Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Joint deliverable D2B.26/D6.13 Recommendations and guidance on methods for the construction and use in impacts applications of probabilistic regional climate projections

Due date of deliverable: month 48
Actual submission date: month 60

Reason for discrepancy: Prolonged delay in availability of probabilistic information from both global models and downscaled information.

Note that this deliverable constitutes the first draft of a contribution to an ENSEMBLES Work Package 6.2 Special Issue of the journal *Natural Hazards and Earth System Science* (*NHESS)*, which is in preparation – see status report in Appendix 1

Start date of project: 1 September 2004 Duration: 60 Months

Lead Partners: SYKE: Finnish Environment Institute, Helsinki, Finland and UEA: University of East Anglia, Climatic Research Unit

Contributing Partners:

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Deliverable D2B.26/D6.13

Recommendations and guidance on methods for the construction of probabilistic regional climate projections and their application in impact modelling

Timothy R. Carter, [1]Clare E. Goodess[2] and Stefan Fronzek[1]

Outline

1. Rationale for examining probabilistic climate projections in ENSEMBLES

1.1. Why are we interested in probabilistic projections?

Projections of future climate are of interest to a wide variety of individuals and organisations throughout the world. While there is considerable scientific interest in simulating the behaviour of the climate system for its own sake, the prime motivation for projecting future climate is to examine how human-induced climate change may lead to unacceptable impacts on the natural environment and on society. Climate change can potentially affect us all, so decision-makers, researchers and the general public alike have a stake in understanding how climate is expected to change, so they can decide how to intervene by mitigating the causes and adapting to the effects.

Projections of future climate that are used to infer the potential consequences of climate change are commonly referred to as *climate scenarios* (see Box 1). Methods of applying climate scenarios to investigate uncertainties in future impacts have been refined over time, but the basic approach has changed little since impact studies were first conducted. Scenarios are usually developed from general circulation model (GCM) outputs, either applied directly or regionalised using high resolution limited area models or statistical downscaling, and impacts are estimated for scenarios that have been selected to embrace as realistic a range of uncertainties as possible (Mearns et al. 2001). In reality, the sample of GCM outputs commonly available for impact analysts to select from (e.g. sets of outputs produced for the IPCC Third and Fourth Assessment Reports) is finite, derived from a range of different model simulations that were not designed to sample model uncertainties systematically or randomly (Meehl et al. 2007a).

GCM outputs can thus be thought of as an "ensemble of opportunity" (Allen and Ingram 2002). Each projection is typically regarded as equally probable, with no effort made to assign

[1] Finnish Environment Institute (SYKE), Box 140, FI-00251, Helsinki Finland
[2] Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, UK

Box 1: Predictions and scenarios

Conventional climate change impact studies usually apply projections of climate variables derived from global climate models under given assumptions of radiative forcing due to atmospheric greenhouse gas and aerosol concentrations. Some of these projections are downscaled from global models to the region of interest, but a common characteristic of each projection is that it offers an internally consistent, physically plausible representation of future climatic conditions. Until recently, climate scientists have been reluctant to ascribe relative likelihoods to different climate model projections due to the difficulties of quantifying uncertainties in future radiative forcing, in the climate response to this forcing, and in various regional downscaling procedures. For this reason, projections are commonly referred to as ***scenarios*** rather than predictions or forecasts.

Several climate scenarios are usually selected by impact analysts in an attempt to sample the uncertainty range, but this procedure tends to be arbitrary, limited either by the availability of scenario information or by guidance on the choice and application of scenarios. The advent of an ensemble climate forecast system means that, in principle, a likelihood can be attached to each individual climate projection. Strictly speaking, once probabilities are attached to projections of future climate, these cease to be scenarios and should be regarded as ***predictions***. However, since the ENSEMBLES project is exploring different methods of quantifying uncertainties, and since some sources of uncertainty may not yet be adequately represented by these methods, the distinction between scenarios and predictions seems likely to remain blurred for some time to come. Moreover, it is important to point out that ENSEMBLES aims to quantify uncertainties in climate processes. The project does not seek to attach probabilities to future greenhouse gas and aerosol emissions, as these are controlled by future human behaviour and are unpredictable. Thus, the likelihoods being estimated for future climates are ***conditional*** on the emissions scenario assumed.

