Climate Projections for the 21st Century Using Random Sets*

E. KRIEGLER
Potsdam Institute of Climate Impact Research, Germany

H. HELD
Potsdam Institute of Climate Impact Research, Germany

Abstract

We apply random set theory to an analysis of future climate change. Bounds on cumulative probability are used to quantify uncertainties in natural and socio-economic factors that influence estimates of global mean temperature. We explore the link of random sets to lower envelopes of probability families bounded by cumulative probability intervals. By exploiting this link, a random set for a simple climate change model is constructed, and projected onto an estimate of global mean warming in the 21st century. Results show that warming estimates on this basis can generate very imprecise uncertainty models.

Keywords

climate change, climate sensitivity, imprecise probability, random sets, belief functions

1 Introduction

It is widely acceped by now that a discernible influence of anthropogenic emissions of greenhouse gases (GHGs) on the earth's climate exists. Greenhouse gas concentrations in the atmosphere have risen by, to name just a few, 30% (carbon dioxide), 250% (methane) and 15% (nitrous oxide) in the industrial era since 1750, mainly due to human activity. Empirical evidence for a growing climate change signal is mounting, and nearly all climate models need the increased radiative forcing due to growing GHG concentrations to reproduce this signal. Still, uncertainty abounds. How sensitive is the climate to growing GHG concentrations? What amount of greenhouse gases will humankind put into the atmosphere in the 21st century?

*This work has been supported in part by the Deutsche Bundesstiftung Umwelt (German Federal Foundation of the Environment).

We believe that the application of imprecise probability concepts carries the potential to greatly improve the situation in climate change forecasting and integrated assessment of climate change policies. However, an obstacle might be the dynamical nature of climate change models, and the large number of uncertain variables which mostly range over continuous possibility spaces. In this paper, we present an application of random set methods to the estimation of global mean temperature (GMT) change in the 21st century. We interprete the corresponding belief functions as a lower envelope of a set of probability measures, and try to respect this interpretation throughout the reasoning process. The uncertain model parameters are initially quantified by lower and upper cumulative probability distribution functions (CDF) on the real line. In section 2, we discuss how this information can be converted into a random set, combined for independent model parameters, and projected onto the model output. In section 3, we present the simple temperature change model, and construct a random set for its uncertain parameters. In section 4, the uncertainty in the input values is projected onto an estimate of global mean temperature change.

2 Methods

2.1 Random Sets of Imprecise CDF Models

Consider an uncertain quantity X that enters a model of some causal relationship, e.g. of the link between GHG emissions and GMT. The imprecise uncertainty about X shall be described by a lower bound $\underline{F}_X : \mathbb{R} \to [0,1]$ and an upper bound $\overline{F}_X : \mathbb{R} \to [0,1]$ for a set of CDFs $F_X(x) := P(X \leq x)$ on the real line $\mathbb{R}$. In the following, such an uncertainty assessment will be called an *imprecise CDF model*

$$\mathcal{M}_X(\underline{F},\overline{F}) := \{\, P \,|\, \forall\, x \in \mathbb{R} \;\; \underline{F}(x) \leq P(-\infty,x] \leq \overline{F}(x) \,\} \tag{1}$$

A monotone set function $\underline{P} : \mathcal{R} \to [0,1]$, $\underline{P}(\emptyset) = 0$, $\underline{P}(\mathbb{R}) = 1$ is a *lower envelope* or *coherent lower probability* on the Borel algebra $\mathcal{R}$ of the real line, if it defines a non-empty set of countably additive probability measures $\mathcal{M}(\underline{P}) := \{P \,|\, \forall\, A \in \mathcal{R} \;\; \underline{P}(A) \leq P(A)\}$, and $\forall\, A \in \mathcal{R} \;\; \underline{P}(A) = \inf_{P \in \mathcal{M}(\underline{P})} P(A)$ [13, theorem 3.3.3]. An ∞-monotone lower envelope is a *belief function* Bel.

In the theory of Dempster [4], belief functions are generated by a multi-valued mapping from an underlying space $\Psi = \{\psi_1, ..., \psi_n\}$ onto a field of sets, in our case the Borel algebra $\mathcal{R}$. By means of the multi-valued mapping, a *probability mass assignment* m on Ψ can be transferred to $\mathcal{R}$, i.e. there exists $m : \mathcal{R} \to [0,1]$, with $m(A) > 0$ for only a finite number of sets $\mathcal{F} = \{E_1, ..., E_n\} \subset \mathcal{R}$ and $\sum_{A \in \mathcal{R}} m(A) = 1$. The pair $(\mathcal{F}, m)$ is called a (finite support) *random set*, and the sets $E_i \in \mathcal{F}$ *focal elements*. A belief function *Bel* and its conjugate *plausibility*

function Pl are connected to a random set by [4, 11]

$$Bel(A) = \sum_{B \subseteq A} m(B) \; = \sum_{i \,|\, E_i \subseteq A} m_i \,, \qquad Pl(A) = \sum_{B \cap A \neq \emptyset} m(A) \; = \sum_{i \,|\, E_i \cap A \neq \emptyset} m_i$$

Thus, knowledge of the random set $(\mathcal{F}, m)$ suffices to determine *Bel* and *Pl* on $\mathcal{R}$.

We explore the relationship between the lower envelope of an imprecise CDF model and a belief function that can be represented by a finite support random set (In the following, the reference to the finiteness of the random set will be omitted). The goal is to capture the information content of an imprecise CDF model with a random set.

Proposition 1 *Let $\mathcal{M}_X(\underline{F}, \overline{F})$ be an imprecise CDF model as defined in (1). Let $\mathcal{A}$ be the algebra generated by the set of half-closed intervals $(a,b]$, $a < b$ of the real line $\mathbb{R}$. Let $(\mathcal{F}, m)$ be a random set, and $Bel_{\mathcal{F}}$, $Pl_{\mathcal{F}}$ the corresponding belief and plausibility functions, respectively.*

If (I) $(\mathcal{F}, m)$ contains only closed intervals $E_i = [\underline{x}_i, \overline{x}_i]$,
(II) $(\mathcal{F}, m)$ includes no pair of focal elements E_i, E_j with $\underline{x}_i < \underline{x}_j < \overline{x}_j < \overline{x}_i$, and
(III) $\forall\, x \in \mathbb{R}$ $Bel_{\mathcal{F}}(-\infty, x] = \underline{F}(x)$, $Pl_{\mathcal{F}}(-\infty, x] = \overline{F}(x)$,

$$\textit{then} \qquad \forall\, A \in \mathcal{A} \quad Bel_{\mathcal{F}}(A) = \underline{P}_X(A) := \inf_{P \in \mathcal{M}_X(\underline{F}, \overline{F})} P(A)$$

