (3)$$

$$B_k = \frac{2}{n} \sum_{t=1}^{n} y(t) \cdot sin\left(\frac{2\pi}{P} \cdot k \cdot 2 \right) \quad (4)$$

as the coefficients of the Fourier-Bessel expansion for harmonic k, $y(t)$ is the observed solar irradiance time series of length n, and P denotes the basic period (WILKS 2006). Here, P equals 176 by default such that the quasi-periodic 88-year Gleissberg and 11-year sun spot cycles are integer factors of P (HOYT and SCHATTEN 1993). This approach for I_k is not necessarily better than any other assessment of solar variability but it allows for reducing the problem to a manageable number of parameters. Note that volcanic eruptions could also be accounted for by an abrupt reduction of I_k and a subsequent relaxation time of several months or years. Yet as future volcanic eruptions cannot be anticipated several years or decades ahead and, hence, no predictability can arise from this forcing component, volcanism will be disregarded in this study.

Surface albedo a is considered as a function of temperature T in the sense that higher (lower) glob-

Tab. 1: Considered means μη and standard deviations ση for the random disturbance of model parameters $\vec{\eta_i}$

Parameter	Symbol	μ_η	σ_η
heat capacity	c_w	$20 \cdot 10^7$	$5 \cdot 10^7$
emissivity	E	0.95	0.008
weather variability	σ_ω	0.02	0.005
initial value	T_0	288.15	0.5
greenhouse forcing	χ_{2011}	0.63625	0.0003
aerosol forcing	$\varkappa_{2003}$	0.00351	0.001
amplitude of albedo feedback	λ	0.027	0.007
transition of albedo feedback	ν	1.548	0.3
phase shift of solar forcing	P	0	1

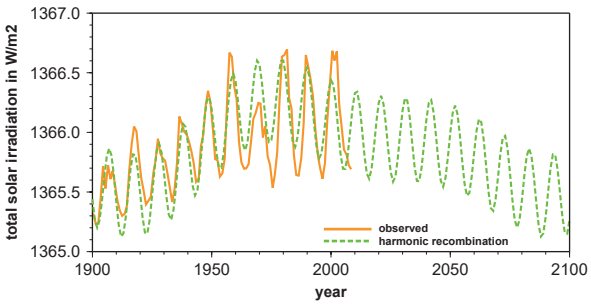

Fig. 2: Observed time series of the solar constant since 1900 and statistical modeling based on three harmonics leading to the future projection until 2100

al-mean temperature is associated with lower (higher) snow and ice cover and, thus, lower (higher) albedo. This relationship can be described with a hyperbolic function in the form

$$\alpha(T) = 0.3 \cdot \left[1 - \lambda \cdot tanh\left(\nu \cdot (T - 288.15)\right)\right] \quad (5)$$

where λ is proportional to the amplitude of the function as a measure of the temperature sensitivity of a, and ν defines the abruptness of the transition from high to low surface albedo, depending on attenuating effects in the climate system, i.e., negative feedbacks. The constants fix the inflexion point of the function at a spatial and temporal mean surface albedo of 0.30 and mean surface temperature of 15 °C. For the mean appearance of the hyperbolic function for a in figure 3b, λ is set to 0.025 and ν to 1.55 (cf. Storch et al. 1999). The positive albedo feedback is an important source of variability in the EBM, enhancing the amplitudes and causing a red shift in the spectrum of temperature variations and affecting the sensitivity of the model (Roe and Baker 2007).

Forcing by atmospheric aerosols, in particular anthropogenic sulphate aerosols, is parameterized by an additional surface albedo $\varkappa$ as a function of time t. According to IPCC (2013), the estimated net radiative forcing due to the direct and first indirect effects of aerosols in 2003 was $F = -1.2 \cdot \frac{W}{m^2}$. In the EBM this is equivalent to

$$\kappa(2003) = \frac{-F}{\frac{\overline{I_k}}{4}} = 0.00351 \quad (6)$$

Referring to the time series of SO_2 emissions from Smith et al. (2004), $\varkappa(2003)$ is assumed to be the maximum and at the beginning of the EBM simulations in 1900 $F = -0.17 \cdot \frac{W}{m^2}$. The third basic value in the year 2100 is taken from the SRES A1B emissions scenario with $F = -0.8 \cdot \frac{W}{m^2}$ (Nakicenovic and Swart 2000). The temporal interpolation of $\varkappa$ is done by fitting a second-order polynomial, leading to the baseline aerosol scenario in figure 3e.

Finally, $\tau(t)$ denotes the time-dependent transmissivity of the Earth's atmosphere with respect to outgoing long-wave radiation. It is composed of a shorter-term fluctuation $\omega(t)$ related to cloudiness and atmospheric water vapour content, and a longer-term trend due to the GHG increase:

$$\tau(t) = \omega(t) + \chi(t) \quad (7)$$

As $\chi(t)$ can neither be measured directly nor deduced theoretically, it is a tuning parameter of the EBM. Under early-industrial conditions it is set to $\chi(1900) = 0.644$, implying a global mean temperature of $T = 288.15K$. Up to today, increasing GHG concentrations may have caused a global warming of 0.8 °C compared with the pre-industrial era (IPCC 2007; Lean and Rind 2009; Schönwiese et al. 2010), which corresponds to $\chi(2011) = 0.636$. The CMIP3 multi-model ensemble mean response to the greenhouse forcing under SRES A1B emissions scenario suggests a warming rate of 2.9 °C until 2100 (IPCC 2007), leading to $\chi(2100) = 0.62$. The baseline GHG scenario in figure 3d arises from a second-order polynomial fit to these three basic values of $\chi(t)$. This non-linear negative trend of atmospheric transmissivity is superimposed by short-term fluctuations at the synoptic time scale. For $\omega(t)$ log-normally distributed random numbers with $\mu_\omega = 0$ and $\sigma_\omega = 0.02$ are drawn (Fig. 3a). This agrees with satellite data revealing that total global cloudiness and atmospheric water vapour content varies by up to 3 % at time scales of days to weeks (Storch et al. 1999). Each simulation receives a different random sequence of short-term weather fluctuations. $\omega(t)$ introduces a stochastic weather term to Eq. 1 and, hence, represents the major source of internal variability in the EBM. Although $\omega(t)$ arises from a white-noise process, the spectrum of temperature variations experiences a red shift (cf. Storch et al. 1999). Internal variability also arises from the initial value of T, which is set to $T_0 = 288.15\,K$ by default.

2.2 Perturbed physics ensemble

The process model described in subsection 2.1 contains parameters for which plausible estimates can be made from measurements and observational data, nevertheless, they are subject to uncertainty.