weights to different climate models according to their relative performance or reliability (see Murphy et al. 2004). When applied to evaluate impacts they are useful for exploring sensitivities to plausible future climates, but they fail to provide information on the relative likelihoods of different outcomes, making it difficult to evaluate the risk of alternative adaptation measures (New et al. 2007).

Few impact studies have addressed the likelihood of climate projections. However, recent progress in developing probability density functions (PDFs) of future climate changes (e.g. Meehl et al. 2007a; Murphy et al. 2004), exploiting improvements in computer power that have facilitated the generation of large ensembles of model simulations, provides an opportunity to move beyond "what if" type studies of potential impacts towards quantitative assessments of the likelihood of pre-specified levels of impact being exceeded under given scenarios of future radiative forcing.

In this paper, we discuss some of the opportunities and challenges offered by a probabilistic framework for studying the impacts of climate change. In particular, we focus on alternative methods that might be applied to climate projections to allow future impacts to be considered in terms of risk. Risk is commonly defined as the product of the probability and consequence of an event. Climate change can be described probabilistically by fitting a PDF to a large ensemble of individual projections of a given climatic variable. Even for global average changes, such a method of representing future climate presents a significant challenge for communicating the uncertainties involved in prediction.

A straightforward way to evaluate impacts is to run an impact model with each of these projections and then to determine the probability of exceeding a prespecified threshold of response (i.e. the risk of a consequence deemed unacceptable) on the basis of ensemble impact outcomes. Given the large number of ensemble projections required for such a procedure, this may be precluded on practical grounds, especially for more complex impact models. An alternative method is to construct an impact response surface from a sensitivity analysis of the impact model with respect to key climatic variables and to superimpose onto this a probabilistic representation of projected changes in these same climatic variables (Jones 2000). In this way, impact risk can be computed as the proportion of the climate PDF exceeding the threshold response.

1.2. Uncertainties in climate projections

Understanding about the Earth's climate system is evolving, and though there have been impressive advances over recent decades in our ability to predict the weather and climate, numerous uncertainties remain. Some of the terms more commonly used to describe different facets of uncertainty are listed in Box 2.

Here some of the key uncertainties are introduced, a number of which are described in more detail in Section 2. These are also illustrated in Figure 1.

1. *Observations* of socio-economic variables affecting emissions, of climate, of atmospheric composition and of impact-relevant variables are all subject to error, introducing uncertainty at all levels of an end-to-end assessment of future climate change and its impacts. The types of uncertainties include observational errors, gaps in spatial/temporal coverage and errors in interpolating point data to a regular grid.
2. *Anthropogenic emissons* of greenhouse gases and aerosols, by definition, originate from human actions, many of which are difficult to predict into the future. The major driving

Box 2: Some definitions of terms related to uncertainties

Uncertainty is an expression of the degree to which a value is unknown. Uncertainty can result from lack of information or from disagreement about what is known or even knowable. It may have many types of sources, from quantifiable errors in the data to ambiguously defined concepts or terminology, or uncertain projections of human behaviour.

Confidence refers to the level of scientific understanding about an occurrence, outcome or result. The level of confidence in a climate projection is a measure of the scientific evidence supporting the outcome, taking into account the set of assumptions upon which it is based. It does not imply a *likelihood* of occurrence.

Likelihood is the quantified probability of an occurrence, outcome or result. The ENSEMBLES project has made some initial attempts at expressing climate projections in terms of probabilities, albeit conditional on a given scenario of future greenhouse gas and aerosol emissions to the atmosphere.