Proof. **Step 1**: Consider an arbitrary $(a,b] \in \mathcal{A}$, $a < b$. We have to show $\underline{P}_X(a,b] = Bel_{\mathcal{F}}(a,b]$ and $\overline{P}_X(a,b] = Pl_{\mathcal{F}}(a,b]$. Since $\underline{P}_X(A) = 1 - \overline{P}_X(A^c)$ and $Bel_{\mathcal{F}}(A) = 1 - Pl_{\mathcal{F}}(A^c)$, this implies that the equalities hold for the complement $(a,b]^c$ as well.

$$\begin{aligned} 1a) \quad \overline{P}_X(a,b] &= \overline{F}(b) - \underline{F}(a) \; = \sum_{i \,|\, E_i \cap (-\infty, b] \neq \emptyset} m_i - \sum_{j \,|\, E_j \subseteq (-\infty, a]} m_j \\ &= \sum_{s(i) \,|\, E_{s(i)} \subseteq (-\infty, a]} m_{s(i)} + \sum_{t(i) \,|\, E_{t(i)} \cap (a,b] \neq \emptyset} m_{t(i)} - \sum_{j \,|\, E_j \subseteq (-\infty, a]} m_j \\ &= Pl_{\mathcal{F}}(a,b] \end{aligned}$$

$$1b) \quad \underline{P}_X(a,b] = \max[0, \underline{F}(b) - \overline{F}(a)] = \max[0, \sum_{i \,|\, E_i \subseteq (-\infty, b]} m_i - \sum_{j \,|\, E_j \cap (-\infty, a] \neq \emptyset} m_j]$$

If $\underline{F}(b) < \overline{F}(a)$, there exists $E_* = [\underline{x}_*, \overline{x}_*] \in \mathcal{F}$ with $E_* \cap (-\infty, a] \neq \emptyset$ and $E_* \not\subseteq (-\infty, b]$. Assume an arbitrary $E' = [\underline{x}', \overline{x}'] \in \mathcal{F}$ with $\underline{x}' > a \geq \underline{x}_*$. By condition (II), $\overline{x}' \geq \overline{x}_* > b$. Thus, $E' \not\subseteq (a,b]$, and $Bel_{\mathcal{F}}(a,b] = 0$.

Assume there exists $E_* \in \mathcal{F}$ with $E_* \cap (-\infty, a] \neq \emptyset$ and $E^* \not\subseteq -(\infty, b]$. By condition (I)+(II), all $E_i \subseteq (-\infty, b] \in \mathcal{F}$ intersect $(-\infty, a]$, and $\underline{F}(b) < \overline{F}(a)$.

Thus, if $\underline{F}(b) \geq \overline{F}(a)$, there is no such focal element $E_* \in \mathcal{F}$. In other words, $\forall\, E_i \in \mathcal{F}\ \ E_i \not\subseteq (-\infty, b] \Rightarrow E_i \cap (-\infty, a] = \emptyset$.

$$\begin{aligned} \Rightarrow \underline{P}_X(a,b] &= \sum_{s(i)\,|\,E_{s(i)} \subseteq (a,b]} m_{s(i)} + \sum_{t(i)\,|\,E_{t(i)} \cap (-\infty,a] \neq \emptyset} m_{t(i)} - \sum_{j\,|\,E_j \cap (-\infty,a] \neq \emptyset} m_j \\ &= Bel_{\mathcal{F}}(a,b] \end{aligned}$$

Step 2: Consider an arbitrary union of k disjoint half-closed intervals $A_k = (a_1, b_1] \cup \ldots \cup (a_k, b_k]$, $a_1 < b_1 < \ldots < a_k < b_k$.

Choose a CDF $F^* : \mathbb{R} \to [0,1]$ with $F^*(a_1) = \min[\overline{F}(a_1), \underline{F}(b_1)]$, $F^*(b_1) = \underline{F}(b_1)$, ..., $F^*(a_k) = \min[\overline{F}(a_k), \underline{F}(b_k)]$, $F^*(b_k) = \underline{F}(b_k)$. Since $F^*(a_1) \leq F^*(b_1) \leq \ldots \leq F^*(a_k) \leq F^*(b_k)$, such a CDF does exist, and is contained in $\mathcal{M}_X(\underline{F}, \overline{F})$.

$$\begin{aligned} P^*(A_k) &= F^*(b_k) - F^*(a_k) + \ldots + F^*(b_1) - F^*(a_1) \\ &= \max[0, \underline{F}(b_k) - \overline{F}(a_k)] + \ldots + \max[0, \underline{F}(b_1) - \overline{F}(a_1)] \\ &= \underline{P}_X(a_k, b_k] + \ldots + \underline{P}_X(a_1, b_1] \end{aligned}$$

Since the lower envelope $\underline{P}_X$ is super-additive on a union of disjoint sets [13, Ch. 2.7.4], $\underline{P}_X(A_k) = P^*(A_k)$. Thus, $\underline{P}_X(\bigcup_{l=1}^{k} (a_l, b_l]) = \sum_{l=1}^{k} \underline{P}_X(a_l, b_l]$. Since $\underline{P}_X(a_l, b_l] = Bel(a_l, b_l]$ as shown in step 1:

$$2a) \quad \underline{P}_X(A_k) = \sum_{l=1}^{k} \sum_{i|E_i \subseteq (a_l,b_l]} m_i = \sum_{i|E_i \subseteq \bigcup_{l=1}^{k} (a_l,b_l]} m_i = Bel_{\mathcal{F}}(A_k)$$

$$\begin{aligned} 2b) \quad \overline{P}_X(A_k) &= \overline{P}_X(-\infty, b_k] - \underline{P}_X((-\infty, a_1] \cup (b_1, a_2] \cup \ldots \cup (b_{k-1}, a_k]) \\ &= \overline{F}(b_k) - \underline{F}(a_1] - \underline{P}_X(b_1, a_2] - \ldots - \underline{P}_X(b_{k-1}, a_k] \\ &= \sum_{i|E_i \cap (a_1,b_k] \neq \emptyset} m_i - \sum_{j|E_j \subseteq \bigcup_{l=1}^{k-1} (b_l, a_{l+1}]} m_j = Pl_{\mathcal{F}}(A_k) \end{aligned}$$