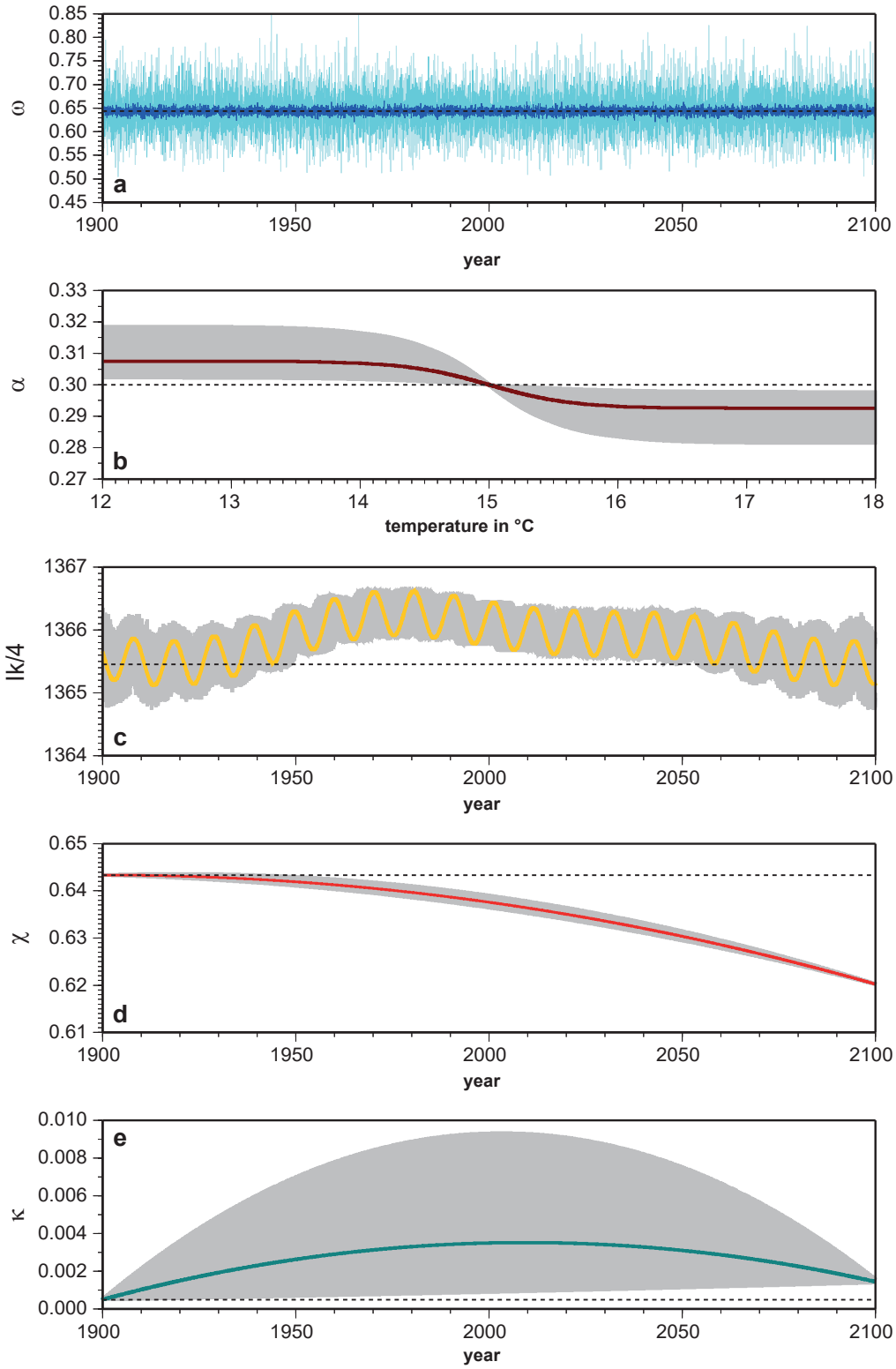

Fig. 3: Mean evolution and minimum-maximum ranges of the considered boundary conditions according to table 1: (a) stochastic weather forcing with the largest, the mean and the smallest value for σ_ω, (b) albedo-temperature feedback, (c) solar constant in $\frac{W}{m^2}$ is, (d) greenhouse forcing translated into transmissivity changes, (e) aerosols forcing translated into albedo changes

A good model simulation arises from an appropriate combination of parameters – appropriate in a sense that the model is close to the empirical data for which it constitutes the theoretical background. The PPE approach is targeted on finding such parameter sets from a large sample of random combinations (Hasselmann 1976; Palmer and Williams 2008). In this study, the PPE is based on $p = 9$ parameters of the EBM listed in table 1. The means and some of the standard deviations of the perturbed model parameters were derived from literature sources (cf. Storch et al. 1999; Smith et al. 2004; IPCC 2007). Some other standard deviations in table 1 are more or less intuitive. Therefore, the sensitivity of the PPE to this choice will also be assessed. Furthermore, it is assumed that the inflexion point in Eq. 5 and the early-industrial GHG and aerosol forcings are assured. In addition, the radiative forcing in 2100 is fixed on the SRES A1B emissions scenario (Nakicenovic and Swart 2000). Thus, uncertainties in the emissions scenarios are not accounted for (cf. Schenk and Lensink 2007).

Up to 100,000 ensemble members, i, are realized each with statistically independent perturbations of the parameter set $\vec{\eta}_i$. In contrast to Murphy et al. (2004) and Stainforth et al. (2005), the prior probability for specific parameter settings is not based on expert knowledge. All model parameters are varied according to appropriate probability density functions (PDFs), as recommended by Lempert et al. (2004). For the phase shift P of the solar cycles, a Gaussian PDF with μ_η and σ_η according to table 1 can be assumed. The extraction of normally distributed random numbers x_i is done by means of the Box-Muller method (Wilks 2006). All other perturbed model parameters η_j in table 1 must not become negative. Accordingly, Gamma distributed variates are derived using the expectations and standard deviations listed in table 1 (Wilks 2006).

2.3 Bayesian model averaging

Given the considerable model spread arising from the PPE approach, an objective tool is required to decide which simulation is more likely than another. If this decision is made with respect to uncertain observational data, Bayesian theory is an appropriate methodical option (Bernardo and Smith 1994; Jackson et al. 2004). In the present case, we have ensemble simulations $\vec{f}_i$ with emerging random parameter set $\vec{\eta}_i$ on the one hand and observational data $\vec{d}$ with measurement error θ as a control parameter on the other hand. Assuming that $\vec{f}_i$ is independent of $\vec{d}$, the Bayesian theorem can be expressed as (Paeth et al. 2011b)

$$\underbrace{p\left(\vec{f}_i \mid \vec{d}\right)}_{I} \propto \underbrace{p\left(\vec{d} \mid \vec{f}_i\right)}_{II} \cdot \underbrace{p\left(\vec{f}_i\right)}_{III} \tag{8}$$

The posterior (term I) indicates the probability of simulation $\vec{f}_i$ with parameter set $\vec{\eta}_i$ and statistical parameter θ given the data $\vec{d}$. It is the product of a likelihood function (term II) as a model for the data and the prior of parameter set $\vec{\eta}_i$ (term III) as the existing knowledge before collecting the data. Thus, the Bayesian approach allows for distinguishing between prior assumptions on the process model and observational constraints (Collins et al. 2006). Note that the same parameter set $\vec{\eta}_i$ always leads to the same simulation $\vec{f}_i$ in the deterministic process model.

Assuming that observational errors are normally distributed, the q-dimensional likelihood function is given by

$$p\left(\vec{d} \mid \vec{f}_i\right) = \frac{1}{\sqrt{(2\pi)^q \, det\Sigma_0}} \cdot exp\left[\frac{-1}{2}\left(\vec{d} - \vec{f}_i\right)^T \Sigma_0^{-1}\left(\vec{d} - \vec{f}_i\right)\right] \tag{9}$$

where q denotes the number of time steps in $\vec{f}_i$ and $\vec{d}$, in this case annual means over the period 1900–2100. θ is the covariance matrix of the observational error. It is a diagonal matrix with parameter θ supposing that observational errors are IID. This implies that no structural uncertainty is accounted for in terms of temperature measurements over time and space. Determining the spatio-temporal structure of observational errors and, hence, indicating all off-diagonal terms of θ is beyond the scope of this study. Even the exact value of θ is unknown: Estimates from Folland et al. (2001) and Brohan et al. (2006) vary between 0.002 °C² and 0.1 °C². A thorough error propagation analysis of the original station data underlying the CRU data set would help to solve this problem or, at least, provide an accurate prior distribution for θ. In the present study, θ is varied as a control parameter in order to assess its impact on the BMA results (cf. Campbell 2005; Tomassini et al. 2007).