Reliability of a climate projection refers to the *confidence* that can be placed in a accuracy of the projection. This is commonly based on performance criteria, such as the accuracy of a climate model's simulation of past or contemporary climate, which can be used to infer its likely accuracy in projecting future climate.

Robustness describes the level of *confidence* in a representation of climate processes or relationships, usually in a model framework, following rigorous testing and validation. It can also refer to the strength of evidence supporting a given outcome or conclusion.

factors determining emissions include human population and economic development, which affect the demand and use of energy, technological development, which influences the emissions per unit of energy production, land use, which affects emissions from the earth's surface, and energy policy, which determines the types of emissions sources.

3. Similarly, the conversion of atmospheric concentrations into estimates of *radiative forcing* involves various radiation schemes and parameterisations that introduce an additional set of uncertainties to projections.

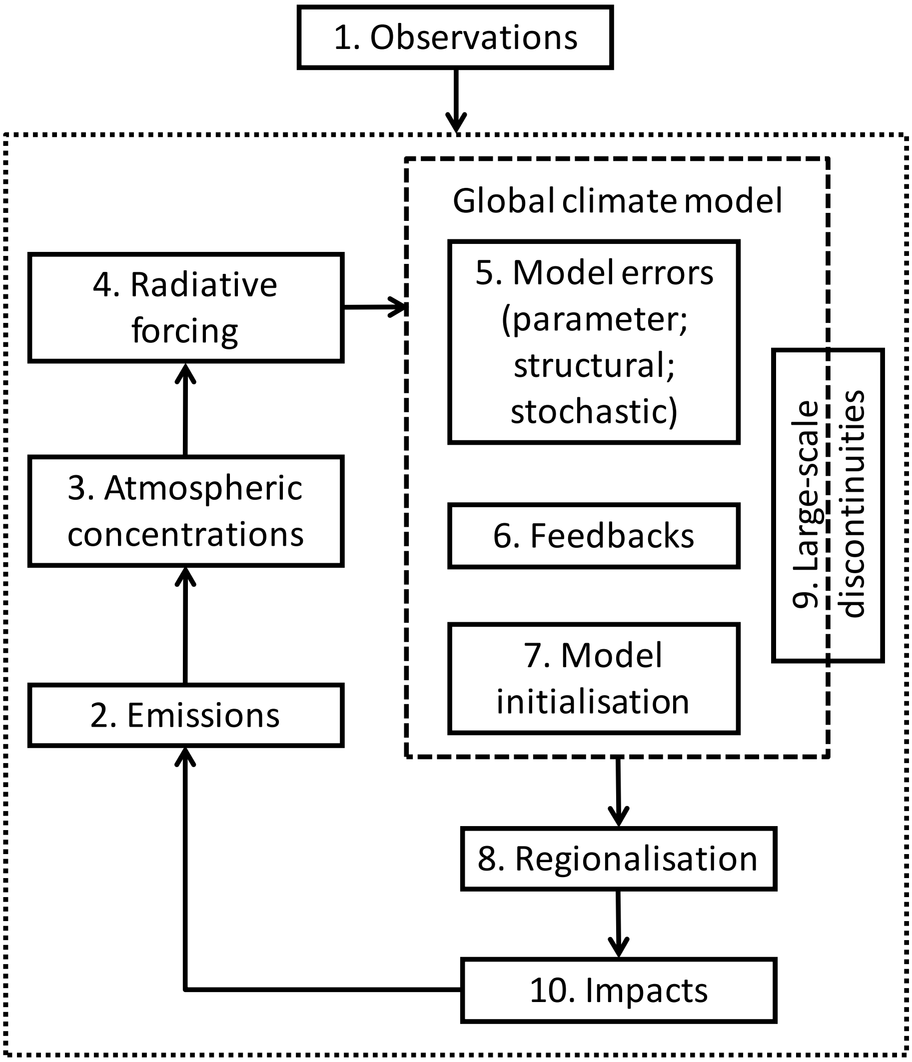

Figure 1. Sources of uncertainty in climate projections and estimates of impacts (boxes). Arrows indicate the direction of uncertainty propagation. Uncertainties in observations are of relevance both in their own right but also for evaluating the other nine sources of uncertainty (in the dotted rectangle), which all relate to representations of the future. See text for details.