Every element of $\mathcal{A}$ is either $\emptyset$, $\mathbb{R}$, a union of $k \in \mathbb{N}$ disjoint half-closed intervals, or its complement. For the latter, $\underline{P}_X(A) = Bel_{\mathcal{F}}(A)$ has been shown in step 1 and 2. For $\emptyset, \mathbb{R}$, $\underline{P}_X(\emptyset) = Bel_{\mathcal{F}}(\emptyset) = 0$ and $\underline{P}_X(\mathbb{R}) = Bel_{\mathcal{F}}(\mathbb{R}) = 1$. □

Since the random set $(\mathcal{F}, m)$ contains only a finite number of focal elements, its corresponding belief and plausibility function cannot fulfil condition (III) of proposition 1 for continuous $\underline{F}$ and/or $\overline{F}$. For application purposes, however, this defect is not disturbing. Every imprecise CDF model with continuous lower and upper bound can be approximated by two step functions approaching the lower

bound from below and the upper bound from above [7, 12]. Consider two step functions $SF_*, SF^* : \mathbb{R} \to [0,1]$ of the form $0 = SF(x_1) < ... < SF(x_k) = 1$,

$$SF_*(x) = \begin{cases} SF_*(x_{*i}) & x_{*i} \le x < x_{*i+1} \\ 0 & x < x_{*1} \\ SF_*(x_{*k}) & x_{*k} \le x \end{cases} \qquad SF^*(x) = \begin{cases} SF^*(x^*_{j+1}) & x^*_j < x \le x^*_{j+1} \\ 0 & x \le x^*_1 \\ SF^*(x^*_k) & x^*_{k'} < x \end{cases}$$

If $\forall\, x \in \mathbb{R} \quad SF^*(x) \ge SF_*(x)$, the two step functions define an imprecise CDF model $\mathcal{M}(SF_*, SF^*) := \{P \,|\, \forall\, x \in \mathbb{R} \quad SF_*(x) \le P(-\infty, x] \le SF^*(x)\}$. The following algorithm can be used to construct a random set $(\mathcal{F}, m)$, which fulfils the requirements of proposition 1, from two arbitrary $SF_* \le SF^*$. Let the lower bound have cumulative probability $SF_*(x_{*i})$ at the "step" points $x_{*1} < ... < x_{*n}$, and the upper bound have cumulative probability $SF^*(x^*_j)$ at $x^*_1 < ... < x^*_m$.

Algorithm 1 *1. Initialize indices $k = 1$ (running over the focal elements of the random set to be constructed), $i = 1$ (running over x_{*i}), $j = 1$ (running over x^*_j). Let p_k denote the cumulative probability already accounted for in step k. Assign $p_0 = 0$.*

2. Construct random set $E_k = [x^*_j, x_{*i}]$.

3. (a) $SF_*(x_{*i}) < SF^*(x^*_j)$: $m_k = SF_*(x_{*i}) - p_{k-1}$, $p_k = SF_*(x_{*i})$. Raise indices $k \to k+1$, $i \to i+1$. Return to step 2.

(b) $SF_*(x_{*i}) > SF^*(x^*_j)$: $m_k = SF^*(x^*_j) - p_{k-1}$, $p_k = SF^*(x^*_j)$. Raise indices $k \to k+1$, $j \to j+1$. Return to step 2.

(c) $SF_*(x_{*i}) = SF^*(x^*_j)$: $m_k = SF^*(x^*_j) - p_{k-1}$. If $SF_*(x_{*i}) = SF^*(x^*_j) = 1$ abort the algorithm.
If $SF_*(x_{*i}) = SF^*(x^*_j) < 1$, set $p_k = SF^*(x^*_j)$. Raise indices $k \to k+1$, $i \to i+1$, $j \to j+1$. Return to step 2.

Algorithm 1 is well defined. For each step k, $x^*_j \le x_{*i}$, $m_k > 0$, and the algorithm will always reach the points x_{*n}, x^*_m with $SF_*(x_{*n}) = SF^*(x^*_m) = 1$ and abort. It constructs a random set $(\mathcal{F}, m)$ with $k \le n + m$ focal elements. The E_k are either closed intervals $[a_k, b_k]$ or singletons $\{a\} = [a_k, a_k]$. The algorithm is also applicable to the case of a precise probability, where $SF_* = SF^* = SF$.

2.2 Combining and Extending Random Sets

In almost all assessments of climate change, uncertainty accumulates from different sources. In general, we need to consider a multivariate uncertainty model that arises from a vector of uncertain quantities $X = \{X_1, ..., X_n\}$, each of which is described by an imprecise CDF model $\mathcal{M}_{X_i}(\underline{F}, \overline{F})$ on the real line $\mathbb{R}$. There are different ways to construct a joint lower envelope $\underline{P}_X$ from the lower envelopes of independent marginals $\underline{P}_{X_i}$. They depend on the concept of independence that is

employed to generate the joint envelope [2, 13]. In general, the resulting envelopes agree only on product sets $A_1 \times ... \times A_n$, $A_i \subseteq \mathbb{R}$.

In our case, the lower envelopes $\underline{P}_{X_i}$ of the independent marginals are represented by belief functions $Bel_{\mathcal{F}_i}$ with corresponding random sets $(\mathcal{F}_i, m_i) = \{(E_{1_i}, m_{1_i}), ..., (E_{k_i}, m_{k_i})\}$. The concept of *random set independence* [4] leads to joint belief functions by applying Dempster's rule of combination to logically independent "marginal" random sets $(\mathcal{F}_i, m_i)$, $1 \leq i \leq n$.

$$(\mathcal{F}, m) = \{ (E_{l_1 \ldots l_n} = E_{l_1} \times ... \times E_{l_n}, m_{l_1 \ldots l_n} = m_{l_1} \cdot ... \cdot m_{l_n}),\ 1 \leq l_i \leq k_i \} \qquad (2)$$

It can be easily checked that $(\mathcal{F}, m)$ generates indeed a belief and plausibility function $Bel_{\mathcal{F}}$ and $Pl_{\mathcal{F}}$ that agree with the joint lower and upper envelopes $\underline{P}_X$ and $\overline{P}_X$ on product sets, no matter under which concept of independence they were generated. However, it is less clear, how $Bel_{\mathcal{F}}$ relates to the different types of the joint lower envelope on sets $A \in \mathcal{R}^n$ that are not product sets. Comparisons of different independence concepts on finite possibility spaces indicate that random set independence yields a lower envelope that is dominated by the envelopes emanating from epistemic or strong independence [2]. It needs to be further investigated how far these findings translate to the special case presented here. For the time being, we use random set independence to construct the joint lower envelope $Bel_{\mathcal{F}}$ from the independent marginals $Bel_{\mathcal{F}_i}$.