The prior of the simulations $\vec{f}_i$ with individual parameter set $\vec{\eta}_i$ (term III) arises from the product of the function values ϱ of the Gamma and normal distribution, respectively, for each model parameter η_j in ensemble member i:

$$p\left(\vec{f}_i\right) = \prod_{j=1}^{p} \rho\left(\eta_{i,j}\right) \tag{10}$$

This prior definition is problematic in the sense that it assigns a very small probability to parameter sets outside a relatively narrow range around the prior mean parameter set. Therefore, this proceeding is only justified, if a good choice of the prior mode is made. Alternatively, Murphy et al. (2004) have used uniform or trapezoidal instead of triangular prior distributions. In the present study, it will be shown that, on the one hand, the chosen prior mode is reasonable and, on the other hand, the Bayesian filter still leads to a noticeable improvement. In addition, the BMA will also be carried out with a uniform prior in order to assess to what extent the filtered ensemble mean is constrained by the data and by the prior, respectively. Finally, the posteriors are weighted by the normalized likelihood in order to ensure that they sum up to one, and are used as weighting factors for the PPE members. The posterior ensemble mean $\overline{\overline{f_B}}(t)$ at each time step t which emanates from the BMA approach, is calculated by

$$\overline{\overline{f_B}}(t) = \sum_{i=1}^{m} \left[p\left(\overline{f_i} \mid \vec{d}\right) \cdot \overline{f_i}(t) \right] \tag{11}$$

Accordingly, the ensemble spread of the BMA approach is defined by the time-dependent confidence interval $\left[\overline{\overline{f_B}}(t) \pm 2 \cdot \sigma_\omega(t)\right]$ with

$$\hat{\sigma}_\omega(t) = \sqrt{\sum_{i=1}^{m} \left[p\left(\overline{f_i} \mid \vec{d}\right) \cdot \left(\overline{f_i}(t) - \overline{\overline{f_B}}(t)\right)^2 \right]} \tag{12}$$

Finally, the so called effective sample size (ESS) is determined by

$$ESS = \frac{1}{\Sigma_{i=1}^{m} p\left(\vec{d} \mid \overline{f_i}\right)^2} \tag{13}$$

with $p(\vec{d}|\overline{f_i})'$ being the normalized likelihood weights according to Eq. 9. The ESS represents a measure of the number of ensemble members effectively contributing to the posterior ensemble mean $\overline{\overline{f_B}}(t)$ that emanates from a Bayesian filtering (Doucet et al. 2000). It ranges between $ESS = n$ in the case that all likelihoods have equal weights $\frac{1}{m}$ and $ESS = 1$ in the case that one of the weights equals *1*.

3 Results

Concerning the prescribed time series of total solar irradiance, figure 2 shows that the observed data $y(t)$ can be well reproduced by a linear combination of three harmonics with periods 176, 88, and 11 years accounting for 92.8 % of the total variability of $y(t)$. Uncertainties in the assessment of solar irradiance towards the beginning of the observed time series and during the prediction period until 2100 are dealt with by a stochastic variation of the phase shift P (see section 2.2) and an imposed random number that grows larger in the past and in the future (see Fig. 3c). The statistical prediction of I_k is also depicted in figure 3c. The panels of figure 3 indicate that the perturbed model parameters are characterized by different amounts of uncertainty. Large uncertainty is associated with the aerosol forcing (Fig. 3e), the greenhouse forcing is more bound (Fig. 3d), and the other parameters denote intermediate uncertainty (Fig. 3a-c), according to the assumptions made in sub-section 2.1.

The results of the PPE approach using the EBM described in subsections 2.1 and 2.2 are summarized in figure 4. The panels refer to different scalings of σ_η in table 1 as a measure of the amplitudes of perturbation. Each panel contains the ensemble mean time series of annual temperature, the total range of temperature values over 10,000 ensemble members, and the best and worst individual simulation as indicated by the root mean-square error (RMSE) (Wilks 2006) in terms of the observed time series of global-mean temperature. The latter is based on the CRU data set (Mitchell and Jones 2005) in the extended version covering the period 1901–2013. The uncertainty related to this data set is documented in subsection 2.3. As a common feature of all three panels, the ensemble mean time series starts rising at the end of the 20th century and reaches a warming rate of almost 3 °C in 2100. The best simulation is close to the ensemble mean but with some crucial exceptions (see below). The worst simulation is rather odd and can be characterized by a prominent transition towards very low or very high temperatures at the beginning of the simulation period. This transition arises from a random combination of initial temperature value, stochastic weather forcing and a distinctive albedo feedback (cf. Roe and Baker 2007). Such unrealistic simulations are due to the fact that the random perturbations imposed on the model parameters are fully independent of each other, in contrast to the systematically attuned PPE approach by Culina et al. (2011).

The spread of the ensemble members is proportional to the scaling of σ_η. The temperature change until 2100 is in the range [-1 °C, +6 °C] when scaling the standard deviations in table 1 with a prefactor of $\Phi = 1$ (Fig. 4b) and in the range [-4 °C, +10 °C] for $\Phi = 2$ (Fig. 3c). The latter is probably too large