4. Different classes of *model error* can contribute to significant uncertainties in projections, including:
    a. structural errors, inherent in basic model construction and expressed by differences between models
    b. parameter uncertainties, relating to the representation of sub-grid-scale processes (for example, cloud physics)
    c. stochastic uncertainties arising from coupling between unresolved sub-grid-scale variability and the resolved grid-scale flow

5. The *initial model state* can influence the detailed outcomes of a climate model simulation, recognising that the coupled atmosphere–ocean–biosphere–cryosphere system is chaotic, and the natural evolution of the climate is sensitive to small perturbations in initial conditions. Natural variations in the climate are commonly experienced as weather, but also reflected in multi-year and multi-decadal shifts in modes of variability of the atmosphere (e.g. features such as ENSO and the North Atlantic Oscillation). Examples of initial conditions that can have an influence on regional climate include ocean temperatures and atmospheric pressure patterns. They are usually examined by conducting ensemble simulations for a number of initial conditions.
6. *Feedbacks* refer to the effects of anthropogenically-induced climate change on climate system processes. Typical examples include the ice-albedo effect, the effect of climate changes on cloud formation, which itself affect the earth's radiation balance, and the changes in the ability of ecosystems to absorb carbon with climate warming. More broadly, human activities and greenhouse gas emissions may be modified in response to the socio-economic impacts of ongoing climate change.
7. *Regionalisation* from GCM-scale to regional- and local-scales embraces all of the sources of uncertainty described above, with additional considerations such as:
    a. for dynamical downscaling between GCMs and RCMs, uncertainties associated with the driving GCM boundary conditions or the choice of GCM/RCM pair;
    b. for statistical downscaling, uncertainties attributable to choice of predictors and the stationarity assumption
8. *Large-scale discontinuities*, such as circulation regime shifts, re-organisation of ocean currents or non-linear surface feedbacks, are often very difficult to predict though potentially of high consequence. They are commonly omitted from model projections, but some qualitative statements about their potential magnitude, timing and likelihood may be required by policy makers.

These sources of uncertainty are summarised in Table 1, along with an indication of the project research themes in which they are discussed. Any application of the ENSEMBLES climate projections (for example, to assess impacts) should bear in mind these sources of uncertainty. Coupled with uncertainties in estimates of the impacts themselves (see Section 3), this frames the treatment of uncertainty recommended for a comprehensive end-to-end analysis.

Table 1. Major sources of uncertainty in climate projections, parts of the ENSEMBLES project in which they are discussed and identification of sources treated probabilistically. Impacts are also included in the table to illustrate the "end-to-end" treatment of uncertainties applied across the project.

Source of uncertainty	ENSEMBLES research theme	Probabilistic treatment
1. Observations	RT5	No
2. Emissions	RT7	No
3. Concentrations	RT2A	No
4. Radiative forcing	RT2A	No
5. Model errors	RT1, RT2A, RT3, RT5	Yes
6. Model initialisation	RT1, RT2A, RT3	Yes
7. Feedbacks	RT1, RT6	Yes
8. Regionalisation	RT2B, RT3	Yes
9. Large-scale discontinuities	RT1	No
10. Impacts	RT2B, RT6, RT7	No

Table 1 also indicates whether there has been an attempt in the ENSEMBLES project to quantify a given type of uncertainty in terms of probabilities. A number of different approaches were adopted for addressing this question, and these are described in more detail in Section 2. In those cases where projections of future climate were expressed probabilistically, an opportunity was thus presented to apply this information in impact assessments. Note that although studies were conducted to estimate impacts on the basis of probabilistic climate projections, none of these impact studies undertook an equivalent probabilistic assessment of uncertainties in the impact estimates themselves. This issue remains to be addressed in subsequent research.