Consider a model of some causal relationship, which generates a transfer function $f : \mathbb{R}^n \to \mathbb{R}^m$, $y = f(x)$. Let the uncertainty in the input variables x be described by $\mathcal{M}_X(Bel_{\mathcal{F}}) := \{ P_X \,|\, \forall A \in \mathcal{R}^n \;\; Bel_{\mathcal{F}}(A) \leq P_X(A) \}$. The corresponding random set $(\mathcal{F}, m) = \{(E_1, m_1), ..., (E_k, m_k)\}$ can be transferred to the model output y by applying the extension principle for random set-valued variables [5]:

$$f(E_i) := \{ y \,|\, \exists\, x \in E_i \;\; y = f(x) \}\,, \qquad m_f(B) := \sum_{f(E_i)=B} m_i \quad B \in \mathbb{R}^m \qquad (3)$$

Let $(f(\mathcal{F}), m_f)$ denote the transferred random set. It corresponds to a belief function $Bel_{f(\mathcal{F})}$ that is the lower envelope of a set of probabilities $\mathcal{M}_Y(Bel_{f(\mathcal{F})})$. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Borel measurable, i.e. $\forall\, B \in \mathcal{R}^m \quad f^{-1}(B) = \{ x \in \mathbb{R}^m : f(x) \in B \} \in \mathcal{R}^n$. Then, every probability measure P on $(\mathbb{R}^n, \mathcal{R}^n)$ is transformed by the mapping f into a probability measure P_f on $(\mathbb{R}^m, \mathcal{R}^m)$ defined by $\forall\, B \in \mathcal{R}^m \;\; P_f(B) := P(f^{-1}(B))$. Using this definition, we can transform each element of $\mathcal{M}_X(Bel_{\mathcal{F}})$ to a probability measure on $(\mathbb{R}^m, \mathcal{R}^m)$, thus generating:

$$f(\mathcal{M}_X(Bel_{\mathcal{F}})) := \{ P_Y \,|\, \exists\, P_X \in \mathcal{M}_X(Bel_{\mathcal{F}}) \;\; \forall\, B \in \mathcal{R}^m \quad P_Y(B) = P_X(f^{-1}(B)) \}$$

Proposition 2 *Let $\mathcal{R}^n$, $\mathcal{R}^m$ be Borel algebras, $f : \mathbb{R}^n \to \mathbb{R}^m$ a Borel measurable transfer function. Let $(\mathcal{F}, m)$, $Bel_{\mathcal{F}}$ describe the set of probabilities $\mathcal{M}_X(Bel_{\mathcal{F}})$. Let $f(\mathcal{M}_X(Bel_{\mathcal{F}}))$ be the f-tranformed set of probabilities as defined above.*

Similarly, let $(f(\mathcal{F}), m_f)$ be the f-extension of $(\mathcal{F}, m)$ calculated from equation (3), and $Bel_{f(\mathcal{F})}$ the corresponding belief function. Then

$$f(\mathcal{M}_X(Bel_{\mathcal{F}})) \subseteq \mathcal{M}_Y(Bel_{f(\mathcal{F})}) := \{ P_Y \mid \forall B \in \mathcal{R}^m \quad Bel_{f(\mathcal{F})}(B) \leq P_Y(B) \}$$

Proof. Consider an arbitrary $P_Y \in f(\mathcal{M}_X(Bel_{\mathcal{F}}))$. There exists a $P_X \in \mathcal{M}_X(Bel_{\mathcal{F}})$ with $\forall B \in \mathcal{R}^m \;\; P_Y(B) = P_X(f^{-1}(B))$. For a particular, yet arbitrary $B \in \mathcal{R}^m$

$$\begin{aligned} P_Y(B) &= P_X(f^{-1}(B)) \geq Bel_{\mathcal{F}}(f^{-1}(B)) = \sum_{E_i \subseteq f^{-1}(B)} m_i \\ &= \sum_{f(E_i) \subseteq B} m_i = Bel_{f(\mathcal{F})}(B) \end{aligned}$$

□

3 A Random Set for a Simple Climate Model

3.1 Global Mean Temperature Model

We use a simple dynamical model to link radiative forcing $F(t)$ to a change ΔT in global mean temperature (GMT) since preindustrial times [14].

$$C_e \cdot \Delta T'(t) = F(t) - F_{2x} \cdot \frac{\Delta T(t)}{T_{2x}} \tag{4}$$

C_e effective ocean heat capacity

F_{2x} radiative forcing for a doubling of atmospheric CO_2

T_{2x} climate sensitivity

Differential equation (4) is the simplest type of energy balance model. It equates the net radiative flux into the system at the top of the atmosphere to oceanic heat uptake $C_e \Delta T'$. If the radiative forcing was kept constant at a value $F(t) = F_{2x}$, the system would undergo an equilibrium temperature change of $\Delta T = T_{2x}$. Climate sensitivity T_{2x} is a crucial parameter to characterize the response of the climate system to an increase in GHG concentrations.

The Intergovernmental Panel on Climate Change (IPCC) gives an estimate of climate sensitivity $T_{2x} = [1.5\text{ K}, 4.5\text{ K}]$ [3]. The panel explicitely refrains from specifying probabilistic information. Recently, models of intermediate complexity (EMICs) were used to establish probability distributions from a comparison of model results with historical atmosphere, surface and deep ocean temperature data [1, 6, 8]. Efforts are hampered by the presence of natural variability, the lack of long-term data and the multitude of forcings.