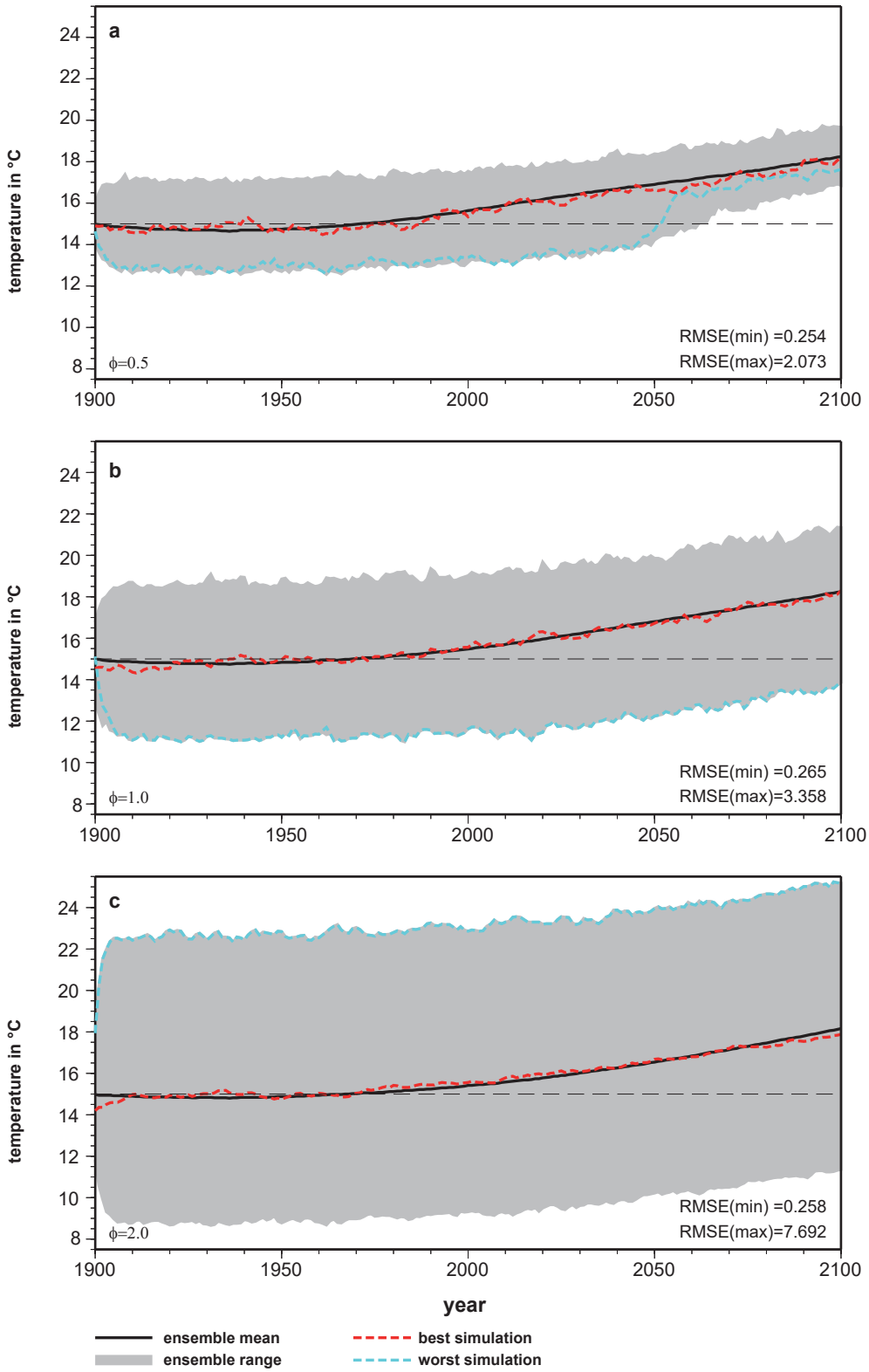

Fig. 4: Ensemble mean, minimum-maximum range and two individual temperature evolutions out of 10,000 Monte Carlo simulations. The best and worst simulation is defined by the smallest and largest RMSE, respectively, compared with the observed time series during the period 1901–2013. Values for σ_η in table 1 multiplied by a prefactor of $\Phi = 0.5$ (a), $\Phi = 1$ (b), $\Phi = 2$ (c)

compared with the PPE approach by STAINFORTH et al. (2005). Intuitively, one would assign rather low probabilities to the extreme simulations (cf. WIGLEY and RAPER 2001) – a matter for the BMA approach (see below). The RMSE is also dependent on Φ (Fig. 4): For the best simulation it is always below 0.3 °C, which is low but still higher than the measurement error of the observations (FOLLAND et al. 2001; BROHAN et al. 2006). Stronger perturbations are generally associated with a reduced RMSE of the best simulation, yet the difference is relatively small. In contrast, the RMSE of the worst simulation increases notably with Φ.

Furthermore, the RMSE is a nonlinear function of the number of Monte Carlo simulations (Fig. 5): The best simulation profits from an increasing ensemble size up to 10,000 members; subsequently the RMSE stabilizes around 0.2 °C. The RMSE of the worst simulation has its stabilization level beyond 60,000 ensemble members. This leads to three important conclusions: (1) Despite large PPE sizes and pronounced perturbations, the simulations are close but cannot become closer to the observations because the processes included in the EBM do not allow for it. (2) In the case of the EBM, 10,000 ensemble members appear to be most expedient. In addition, a prefactor of $\Phi = 1$ is retained because it best matches the parameter uncertainty deduced from the literature (see subsections 2.1 and 2.2). (3) For more complex models it can be expected that the optimal number of Monte Carlo simulations is much larger than 10,000 (cf. JACKSON et al. 2004).

The results of the BMA approach are illustrated in figure 6, showing the posterior ensemble mean time series $\overline{\overline{f_B}}$ from the BMA approach, using a Gamma distributed and a uniform prior, respectively, along with the observed time series until 2013 and the unweighted (prior) ensemble mean over 10,000 PPE simulations. The posterior ensemble means are very close to the observations, at least at the decadal to inter-decadal time scale. In particular, the combined dynamical-statistical approach of EBM and BMA is able to reproduce the warm 1940s, the turnaround with slight cooling in the 1950s and 1960s, and the distinct temperature rise thereafter (cf. IVANOV and EVTIMOV 2010; THOMPSON et al. 2010). The prior ensemble mean, which has not learned from the data, only agrees in terms of the recent temperature trend and slightly underestimates the warming rate, just like most state-of-the-art GCMs (STAINFORTH et al. 2005). Both posterior ensemble means are quite similar to each other, at least they reflect the same decadal component of temperature variability. The RMSE is somewhat lower for the Gamma distributed prior compared with the uniform prior. This indicates that the dominant constraint in the BMA approach is given by the observations, while a second order effect arises from the Gamma distributed prior.

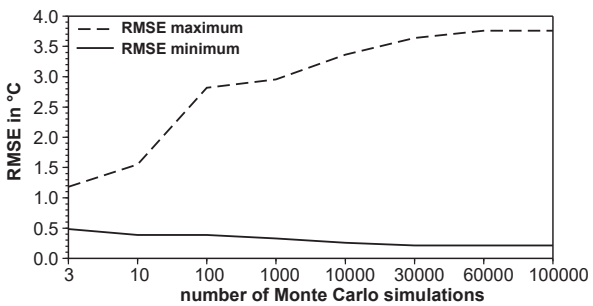

Fig. 5: Minimum and maximum RMSE for Φ = 1 between simulated and observed global-mean temperature over the 1901–2013 validation period as a function of number of Monte Carlo simulations

The remarkable performance of the BMA filter as displayed in figure 6 may also be interpreted as the consequence of an overfitting problem. In fact, 9 model parameters are fitted to an observed time series of 113 annual mean values. Given some expectable serial autocorrelation imposed by the combination of weather forcing and albedo feedback (STORCH et al. 1999), as well as by solar and radiative forcing, it cannot be ruled out a priori that the number of fitted parameters exceeds the temporal degrees of freedom of the considered time series. To determine the temporal degrees of freedom of the model system, an empirical orthogonal function (EOF) analysis is applied to the 10,000 member ensemble used for the BMA approach and the number of EOFs required to account for a certain part of global-mean total temperature variability is assessed (Fig. 7). The individual runs were filtered by the ensemble mean in order to remove some of the common warming signal due to greenhouse gas forcing. It turns out that at least 25 EOFs are needed to represent 90 % of total variability, for 95 % it is 41 EOFs and for 99 % even 82 EOFs. Depending on the threshold of total temperature variability to be represented, the number of mutually independent EOFs that can be taken as a measure of degrees of freedom (WILKS 2006) is definitely larger than the number of fitted model parameters. Note that the overfitting problem can be neglected entirely when some kind of spatial differentiation is carried out (using spatially undifferentiated model parameters). In figure 6 $\theta = 0.04$ was set, which is at the lower boundary, as compared with the range spanned by FOLLAND et al. (2001) and BROHAN et al. (2006).