1.3. How can impact and adaptation studies make use of probabilistic climate projections?

In recent years there has been a proliferation of projections of future climate from climate models at varying scales of temporal and spatial resolution. Model simulations are usually conducted for periods in the 21st century, assuming a variety of emissions scenarios (commonly selected from the SRES set of forcings). Ensemble simulations are frequently carried out, assuming different initial conditions, and more recently systematic analyses of parameter and structural uncertainties have also been conducted at some centres. For regional applications, many impact analysts prefer to adopt projections that have been dynamically or statistically downscaled from global model outputs to a region of interest. The ENSEMBLES projections described in Section 2 complement the existing datasets of projections already available from earlier projects such as STARDEX, PRUDENCE, ATEAM and ALARM as well as the GCM outputs archived for the IPCC AR4, a subset of which is to be found at the IPCC Data Distribution Centre.

Most of these projections are stored as outputs from individual models, giving estimates of key variables required in impact assessments (e.g. surface air temperature, precipitation, global irradiance, atmospheric humidity and windspeed). Each model output, or derivative downscaled output, is commonly treated as a separate *scenario*, indicating that the projection is a physically plausible and internally consistent representation of the future climate. This is typically conditional on a given set of assumptions about future emissions of greenhouse gases and aerosols. Scenarios are usually selected by impact analysts to offer a representative range of future conditions (i.e. accounting for a range of uncertainties in projections – cf. Table 1). In reality, the character and number of scenarios selected may be limited by access to projections and resource limitations for extracting and then applying multiple scenarios in impact studies. Hence the number of climate scenarios applied by impact analysts working in Europe is usually a very small subset (typically 2-5) of the available global climate model "ensemble of opportunity" archived for Europe (cf. Section 1.1). Procedures for scenario selection are often arbitrary, and tend to vary between impact studies unless efforts are made to impose common scenarios for promoting the inter-comparison of study outcomes.

Many of the impact studies undertaken in ENSEMBLES apply climate scenarios drawn from the ensemble of model outputs provided in the project. Some studies attempt to apply all available climate model outputs (e.g. a study of wind damage potential by Donat et al. 2010 – see Section 3), but where this is not feasible, selection procedures have been applied to select a subset of models (e.g. a study of space heating and cooling demand in Europe by Giannakopoulos et al. 2009). Some criteria that might be applied for selecting climate scenarios are suggested in Section 2.

A unique feature of the ENSEMBLES project is that projections for some variables have been expressed in terms of probabilities, using as much of the available information about uncertainties as possible. In effect, the procedures followed to generate probability density functions (PDFs) have involved some attempt at quantifying the spread of plausible regional outcomes for a given forcing emissions scenario (see section 2). However, the question posed by some impact analysts in Research Theme 6 has been whether such summary PDFs can be applied *directly* in impact studies.

A method explored in several impact studies involves undertaking a sensitivity analysis of an impact model with respect to key climate variables and thereby constructing an impact response surface showing how impacts vary with changes in climate in a given region. An impact threshold is then selected, which is a level of impact the exceedance of which is judged to be unacceptable by a decision maker (e.g. a crop yield shortfall exceeding a certain percentage; a minimum depth of water in a reservoir for safe extraction). A probabilistic representation of the climate of the study region for the same climatic variables used to plot the impact response surface is then superimposed onto the impact response surface. The risk of the impact threshold being exceeded can then be estimated as the proportion of the superimposed climate PDF in the exceedance zone.

A number of examples are provided in section 3 to illustrate the application of the response surface approach. The study areas considered range in size from site-scale through to Europe-wide. In all cases, joint PDFs of projected seasonal or annual air temperature and precipitation changes (representing a wide set of uncertainties across climate models of varied complexity and incorporating observational constraints) were obtained from the Met Office Hadley Centre (MOHC) for GCM grid boxes over Europe. In principle, probablistic information could also be generated at higher time resolutions (e.g. monthly, daily or sub-daily), though these have not been applied in ENSEMBLES studies. For example, the recent UKCP09 climate projections include an option to generate climate projections for regions of the UK at daily and sub-daily time steps by stochastically sampling from the MOHC PDFs described above using a weather generator (Jones et al., 2009).