In this analysis, we use the probability distributions of [1, 6] to generate an imprecise CDF model for T_{2x} (fig. 1). The estimates of [1] are shifted to considerably higher values of climate sensitivity compared to [6], ranging up to values of

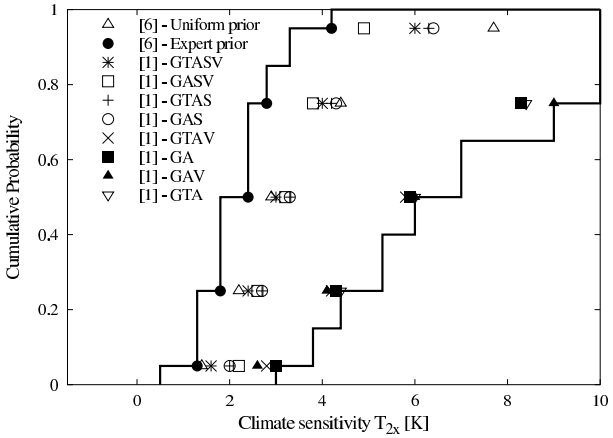


Figure 1: Imprecise CDF model for T_{2x}: Shown are 5%, 25%, 50%, 75% and 95% quantiles of probability distributions from [1, 6]. Estimates of [6] depend on a prior probability for T_{2x}. Estimates of [1] depend on whether solar forcing (S), volcanic aerosol forcing (V) and tropospheric ozone (T) was added to greenhouse gas (G) and aerosol forcing (A). The capital letters G, A, T, S, V in the figure key specify the radiative forcing components that were considered for the particular estimate of [1].

$T_{2x} = 22$ K. One reason could be that [1] does not compare their results with deep ocean temperature data. [6] requires the ocean record to restrict T_{2x} from above. However, [8] considers ocean heat uptake, and fails to discriminate between climate sensitivity in the range $T_{2x} = [1\text{ K}, 10\text{ K}]$. In this situation, we simply cut of the probability distributions of [1] at $T_{2x} = 10$ K, and allocate their total probability mass $P(T_{2\times} \geq 10\text{ K})$ to this value.

Fig. 1 depicts the resulting ranges for 5%, 25%, 50%, 75% and 95% quantile estimates in [1, 6]. We interpolate the extreme values of the ranges to generate a lower and upper CDF, and approximate the resulting imprecise CDF model with two step functions SF_* and SF^* (Fig. 1). There is some arbitrariness here. It could be resolved by fixing the number of "step" points $T_{2x,i*}$ and $T^*_{2x,j}$, and calculating the optimal approximation according to some accuracy measure [7, 12]. Algorithm 1 is applied to construct a random set $(\mathcal{F}_{T_{2x}}, m_{T_{2x}})$ that corresponds to $\mathcal{M}_{T_{2x}}(SF_*, SF^*) := \{\, P \,|\, \forall\, T_{2x} \in \mathbb{R} \quad SF_*(T_{2x}) \leq P(-\infty, T_{2x}] \leq SF^*(T_{2x}) \,\}$ (in the sense of proposition 1). $\mathcal{M}_{T_{2x}}(SF_*, SF^*)$ can be compared with the IPCC estimate [1.5 K, 4.5 K] for climate sensitivity. The probability for $T_{2x} \in$ [1.5 K, 4.5 K] lies in the interval $[0, 1]$, for $T_{2x} < 1.5K$ in $[0, 0.25]$, and for $T_{2x} > 4.5$ K in $[0, 0.75]$. The numbers show that $\mathcal{M}_{T_{2x}}(SF_*, SF^*)$ does not support the IPCC estimate, especially for high climate sensitivities $T_{2x} > 4.5$ K. This reflects the fact that the upper bound of the IPCC estimate is not supported by [1, 6, 8].

Effective ocean heat capacity C_e is an artificial quantity that arises from the simple form of the energy balance model (4). It depends on ocean characteristics,

but also on climate sensitivity [3]. A comparison of model (4) with emulations of different AOGCMs suggest a functional dependence of C_e on T_{2x} of the form $C_e \sim T_{2x}^{\gamma_c}$ with $0 < \gamma_c \leq 1$. We specify an interval uncertainty for the parameters $\bar{C} = C_e(T_{2x} = 3\text{ K})$ and γ_c, which is an adequate choice in the light of the large uncertainty surrounding ocean characteristics like vertical diffusivity [6]. Interval uncertainty is the simplest form of an imprecise CDF model. Lower and upper CDF are either 0 or 1. The model can be immediately captured by a random set $(\mathcal{F}_{\bar{C},\gamma_c}, m_{\bar{C},\gamma_c})$ containing just one focal element $E = [40\text{ Wa/m}^2\text{K}, 50\text{ Wa/m}^2\text{K}] \times [0.6,1]$ with probability mass assignment $m(E) = 1$.

An additional uncertainty concerns the present day global mean warming ΔT_o since 1860, which enters model (4) as initial value. Estimates of ΔT_o lie in the range 0.6 ± 0.2 K. We adopt the interval uncertainty [0.4 K, 0.8 K] for ΔT_o, since its small influence on future GMT projections does not justify a more complicated imprecise CDF model.

3.2 Radiative Forcing Model

We group the anthropogenic sources of radiative forcing $F(t)$ into carbon dioxide, which is the most important GHG, the "other" greenhouse gases (OGHG) including both the remaining direct as well as indirect GHGs, and aerosols. Solar and volcanic sources are neglected since we are interested in estimating the anthropogenic climate change signal.

$$F(t) = F_{2x} \ln\left(\frac{C_{CO_2}(t)}{C_{CO_2}(1750)}\right) / \ln 2 + F_{Aer}\, g(E_{Aer}(t)) + F_{OGHG}\, h(t) \quad (5)$$

C_{CO_2} atmospheric CO_2 concentration

E_{Aer} anthropogenic sulfate aerosol emissions

F_{Aer} Total aerosol forcing in the period 1990-2000

F_{OGHG} Total OGHG forcing in the period 1990-2000

The radiative properties of aerosol particles are most uncertain. Aerosols influence the radiation balance not only directly, but also indirectly by altering cloud formation processes. The IPCC estimates that the negative forcing of aerosols has been in the range [-0.8 W/m^2, -0.2 W/m^2] (direct effect) and [-2 W/m^2, 0 W/m^2] (indirect effect) for the period 1990-2000 [10]. [1, 6, 8] have investigated F_{Aer} in their comparison of model results with historical data. Fig. 2 shows the ranges for the 5%, 25%, 50%, 75% and 95% quantile estimates from [1, 6]. [8] presents a histogram probability which can be converted into two step functions for the lower and upper bound on the CDFs that are supported by the probability masses allocated to the bins of the histogram. Analogous to the case of climate sensitivity, we construct a lower CDF SF_* and upper CDF SF^* (solid lines in fig. 2). Algorithm 1 is used to generate the random set $(\mathcal{M}_{F_{Aer}}, m_{F_{Aer}})$ that corresponds to the imprecise CDF model $\mathcal{M}_{F_{Aer}}(SF_*, SF^*)$.