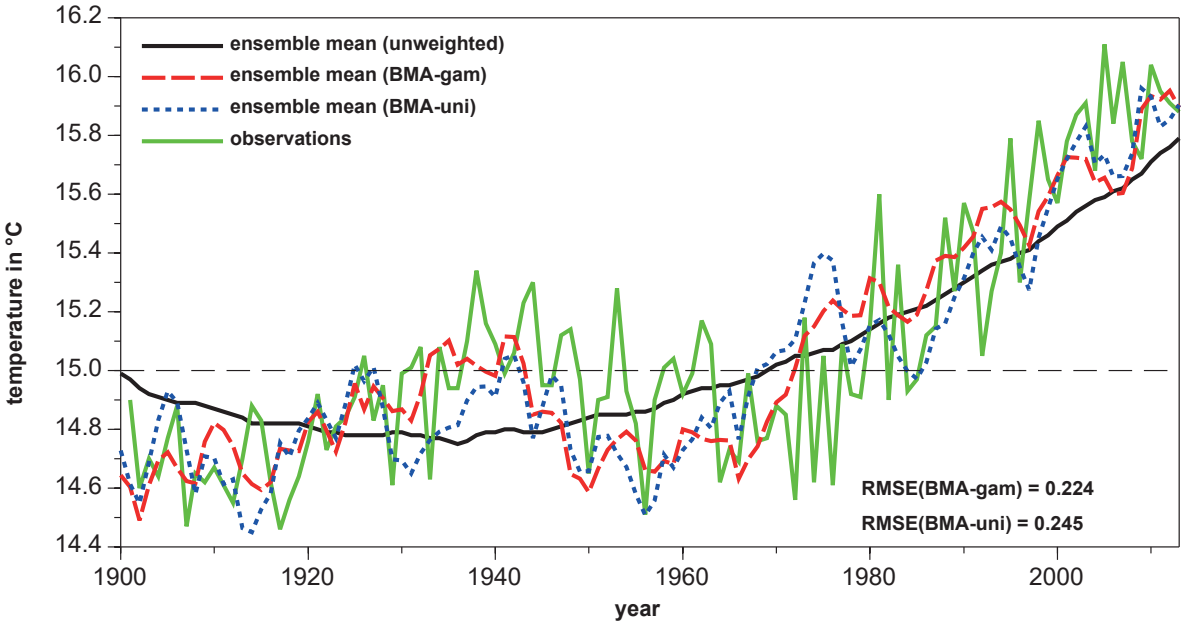

Fig. 6: Simulated and observed global-mean temperature over the hindcast period 1901–2013, once from the unweighted ensemble mean and once from the data-filtered ensemble mean arising from the BMA approach, once for a Gamma distributed prior (BMA-gam) and once for a uniform prior (BMA-uni), using prefactor $\Phi = 1$ and observational parameter $\theta = 0.04$

As θ determines the relative values of the likelihood function among simulations $\overline{f_i}$, it dictates how the posterior ensemble mean is composed (cf. Campbell 2005; Tomassini et al. 2007). Figure 8 depicts some aspects of the BMA approach in relation to θ. The ESS (cf. Eq. 13) strictly increases with θ. This implies that for small measurement errors, the posterior ensemble mean is highly focussed on the best simulations, while more simulations contribute to $\overline{\overline{f_B}}(t)$ for larger values of θ. In general, ESS is small, i.e., closer to *1* than to $m = 10{,}000$, because it is constrained by a relatively low measurement error compared with the model bias. At the same time, the RMSE of the posterior ensemble mean tends to decrease slightly with increasing data uncertainty since more aspects from different simulations are taken into account and may enable the dynamical-statistical model to match the observations more accurately. For all considered estimates of θ, the posterior ensemble mean clearly outperforms the best individual simulation (see dashed line in figure 8). Note that $\theta > 0.1$ is quite unrealistic.

In figure 9 the posterior ensemble mean and ensemble range are compared with the CMIP3 and CMIP5 multi-model ensembles of coupled global GCMs as the current standard to assess future climate change (IPCC 2007, 2013; Taylor et al. 2012). Individual simulations from 23 CMIP3 climate models are available for the emissions scenario A1B 48. CMIP5 provides 53 simulations from 25 CGCMs under the RCP4.5 emissions scenario, which is of intermediate nature and comparable (but not identical) to SRES A1B (Meinshausen et al. 2011). The ensemble range of the unweighted PPE approach presented here is much larger than in CMIP3 and CMIP5. Obviously, model parameters differ more in the PPE than in CMIP multi-model approaches (cf.

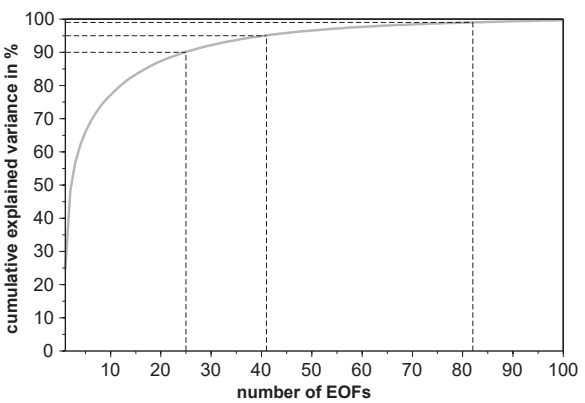

Fig. 7: Number of EOFs and related cumulative explained variance derived from the ensemble of 10,000 Monte Carlo simulations used for the BMA-gam approach, filtered by the ensemble mean. The vertical dashed lines denote the number of EOFs necessary to represent at least 90%, 95% and 99%, respectively, of simulated global-mean total temperature variability over the period 1900–2013

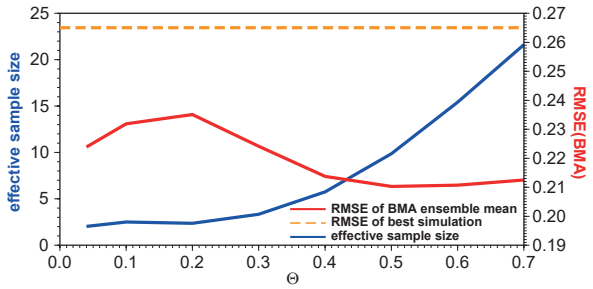

Fig. 8: Effective sample size of the BMA-gam approach and associated RMSE of the BMA ensemble mean over the 1901–2013 validation period as a function of observational parameter θ

Stainforth et al. 2007). The data filtering associated with the BMA leads to some crucial differences between $\overline{\overline{f}}_B(t)$ and conventional CGCMs ensemble means, simply because the latter are uninitialized and, hence, not bound to the observed decadal variations of global-mean temperature. Rather they are characterized by a continuous warming trend, which accelerates in the 21st century, but still underestimates the recently observed trend (Stainforth et al. 2007), making it comparable with the prior ensemble means in figure 4. It is arguable whether a BMA approach applied to the CMIP3 and CMIP5 simulations would be able to overcome these deficiencies. In 2100, the posterior ensemble mean matches the CMIP3 ensemble mean because the GHG forcing parameter $\chi(t)$ has been tuned to the mean temperature response of CMIP3 under the A1B scenario (see subsection 2.1). Another important finding is that the ensemble spread of the BMA approach, which has learned from the data, is by far smaller than in all other ensembles, suggesting a lower level of uncertainty in future climate projection. Of course, this is related to the relatively small effective ensemble size (cf. Fig. 8). Anyhow, it represents a clear pleading for weighting individual climate simulations according to their performance.