2. Development of probabilistic climate projections for Europe to apply in IAV analysis based on ENSEMBLES global and downscaled model outputs

2.1. The rationale for and availability of ENSEMBLES model projections

The main starting point for ENSEMBLES work on developing probabilistic climate projections for Europe was the GCM "ensemble of opportunity" (see Section 1.1) produced for the IPCC Fourth Assessment Report (AR4) – in particular, the six GCMs run by groups participating in ENSEMBLES RT2A. Thus uncertainties due to structural GCM errors were sampled to some extent – although the models were selected on non-scientific grounds (the location of the modelling centres). In cases such as this, where it is not possible to work with the full "ensemble of opportunity", it is still important to be aware of the extent to which the selected models reflect the larger ensemble range. The ENSEMBLES GCMs project a smaller range in global mean warming than the full IPCC AR4 ensemble. Thus runs from the METO-HC GCM perturbed physics ensemble were also employed. As well as the 'reference' scenario, two simulations with low and high climate sensitivity respectively, and showing very different climate responses, were used. Thus eight different European GCMs were available, together with one Canadian GCM from an ENSEMBLES affiliated partner, allowing consideration of *uncertainty in boundary conditions* for dynamical downscaling.

ENSEMBLES provided the opportunity to perform a co-ordinated set of RCM simulations using a common European domain and providing common outputs. As with the choice of GCMs, the choice of RCMs was constrained by the geographical location of modelling centres and participation in ENSEMBLES RT3 and RT2B – thus RCMs from 15 European and one Canadian modelling centre were available.

Ideally, every RCM would have been run using every GCM – but this was not possible due to limitations in computing resources. While the sampling of uncertainty was an important consideration in the experimental design of ENSEMBLES, other considerations – particularly relating to user needs – also had to be taken into account. Thus it was agreed before the start of the project to run the RCM simulations at 25 km resolution and to run transient simulations from 1950-2050 (and where possible to 2100) – thus providing users with higher-resolution, transient information for the next few decades (compared with the earlier PRUDENCE project which ran RCMs at 50km resolution for two snapshot periods – 1961-1990 and 2071-2100). The period out to 2050 is generally of greater interest to most stakeholders than the end of the century, although the latter period does have the advantage of maximising the signal-to-noise ratio (and is, for example, of interest to those planning large-scale infrastructure such as dams, reservoirs and flood protection measures). Inevitably, experimental design reflects a compromise between a number of different considerations. The length and spatial resolution desired of the ENSEMBLES RCM simulations constrained the number of simulations that could be undertaken and thus the sampling of uncertainty.

Within the constraints outlined above, ENSEMBLES RT3 and RT2B designed a GCM/RCM experimental matrix, i.e., the pairing of available GCM runs and RCMs, focusing on the sampling of *uncertainty in boundary conditions* and *uncertainty in RCM model formulation* – both of which were shown to be important in the earlier PRUDENCE project. They can be viewed as different aspects of uncertainty due to model error and regionalisation (see Section 1.2).

At the start of ENSEMBLES, most of the selected RCMs were either new versions or had not been tested for the ENSEMBLES set-up. Thus there was no *a priori* information as to which

RCMs might amplify or weaken the GCM climate change signals and thus artificially inflate or downplay uncertainty. It was therefore decided to spread efforts over all available models within the limits of the available computational resources. One of the six European GCMs was not used because it was of lower spatial resolution, but, so far as possible, all other GCMs were downscaled by at least two RCMs. The final GCM/RCM matrix encompassing 25 simulations is shown in Table 2.1.