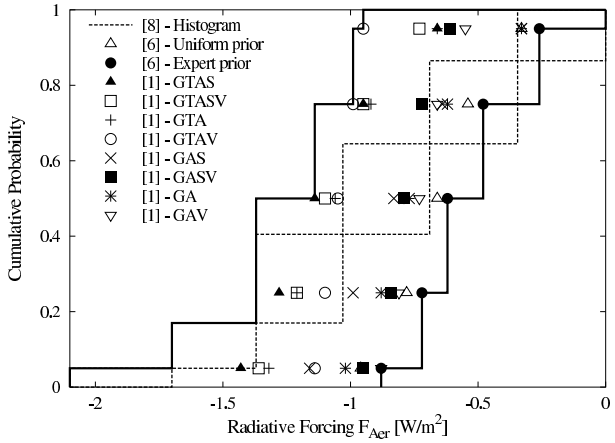


Figure 2: Imprecise CDF model for F_{Aer}: Shown are 5%, 25%, 50%, 75% and 95% quantiles of probability distributions from [1, 6], and a histogram probability from [8]. See Fig. 1 for additional explanation of the figure key.

The probability that F_{Aer} is contained in the IPCC estimate [-2.8 K, -0.2 K] (direct and indirect effect combined) lies in the range $[0.95, 1]$. In contrast to climate sensitivity, the IPCC range includes $\mathcal{M}_{F_{Aer}}(SF_*, SF^*)$ almost entirely. The results in [1, 6, 8] support a more narrow range, where in particular the potential of a very strong negative aerosol forcing contribution is discarded.

Estimates for the radiative forcing contributions of indirect GHGs, in particular troposheric and stratospheric ozone, exhibit relative errors between 40%-70%. The indirect GHGs have contributed around 30-40% to F_{OGHG} in the last decade. We capture the uncertainty by the interval $F_{OGHG} \in [0.8\ \mathrm{W/m^2}, 1.2\ \mathrm{W/m^2}]$.

We link the uncertainty in the time-dependent paths of atmospheric CO_2 concentration $C_{CO_2}(t)$, future changes in the radiative forcing of the OGHG $h(t)$, and anthropogenic aerosol emissions $E_{Aer}(t)$ directly to the socio-economic sphere. Thereby, we neglect any uncertainty about the response of the biogeochemical cycles to anthropogenic emissions. In a special report on emissions scenarios (SRES) [9], the IPCC has formulated a range of scenarios describing future pathways of society and economy on a global scale. The major branching points of these scenarios are globalization vs. regionalization and sustainability orientation vs. growth orientation. In this analysis, we specify just two parameters G ("Growth") and S ("Shift"), with $C_{CO_2}(t)$, $h(t) \sim e^{Gt-St^2}$. We restrict $S \leq G/200$, so that the growth in atmospheric CO_2 concentration and radiative forcing of OGHGs can be dampened, but not reversed by a "shift" S in the 21st century.

As the future socio-economic development is entirely uncertain, it is appropriate to specify interval uncertainties for $G \in [0.004\,/a, 0.012\,/a]$ and $S \in [0, G/200]$.

Growth rates from 0.4% to 1.2% per year lead to atmospheric CO_2 concentrations from 480 ppmv to 1230 ppmv in 2100 (present day: 370 ppmv), and to a forcing contribution of the OGHG from 1 W/m^2 to 4 W/m^2. This covers the full range of the SRES scenarios including uncertainty in the biogeochemical cycles [3].

3.3 Combining the Random Set Information

Most parameter pairs are physically and epistemically independent. Present day warming T_o depends physically on climate sensitivity and ocean heat capacity, but knowledge of T_o alone does not constrain the assessment of T_{2x} and C_e. A more critical issue is the epistemic dependence of F_{Aer} and T_{2x}. Although physically independent, comparisons of model results with historical data will have a tendency to produce high estimates of T_{2x} for a large negative radiative forcing F_{Aer} of aerosols, and vice versa [6]. Neglecting this dependence will yield a more imprecise estimate of future GMT change, since the probability weight of combinations with large negative F_{Aer} and low T_{2x} leading to a weak GMT increase, and with small negative F_{Aer} and high T_{2x} leading to a strong rise of GMT, will be overestimated. This issue needs to be investigated in further studies. For the time being, we use equation (2) based on random set independence to combine the random sets for all eight parameters $par := (\Delta T_0, T_{2x}, \bar{C}, \gamma_c, F_{Aer}, F_{OGHG}, G, S)$ to a joint random set $(\mathcal{F}_{par}, m)$.

4 Estimation of Global Mean Temperature Change

Differential equation (4) and radiative forcing model (5) generate a continuous transfer function that maps the uncertain model parameters to an increase ΔT in GMT since 1860. The extension principle for random sets ([5], equation 3) transfers the random set $(\mathcal{F}_{par}, m)$ for the uncertain parameters to a random set $(\mathcal{F}_{\Delta T}, m)$ for GMT increase. In our specific case, the images $f(E_{i,par}) = [\underline{\Delta T}_i(t), \overline{\Delta T}_i(t)]$ can be calculated with standard gradient-based optimization methods. After discretizing time in sufficiently small time steps Δt, the boundaries of the range at time $t_k = k\Delta t + t_o$ are found by solving

$$\underline{\Delta T}_i(t_k) = \min_{(\Delta T_0, T_{2x}, \bar{C}, \gamma_c, F_{Aer}, F_{OGHG}, G, S) \in E_{i,par}} \Delta T(t_k) \tag{6}$$

$$\text{subject to} \quad \Delta T(t_l) = \Delta T(t_{l-1}) + \Delta t \cdot \left(\frac{F(t_{l-1})}{C_e} - \frac{F_{2x}}{C_e} \cdot \frac{\Delta T(t_{l-1})}{T_{2x}} \right) \quad 1 \le l \le k$$

$$\overline{\Delta T}_i(t_k) = \max_{(\Delta T_0, T_{2x}, \bar{C}, \gamma_c, F_{Aer}, F_{OGHG}, G, S) \in E_{i,par}} \Delta T(t_k) \tag{7}$$

$$\text{subject to} \quad \Delta T(t_l) = \Delta T(t_{l-1}) + \Delta t \cdot \left(\frac{F(t_{l-1})}{C_e} - \frac{F_{2x}}{C_e} \cdot \frac{\Delta T(t_{l-1})}{T_{2x}} \right) \quad 1 \le l \le k$$