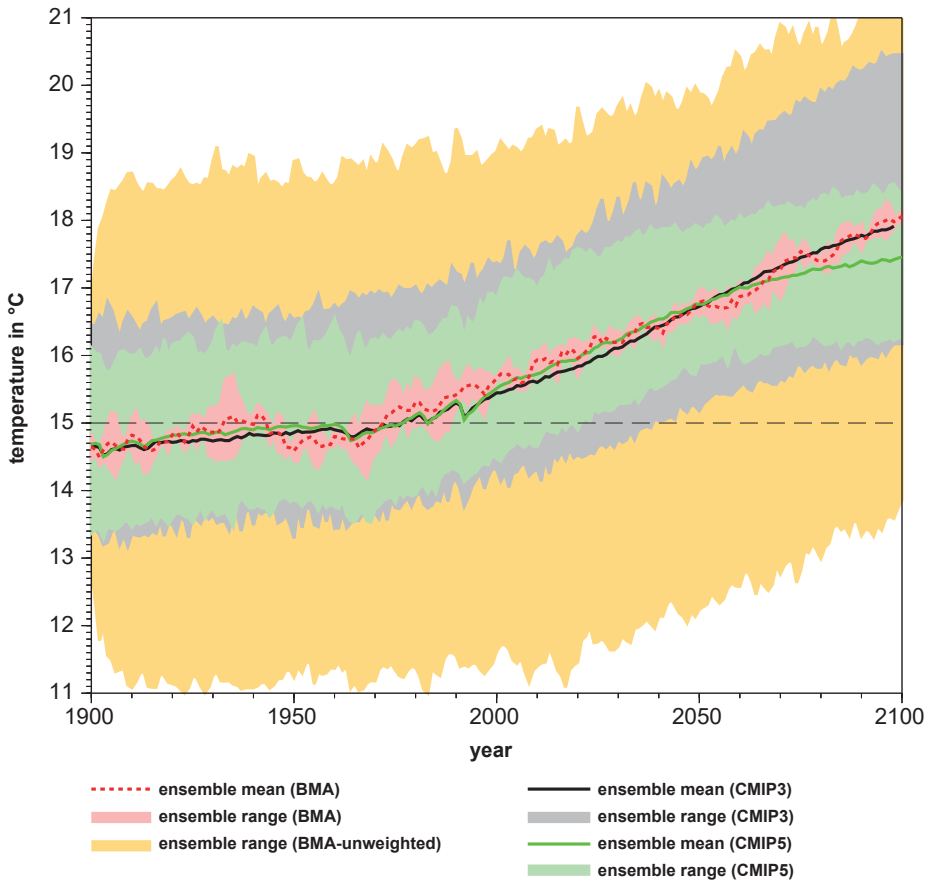

Fig. 9: Ensemble mean and ensemble range compared between the BMA-gam approach, the perturbed physics ensemble (both presented here), and CMIP3 and CMIP5 multi-model ensembles of coupled global GCMs over the 1900–2100 period under emissions scenario A1B and RCP4.5, respectively

4 Discussion

Applying the PPE approach to an extended EBM with natural and anthropogenic forcings leads to a large variety of simulated temperature evolutions over the period 1900–2100. Some simulations are rather odd, others match the observed decadal and interdecadal variations of global-mean temperature quite well. The limited complexity of the EBM entails that the RMSE of the best simulations cannot decrease below a certain threshold, which is reached with about 10,000 ensemble members. The BMA method is associated with a distinct learning effect of the EBM, using the data themselves as a benchmark. The posterior ensemble mean time series reproduces the observed early 20th century warming phase, the slight cooling period until the 1970s and the recent temperature rise excellently. At a first glance, the correct decadal variability may arise from the prescribed solar forcing, while the long-term changes may relate to the interplay between greenhouse gas and aerosol forcings. However, the EBM is nonlinear and, hence, the frequency of the forcing is not necessarily identical with the frequency of the climate response. This is supported by the fact that all ensemble members are subject to about the same time series of solar irradiance, but the unweighted ensemble means in figure 4 and 6 do not reflect the observed decadal fluctuations mentioned above. Overfitting does not play a noticeable role in the efficiency of the dynamical-statistical model approach. The success of the BMA concept not only depends on the performance of the process model but also on the measurement error. As the prior distribution of this error is unknown, the BMA was carried out conditionally on θ. Regardless of the choice of θ, the posterior ensemble mean clearly outperforms the best individual temperature simulation. Compared with the CMIP3 and CMIP5 multi-model ensembles of global GCMs, the combined dynamical-statistical model presented here captures many more features of the observed temperature evolution in the 20th century and is characterized by a smaller range of uncertainty in future temperature projection.

The usefulness of the BMA approach has already been reported by many authors (e.g., Stainforth et al. 2005; Tebaldi et al. 2005; Min and Hense 2006; Tomassini et al. 2007; Paeth et al. 2011b) as well as the need for large ensembles with randomly disturbed model physics (e.g., Hasselmann 1976; Räisänen and Palmer 2001; Palmer and Williams 2008). The new point about the present study is that the PPE concept was applied to a physically based, though zero-dimensional, climate model, which has a good performance in terms of global-mean temperature and allows for the realization of more than 10,000 ensemble members in order to meet the criterion by Jackson et al. (2004). Several of such and even larger ensembles were created to assess how the dynamical-statistical model depends on crucial constraints such as the amplitudes of parameter perturbations, the ensemble size, the uncertainty of the data and different shapes of the prior distribution (cf. Frame et al. 2005).

In terms of $\overline{\overline{f_B}}$, a comparable reproduction of 20th century global-mean temperature was also achieved by statistical modeling (Schönwiese et al. 2010), but with a limited prospect of climate prediction, and by global GCMs with natural and anthropogenic forcings (Stott et al. 2000; IPCC 2007), but with much more computational effort and a much smaller ensemble size. The size of the PPE is related to the probability of getting a 'good' simulation. Possibly the required number of runs can be reduced by implementing systematic rather than purely random perturbations (cf. Culina et al. 2011). Using the so-called discrepancy as a measure of the model's prior structural imperfection may also help to exclude inadequate simulations (Sexton et al. 2012; Sexton and Murphy 2012). Nevertheless, a large ensemble also ensures a more accurate shaping of the PDFs for probabilistic forecasts (Räisänen and Palmer 2001; Wigley and Raper 2001).

5 Conclusions

Although the presented dynamical-statistical model appears to have some skill at the level of global-mean temperature, the resulting temperature prediction should not be over-interpreted in a quantitative sense. Therefore, the complexity of the EBM is not sufficient. Especially, important feedbacks of the real climate system, which affect climate sensitivity, are missing (cf. Hansen et al. 2006; Roe and Baker 2007). For instance, it cannot be ruled out that the model correctly reproduces the observed decadal-scale fluctuations of global mean temperature for the wrong reason, because the relevant processes, e.g., like ocean circulation as suggested by Keenlyside et al. (2008), are not accounted for in the EBM. While some additional processes such as a temperature-dependent term in the equation for $\tau(t)$ (Eq. 7) could be incorporated in the EBM as a measure of the temperature-water vapour feedback or the oceanic carbon sink feedback, the strength of this approach lies more in the prospect of sensitivity studies. One promising application could be the testing of a large number

of different scenarios for anthropogenic forcing (cf. Kettleborough et al. 2007; Schenk and Lensink 2007) and their interplay with natural climate drivers or the investigation of individual extreme events such as severe volcanic eruptions and meteor strikes.

Undoubtedly, GCMs and even more Earth system models represent the only tool to predict climate in a reliable way, because they – hopefully – account for most of the relevant processes in the climate system (Murphy et al. 2004). However, the insights into PPEs gained from this study suggest that the GCM ensemble sizes generated until now in the PPE context (e.g. Stainforth et al. 2005; Collins et al. 2006) and in the CMIP framework of the IPCC (IPCC 2007, 2013) are nowhere near large enough (cf. Jackson et al. 2004). It is likely that we have to long for computing resources allowing for ensemble sizes beyond one million members.

Acknowledgements

Observed temperature data were provided by the Climatic Research Unit and GCM data were derived from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) in the framework of the WCRP CMIP3 and CMIP5 projects. Solar irradiance reconstructions were downloaded from KNMI. I thank Felix Pollinger for pre-processing the GCM data and Andreas Hense for helpful discussions. Furthermore, special thanks go to Jonathan Rougier and a second anonymous reviewer who helped to improve the manuscript considerably.