While the RCM ensemble encompasses uncertainty in boundary conditions (multiple GCMs) and RCM model formulation (multiple RCMs), it does not encompass *uncertainty due to initial conditions* (initial model state – Section 1.2). All simulations are based on the A1B emissions scenario thus *uncertainty in radiative forcing* related to greenhouse gas emissions and atmospheric concentrations (see Section 1.2) is not considered. The ENSEMBLES-wide decision to use A1B for the development of regional projections was made to allow focus on the 'downstream' uncertainties – and recognising that the choice of emissions scenarios is less important for the next 40 or 50 years than towards the end of the century. Parameter uncertainties (an aspect of model error – see Section 1.2) are only addressed through use of three members of the METO-HC perturbed physics simulation (Standard, Low and High sensitivity – HadRM uses the same parameterisations as the GCM in each case).

Although there are gaps in the GCM/RCM matrix, the ENSEMBLES RCM ensemble is the largest currently available for any region of the world and the resulting TeraBytes of common output are readily accessible to users through the central archive hosted by the Danish Meteorological Institute (DMI).

As noted above, no suitable *a priori* criteria were available for designing the GCM/RCM matrix (Table 2.1). However, subsequent analyses provide some *post hoc* justification for the matrix experimental design which implicitly assumes that any GCM can be used to drive any RCM. This assumption is supported by analysis of large-scale consistency between the driving GCM and RCM and analysis of the fine-scale skill of the RCMs. These analyses do not reveal any examples of poor quality RCM integrations caused by poor matching with the driving GCM. Thus all the GCMs are considered to be valid choices for driving any RCM.

It is recommended that the full RCM ensemble be used wherever possible. However, where this is not possible, because a limited number of time series are required to run a computationally expensive impacts model, for example, ENSEMBLES RT3 has produced some guidance on how to select a subset. Choosing a subset based on the weighting scheme (see Section 2.4) is not necessarily the best strategy since it might lead to an undersampling of uncertainty. The minimum recommended requirement is to use results based on two or more RCMs that are forced by at least two GCMs (i.e., an absolute minimum of four simulations). The full GCM-RCM matrix should then be used as supporting information on how the subset relates to the other cases.

In part, any sampling strategy (for identifying a subset of the RCM ensemble or for setting priorities for filling out the matrix with additional RCM simulations), should take account of the demonstration by ENSEMBLES that the balance of uncertainty varies depending on the future time period considered. The general message emerging from the new ENSEMBLES studies (Déqué et al., 2009; Kendon et al., 2008, 2009), and the earlier PRUDENCE work (Déqué et al., 2007), can be summarised as: the higher the climate change signal, the more important the GCM spread, the lower the signal, the more important the RCM. This implies that, for the end of the century, it is important to fully sample the range of GCM uncertainty, whereas, for periods closer to the present day, more RCMs should be sampled. The design of the ENSEMBLES GCM/RCM matrix gives users the opportunity to adopt such different sampling strategies.

- For further information see:
    - Sections 5.2.6 and 6.2.1 of the ENSEMBLES final report
    - D3.2.1, D3.2.2 Appendix, D3.3.1, D3.3.2, D3.4.1, D3.4.2, D2B.1
    - Déqué et al., 2009; Kendon et al., 2008, 2009

2.2 Scaling methods for enhancing the representation of uncertainties (filling gaps in the matrix)

The RCM simulations conducted by ENSEMBLES were a finite set of experiments which explored some, but not all, of the uncertainty space described by different GCMs and different RCMs – all conditional on the A1B emissions scenario. Each simulation occupies its own position in the matrix of experiments, but some elements of this matrix remain empty due to constraints on resources for undertaking model simulations (Table 2.1). Pattern-scaling is potentially a technique that can be used to substitute for these missing data in the RCM matrix. The conventional approach is to estimate the local climate response using global mean temperature change as the scalar and has been shown to be skilful for temperature but not for precipitation.