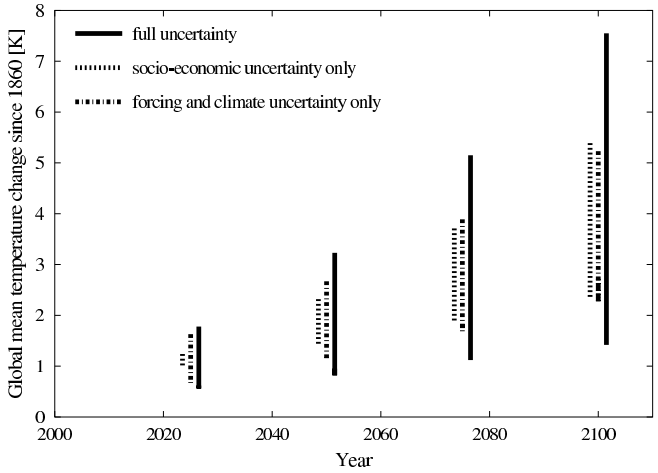
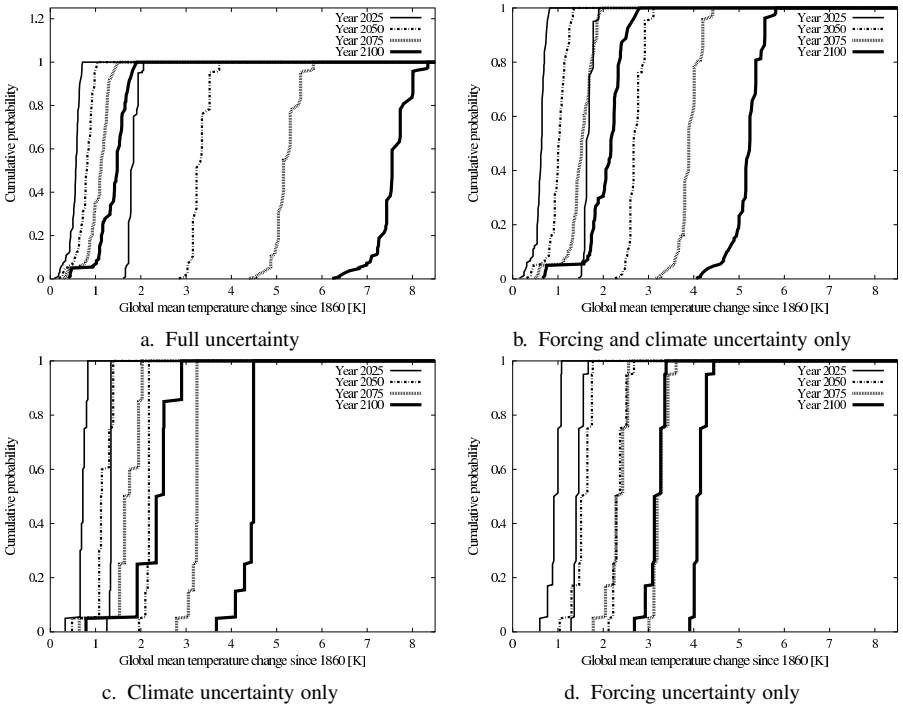


Figure 3: Image $[\underline{\Delta T}(t), \overline{\Delta T}(t)]$ of a single focal element $E^* =$ [1.8 K, 6.0 K] $\times$ [-1.37 W/m^2,-0.62 W/m^2] $\times(\Delta T_o, \bar{C}, \gamma_c, F_{OGHG}, G, S) \in \mathcal{F}_{par}$ for the years 2025, 2050, 2075 and 2100. Shown are also the cases with solely socio-economic or solely forcing and climate uncertainty.

It can be checked that $\Delta T(t)$ is monotone in $\Delta T_o, \bar{C}, \gamma_c, F_{Aer}, F_{OGHG}, G, S$ and convex in T_{2x}. The latter is due to the fact that T_{2x} influences ΔT both directly and indirectly through its connection to effective ocean heat capacity. Thus, program (7) is a well-defined convex optimization problem. Care has to be taken with program (6). The solution will be a boundary point of the focal element $E_{i,par}$, and we have to check both for the lower and upper bound of T_{2x}.

Fig. 3 shows the image $[\underline{\Delta T}(t), \overline{\Delta T}(t)]$ of a single focal element. The range of the image grows considerably in time. We performed a sensitivity analysis with partly resolved uncertainty. Uncertainty in the radiative forcing and climate parameters dominates the overall uncertainty in the first half of the 21st century, but socio-economic uncertainty becomes equally important in the second half of the 21st century. Most strikingly, the uncertainties on the subspaces combine in a nonlinear way. A much larger overall uncertainty is found in particular for cases where the natural systems and socio-economic uncertainties are of similar size.

The projected random set $(\mathcal{F}_{\Delta T}, m)$ for GMT increase can be used to construct the lower CDF $\underline{F}_{\Delta T}$ and upper CDF $\overline{F}_{\Delta T}$. It is important to note that the corresponding imprecise CDF model $\mathcal{M}_{\Delta T}(\underline{F}, \overline{F}) := \{ P \,|\, \forall x \in \mathbb{R} \;\; \underline{F}_{\Delta T}(x) \leq P(-\infty, x] \leq \overline{F}_{\Delta T}(x) \}$ can be more imprecise than $\mathcal{M}_{\Delta T}(Bel_{\mathcal{F}_{\Delta T}}) := \{ P \,|\, \forall A \in \mathcal{A} \;\; Bel_{\mathcal{F}_{\Delta T}}(A) \leq P(A) \}$, i.e. $\mathcal{M}_{\Delta T}(\underline{F}, \overline{F}) \supseteq \mathcal{M}_{\Delta T}(Bel_{\mathcal{F}_{\Delta T}})$. This is due to the fact, that after applying the extension principle, the focal elements $E_{i,\Delta T} = [\underline{\Delta T}_i(t), \overline{\Delta T}_i(t)] \in \mathcal{F}_{\Delta T}$ might violate condition (II) of proposition 1. In this case, the lower envelope $\underline{P}_{\Delta T}$ of $\mathcal{M}_{\Delta T}(\underline{F}, \overline{F})$ is strictly smaller than $Bel_{\mathcal{F}_{\Delta T}}$ for some $A \in \mathcal{A}$. Recalling proposi-



a. Full uncertainty

b. Forcing and climate uncertainty only

c. Climate uncertainty only

d. Forcing uncertainty only

Figure 4: Lower and upper CDFs for GMT increase ΔT in the years 2025, 2050, 2075, 2100

tion 2, it can be seen that $\mathcal{M}_{\Delta T}(\underline{F},\overline{F})$ does not contain more information than $(\mathcal{F}_{par},m)$, which captures the uncertainty in the model parameters, would allow.