References

Allen, M. R.; Stott, P. A.; Mitchell, J. F. B.; Schnur, R. and Delworth, T. L. (2000): Quantifying the uncertainty in forecasts of anthropogenic climate change. In: Nature 407, 617–620. DOI: 10.1038/35036559

Bernardo, J. M. and Smith, A. F. M. (1994): Bayesian theory. New York.

Brohan, P.; Kennedy, J. J.; Harris, I.; Tett, S. F. B. and Jones, P. D. (2006): Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. In: Journal of Geophysical Research 111, D12106. DOI: 10.1029/2005JD006548

Campbell, E. P. (2005): Statistical modeling in nonlinear systems. In: Journal of Climate 18, 3388–3399. DOI: 10.1175/JCLI3459.1

Collins, M.; Booth, B. B. B.; Harris, G. R.; Murphy, J. M.; Sexton, D. M. H. and Webb, M. J. (2006): Towards quantifying uncertainty in transient climate change. In: Climate Dynamics 27, 127–147. DOI: 10.1007/s00382-006-0121-0

Craig, P. S.; Goldstein, M.; Rougier, J. C. and Seheult, A.H. (2001): Bayesian forecasting for complex systems using computer simulators. In: Journal of the American Statistical Association 96, 717–729. DOI: 10.1198/016214501753168370

Culina, J.; Kravtsov, S. and Monahan, A. H. (2011): Stochastic parameterization schemes for the use in realistic climate models. In: Journal of the Atmospheric Sciences 68, 284–299. DOI: 10.1175/2010jas3509.1

Doblas-Reyes, F. J.; Hagedorn, R. and Palmer, T. N. (2005): The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. In: Tellus 57A, 234–252. DOI: 10.1111/j.1600-0870.2005.00104.x

Doucet, A.; Godsill, S. and Andrieu, C. (2000): On sequential Monte Carlo sampling methods for Bayesian filtering. In: Statistics and Computing 10, 197–208. DOI: 10.1023/A:1008935410038

Folland, C. K.; Rayner, N. A.; Brown, S. J.; Smith, T. M.; Shen S. S. P.; Parker, D. E.; Macadam, I; Jones, P. D.; Jones, R. N.; Nicholls, N. and Sexton, D. M. H. (2001): Global temperature change and its uncertainties since 1861. In: Geophysical Research Letters 28, 2621–2624. DOI: 10.1029/2001GL012877

Forest, C. E.; Stone, P. H.; Sokolov, A. P.; Allen, M. R. and Webster, M. D. (2002): Quantifying uncertainties in climate system properties with the use of recent climate observations. In: Science 295, 113–117. DOI: 10.1126/science.1064419

Frame, D. J.; Booth, B. B. B.; Kettleborough, J. A.; Stainforth, D.A.; Gregory, J. M.; Collins, M. and Allen, M. R. (2005): Constraining climate forecasts: the role of prior assumptions. In: Geophysical Research Letters 32, L09702. DOI: 10.1029/2004GL022241

Hansen, J. A. and Emanuel, K. A. (2003): Forecast 4D-Var: exploiting model output statistics. In: Quarterly Journal of the Royal Meteorological Society 129, 1255–1267. DOI: 10.1256/qj.01.75

Hansen, J.; Sato, M.; Ruedy, R.; Lo, K.; Lea, D. W. and Medina-Elizade, M. (2006): Global temperature change. In: Proceedings of the National Academy of Science 103, 14288–14293. DOI: 10.1073/pnas.0606291103

Hasselmann, K. (1976): Stochastic climate models. Part I: Theory. In: Tellus 28, 473–484. DOI: 10.1111/j.2153-3490.1976.tb00696.x

Hoyt, D. V. and Schatten, K. H. (1993): A discussion of plausible solar irradiance variations, 1700–1992. In: Journal of Geophysical Research 98, 18895–18906. DOI: 10.1029/93JA01944

IPCC (2007): Climate change 2007: the physical science basis. Contribution of working group I to the fourth Assessment Report of the Intergovernmental Panel on Climate

Change [Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K. B.; Tignor, M. and Miller, H. L. (eds.)]. Cambridge.

– (2013): Climate change 2013: the physical science basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F.; Qin, D.; Plattner, G.-K.; Tignor, M.; Allen, S. K.; Boschung, J.; Nauels, A.; Xia, Y.; Bex, V. and Midgley, P. M. (eds.)]. Cambridge.

Ivanov, M. A. and Evtimov, S. N. (2010): 1963: The break point of the Northern Hemisphere temperature trend during the twentieth century. In: International Journal of Climatology 30, 1738–1746. DOI: 10.1002/joc.2002

Jackson, C.; Sen, M. K. and Stoffa, P. L. (2004): An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. In: Journal of Climate 17, 2828–2841. DOI: 10.1175/1520-0442(2004)017<2828:AESBAT>2.0.CO;2

Keenlyside, N.; Latif, M.; Jungclaus, J.; Kornblueh, L. and Roeckner, E. (2008): Advancing decadal-scale climate prediction in the North Atlantic sector. In: Nature 453, 84–88. DOI: 10.1038/nature06921

Kettleborough, J. A.; Booth, B. B. B.; Stott, P. A. and Allen, M. R. (2007): Estimates of uncertainty in predictions of global mean surface temperature. In: Journal of Climate 20, 843–855. DOI: 10.1175/jcli4012.1

Krishnamurti, T. N.; Kishtawal, C. M.; La Row, T. E.; Bachiochi, D. R.; Zhang, Z.; Williford, C. E.; Gadgil, S. and Surendan, S. (1999): Improved weather and seasonal climate forecasts from multimodel superensemble. In: Science 285, 1548–1550. DOI: 10.1126/science.285.5433.1548

Lean, J. L. (2000): Evolution of the sun's spectral irradiance since the Maunder Minimum. In: Geophysical Research Letters 27, 2425–2428. DOI: 10.1029/2000GL000043

Lean, J. L. and Rind, D. H. (2009): How will Earth's surface temperature change in future decades? In: Geophysical Research Letters 36, L15708. DOI: 10.1029/2009GL038932

Lempert, R.; Nakicenovic, N.; Sarewitz, D. and Schlesinger, M. (2004): Characterizing climate-change uncertainties for decision-makers. In: Climatic Change 65, 1–9. DOI: 10.1023/B:CLIM.0000037561.75281.b3

Leroy, S. S. (1998): Detecting climate signals: some Bayesian aspects. In: Journal of Climate 11, 640–651. DOI: 10.1175/1520-0442(1998)011<0640:dcssba>2.0.co;2

Meinshausen, M.; Smith, S. J.; Calvin, K.; Daniel, J. S.; Kainuma, M. L. T.; Lamarque, J.-F.; Matsumo, K.; Montzka, S. A.; Raper, S. C. B. and Riahi, K. (2011): The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. In: Climatic Change 109, 213–241. DOI: 10.1007/s10584-011-0156-z

Min, S. K. and Hense, A. (2006): Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. In: Geophysical Research Letters 33, L08708. DOI: 10.1029/2006GL025779

Mitchell, T. D. and Jones, P. D. (2005): An improved method of constructing a database of monthly climate observations and associated high-resolution grids. In: International Journal of Climatology 25, 693–712. DOI: 10.1002/joc.1181

Murphy, J. M.; Sexton, D. M. H.; Barnett, D. N.; Jones, G. S.; Webb, M. J.; Collins, M. and Stainforth, D. A. (2004): Quantification of modelling uncertainties in a large ensemble of climate change simulations. In: Nature 430, 768–772. DOI: 10.1038/nature02771

Nakicenovic, N. and Swart, R. (eds.) (2000): Emission scenarios. Special Report of the Intergovernmental Panel on Climate Change. Cambridge.