$$\mathcal{M}_{\Delta T}(\underline{F},\overline{F}) \supseteq \mathcal{M}_{\Delta T}(Bel_{\mathcal{F}_{\Delta T}}) \supseteq \mathcal{M}_{par}(Bel_{\mathcal{F}_{par}})$$

Fig. 4 shows the lower and upper CDFs that are generated by the random set $(\mathcal{F}_{\Delta T},m)$. We consider the area between lower and upper CDF as an indicator for the *imprecision* in the uncertainty. It can be seen that the imprecision in the GMT estimate for the case of full uncertainty in the model parameters is enormous. This is partly due to the large number of uncertain parameters, as a comparison with the other cases shows. However, the cases (4.b) and (4.c) also exhibit large imprecision. This reflects the fact that the underlying imprecise CDF models for the climate parameters are already very imprecise. Certainly, they are conservative estimates, as the results of different studies were not weighed against each other. Some imprecision is also induced by the combination of the uncertainty for single parameters using random set independence (sec. 3.3).

The results can be compared with the IPCC estimate [1.8 K, 6.6 K] for GMT increase in 2100 relative to 1860 [3]. The probability for $\Delta T \in$ [1.8 K, 6.6 K] lies

in the interval $[0,1]$, for $\Delta T < 1.8$ K in $[0,0.95]$, and for $\Delta T > 6.6$ K in $[0,0.965]$. Despite the large range of the IPCC estimate, the uncertainty in GMT increase is too imprecise to discriminate against values outside this range. The probability mass allocated to values smaller than 1.8 K stems from random sets allowing for climate sensitivity values that are below the IPCC estimate for climate sensitivity. Similarly, the probability mass allocated to GMT increases higher than 6.6 K is due to climate sensitivity values above the IPCC estimate. As a comparison of (4.c) and (4.d) underlines, the uncertainty in climate parameters is the most influential factor on the uncertainty in GMT increase.

5 Conclusion

Imprecise probability concepts carry the potential to consistently capture the different types of uncertainties and different degrees of knowledge that are encountered in climate change analysis. However, they need to be applicable to dynamical problems with a large number of continuous uncertain variables. We suggest that imprecise CDF models are conceptually flexible and mathematically tractable enough to fulfil these competing requirements to some extent. When the imprecise CDF model is bounded by lower and upper step functions on the real line, the information about the encompassed set of additive probabilities can be condensed in a random set $(\mathcal{F},m)$. The corresponding belief function $Bel_{\mathcal{F}}$ is the lower envelope of the imprecise CDF model on the algebra generated by the half-closed intervals of the real line. Moreover, if the random set extension principle is used to project a random set onto the range of a measurable function, no information is added in the sense that every additive probability dominating $Bel_{\mathcal{F}}$ is transferred into a probability dominating the "extended" belief function.

We have constructed a random set for a simple climate model, and projected it onto an estimate for global mean temperature increase. The resulting estimate is very imprecise, with uncertainties about socio-economic development, radiative forcing and climate characteristics combining in a nonlinear way. The large imprecision of the estimate has different reasons and implications. Firstly, we incorporated a very broad range of factors in the analysis. Imprecision will be reduced if the range of factors is limited by formulating more specific questions. Secondly, we combined the random sets of single uncertain factors by assuming random set independence. This has increased the imprecision in the overall estimate, since aerosol forcing and climate sensitivity are not epistemically independent, when estimated from the present day climate change signal. Thirdly, the CDF models for the single parameters should be considered conservative estimates, which can be improved upon, when more comparisons of model results with historical data become available. Imprecision can be reduced in particular, if it is discriminated between the reliability of different models and methods.

Nevertheless, the results show that uncertainty is a key issue in the integrated

assessment of climate change. Random set methods provide new insights into the structure of the uncertainty, particularly into its imprecision. The link to imprecise CDF models seems to be an important yardstick for assessing information losses when combining random sets, and applying the extension principle. More theoretical work is needed here to enhance the applicability of random sets to climate change analysis. In addition, methods need to be developed to determine imprecise CDF models directly from a comparison of model results with historical data.

References

[1] Andronova, N. G., and Schlesinger, M. E. Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research 106* (2001), 22605–22611.

[2] Couso, I., Moral, S., and Walley, P. Examples of independence for imprecise probabilities. In *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications* (1999), pp. 121–130.

[3] Cubasch, U., and Meehl, G. Projections of future climate change. In *Climate Change 2001: The Scientific Basis*, J. Houghton and Y. Ding, Eds. Cambridge University Press, Cambridge, 2001, pp. 525–582.

[4] Dempster, A. P. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist. 38* (1967), 325–339.

[5] Dubois, D., and Prade, H. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems 42* (1991), 87–101.

[6] Forest, C. E., Stone, P. H., Sokolow, A. P., Allen, M. R., and Webster, M. D. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science 295* (2002), 113–117.

[7] Hall, J., and Lawry, J. Generation, combination and extension of random set approximations to coherent lower and upper probabilities. *Reliability Engineering and System Safety* (2003), in press.

[8] Knutti, R., Stocker, T. F., Joos, F., and Plattner, G. K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature 416* (2002), 719–723.

[9] Nakićenović, N., and Swart, R. *Emissions Scenarios. Special Report of the IPCC*. Cambridge University Press, Cambridge, 2000.

[10] Ramaswamy, V. Radiative forcing of climate change. In *Climate Change 2001: The Scientific Basis*, J. Houghton and Y. Ding, Eds. Cambridge University Press, Cambridge, 2001, pp. 289–348.

[11] Shafer, G. *A Mathematical Theory of Evidence*. Princeton U. Press, Princeton, 1976.

[12] Tonon, F. Using random set theory to propagate uncertainty through a mechanical system. *Reliability Engineering and System Safety* (2003), in press.

[13] Walley, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[14] Watterson, I. G. Interpretation of simulated global warming using a simple model. *Journal of Climate 13* (2000), 202–215.

H. Held is with the Potsdam Institute of Climate Impact Research, Germany. E-mail: kriegler@pik-potsdam.de

E. Kriegler is with the Potsdam Institute of Climate Impact Research, Germany. E-mail: kriegler@pik-potsdam.de