North, G. R.; Cahalan, R. F. and Coakley, J. A. (1981): Energy balance models. In: Reviews of Geophysics and Space Physics 19, 91–121. DOI: 10.1029/RG019i001p00091

Paeth, H. (2011): Postprocessing of simulated precipitation for impact studies in West Africa – Part I: model output statistics for monthly data. In: Climate Dynamics 36, 1321–1336. DOI: 10.1007/s00382-010-0760-z

Paeth, H. and Diederich, M. (2011): Postprocessing of simulated precipitation for impact studies in West Africa – Part II: a weather generator for daily data. In: Climate Dynamics 36, 1337–1348. DOI: 10.1007/s00382-010-0840-0

Paeth, H. and Hense, A. (2002): Sensitivity of climate change signals deduced from multi-model Monte Carlo experiments. In: Climate Research 22, 189–204. DOI: 10.3354/cr022189

Paeth, H. and Pollinger, F. (2010): Enhanced evidence for changes in extratropical atmospheric circulation. In: Tellus 62A, 647–660. DOI: 10.1111/j.1600-0870.2010.00455.x

Paeth, H.; Rauthe, M. and Min, S. K. (2008): Multi-model Bayesian assessment of climate change in the northern annular mode. In: Global and Planetary Change 60, 193–206. DOI: 10.1016/j.gloplacha.2007.02.004

Paeth, H.; Hall, N. M. J.; Gaertner, M. A.; Dominguez Alonso, M.; Moumouni, S.; Polcher, J.; Ruti, P. M.; Fink, A. H.; Gosset, M.; Lebel, T.; Gaye, A. T.; Rowell, D. P.; Moufouma-Okia, W.; Jacob, D.; Rockel, B.; Giorgi, F. and Rummukainen, M. (2011a): Progress in regional downscaling of West African precipitation. In: Atmospheric Science Letters 12, 75–82. DOI: 10.1002/asl.306

Paeth, H.; Lindenberg, J.; Kschischo, M. and Hense, A. (2011b): How dynamical models can learn from the data – an example with a simplified ENSO model. In: Theoretical and Applied Climatology 104, 221–231. DOI: 10.1007/s00704-010-0333-4

Paeth, H.; Steger, C. and Merkenschlager, C. (2013): Climate change – it's all about probability. In: Erdkunde 67, 203–222. DOI: 10.3112/erdkunde.2013.03.01

Palmer, T. N. and Anderson, D. L. T. (1994): The prospects for seasonal forecasting – A review paper. In: Quarterly Journal

of the Royal Meteorological Society 120, 755–793. DOI: 10.1002/qj.49712051802

Palmer, T. N. and Williams, P. D. (2008): Introduction. Stochastic physics in climate modelling. In: Philosophical Transactions of the Royal Society A 366, 2421–2427. DOI: 10.1098/rsta.2008.0059

Piani, C.; Frame, D. J.; Stainforth, D. A. and Allen, M. R. (2005): Constraints on climate change from a multi-thousand member ensemble of simulations. In: Geophysical Research Letters 32, L23825. DOI: 10.1029/2005GL024452

Räisänen, J. and Palmer, T. N. (2001): A probability and decision-model analysis of a multimodel ensemble of climate change simulations. In: Journal of Climate 14, 3212–3226. DOI: 10.1175/1520-0442(2001)014<3212:APADMA>2.0.CO;2

Roe, G. and Baker, M. (2007): Why is climate sensitivity so unpredictable? In: Science 318, 629–632. DOI: 10.1126/science.1144735

Rougier, J. C. (2007): Probabilistic inference for future climate using an ensemble of climate model evaluations. In: Climatic Change 81, 247–264. DOI: 10.1007/s10584-006-9156-9

Schenk, N. J. and Lensink, S. M. (2007): Communicating uncertainty in the IPCC's greenhouse gas emission scenarios. In: Climatic Change 82, 293–308. DOI: 10.1007/s10584-006-9194-3

Schönwiese, C. D.; Walter, A. and Brinckmann, S. (2010): Statistical assessment of anthropogenic and natural global climate forcing. An update. In: Meteorologische Zeitschrift 19, 3–10. DOI: 10.1127/0941-2948/2010/0421

Sexton, D. M. H. and Murphy, J. M. (2012): Multivariate probabilistic projections using imperfect climate models part II: robustness of methodological choices and consequences for climate sensitivity. In: Climate Dynamics 38, 2543–2558. DOI: 10.1007/s00382-011-1209-8

Sexton, D. M. H.; Murphy, J. M.; Collins, M. and Webb, M. J. (2012): Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. In: Climate Dynamics 38, 2513–2542. DOI: 10.1007/s00382-011-1208-9

Smith, S. J.; Conception, E.; Andres, R. and Lurz, J. (2004): Historical sulfur dioxide emissions 1850–2000: methods and results. Pacific Northwest National Laboratory, PNNL-14537.

Stainforth, D. A.; Aina, T.; Christensen, C.; Collins, M.; Faull, N.; Frame, D. J.; Kettleborough, J. A.; Knight, S.; Martin, A.; Murphy, J. M.; Piani, C.; Sexton, D.; Smith, L. A.; Spicer, R. A.; Thorpe, A. J. and Allen, M. R. (2005): Uncertainty in predictions of the climate response to rising levels of greenhouse gases. In: Nature 433, 403–406. DOI: 10.1038/nature03301

Storch, H. von; Güss, S. and Heimann, M. (1999): Das Klimasystem und seine Modellierung. Berlin.

Stott, P. A. and Kettleborough, J. A. (2002): Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. In: Nature 416, 723–726. DOI: 10.1038/416723a

Stott, P. A.; Tett, S. F. B.; Jones, G. S.; Allen, M. R.; Mitchell, J. F. B. and Jenkins, G. J. (2000): External control of 20th century temperature by natural and anthropogenic forcings. In: Science 290, 2133–2137. DOI: 10.1126/science.290.5499.2133

Taylor, K. E.; Stouffer, R. J. and Meehl, G. A. (2012): An overview of CMIP5 and the experiment design. In: Bulletin of the American Meteorological Society 93, 485–498. DOI: 10.1175/bams-d-11-00094.1

Tebaldi, C.; Smith, R. L.; Nychka, D. and Mearns, L. O. (2005): Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. In: Journal of Climate 18, 1524–1540. DOI: 10.1175/JCLI3363.1

Thompson, D. W. J.; Wallace, J. M.; Kennedy, J. I. and Jones, P. D. (2010): An abrupt drop in Northern Hemisphere sea surface temperature around 1970. In: Nature 467, 444–447. DOI: 10.1038/nature09394

Tomassini, L.; Reichert, P.; Knutti, R.; Stocker, T. F. and Borsuk, M. E. (2007): Robust Bayesian uncertainty analysis of climate system properties using Markov Chain Monte Carlo methods. In: Journal of Climate 20, 1239–1254. DOI: 10.1175/JCLI4064.1

Wang, Y.-M.; Lean, J. L. and Sheeley, N. R. (2005): Modeling the sun's magnetic field and irradiance since 1713. In: The Astrophysical Journal 625, 522–538. DOI: 10.1086/429689

Wigley, T. M. L. and Raper, S. C. B. (2001): Interpretation of high projections for global-mean warming. In: Science 293, 451–454. DOI: 10.1126/science.1061604

Wilks, D. S. (2006): Statistical methods in the atmospheric sciences. Amsterdam.

Author

Prof. Dr. Heiko Paeth
Institute of Geography and Geology
University of Würzburg
Am Hubland
97074 Würzburg
Germany
heiko.paeth@uni-wuerzburg.de