In ENSEMBLES, a new local scaling approach was developed, using the large-scale GCM change as a predictor of the 2080s RCM response (Kendon et al., 2009). Initial results suggest this may be skilful for different driving GCMs (i.e., the RCM response can be predicted for untested GCM-RCM pairs) and is therefore skilful for filling the GCM-RCM matrix. In particular, it performs well for precipitation (mean, variance and extremes) across much of Europe in winter; and for temperature (mean and extremes) in summer and winter, with the exception of central Europe in summer. Internal variability may, however, lead to to a substantial apparent reduction in scaling skill for precipitation, with scaling relationships only being reliable where the local change is robust compared to internal variability. Although it shows potential, due to limited time and resources, it was not possible to apply this approach to complete the matrix (Table 2.1) as part of ENSEMBLES.

An alternative approach to matrix filling was developed in ENSEMBLES based on the ANOVA technique employed in PRUDENCE (Déqué et al., 2007). The refined approach is based on weather regime decomposition (four clusters are defined using daily 500 hPa height values over the North Atlantic-Europe domain) and hence also accounts for the GCM providing large-scale forcing on the RCM. The method assumes that the way the RCM behaves under a given regime only depends on the RCM. The matrix completion was tested by removing an RCM-GCM pair and comparing its reconstruction with the original response. This could be done only for the RCMs that were run with two GCMs. For these runs, the reconstructed results show reasonable skill for temperature, but demonstrate only little skill for precipitation. Seventeen RCM runs were available at the time this matrix-filling approach was developed and tested. Now that additional runs are available, further testing and analysis is underway. Nonetheless, the approach has been used to complete the matrix using these first 17 runs. The average of the 17 original responses in mean seasonal temperature for 2021-2050 compared with 1961-1990 is found to be little changed when considering the full matrix of 98 responses.

- For further information see:
  - Sections 5.2.6 and 6.3.4 of the ENSEMBLES final report
  - D2B.7, D2B.25
  - Déqué et al., 2009; Kendon et al., 2009

2.3 Enhancing the suite of ensemble projections with statistical downscaling

The ENSEMBLES RCM simulations provide downscaled information at 25 km grid-box resolution for a limited number of elements in the GCM/RCM matrix (Table 2.1). The potential for the statistical downscaling approach to enhance this information has been explored in ENSEMBLES from a number of different perspectives:

- The provision of higher spatial-resolution (i.e., point scale) information
- Filling gaps in the GCM/RCM matrix
- Bias correction.

Some users prefer to work with point rather than grid-box values, i.e., something closer to what is measured at a weather station. This may be particularly desirable when the interest is in extreme weather events, since area-average extremes will always be less intense than station/point values. A comparison of the ENSEMBLES observed 25 km daily gridded dataset (E-OBS) with station values, for example, indicates about a 30% reduction in the intensity of the 10 year return period rainfall event for the gridded series (Haylock et al., 2008). Provided that appropriate station data are available for calibration and validation (i.e., long, homogeneous series), statistical downscaling can provide the desired station-scale information.

In the traditional approach to statistical downscaling for climate change applications (the perfect prog approach, rather than the model output statistics approach which is often used in weather and seasonal forecasting applications), the statistical model is constructed using large-scale GCM predictors and local/point-scale observed predictands. This approach has been used in ENSEMBLES, developing models which are tailored towards rather specific (both in terms of geographical location and variables) user applications:

A1 A non-homogeneous hidden Markov model used to explore changes in Danube riverflow extremes (developed by NIHWM)
A2 A canonical correlation analysis (CCA) based method used to construct PDFs of changes in temperature extremes for Bologna and northern Italy (developed by ARPA-SMR)
A3 A conditional stochastic weather generator used to investigate changes in precipitation extremes over Romania (developed by NMA).

Another group of statistical downscaling approaches developed in ENSEMBLES makes use of RCM rather than GCM output:

B1 Time-series resampling from RCMs to simulate changes in return levels (up to 1000 years) of extreme precipitation events in the Rhine basin. Three different non-linear bias correction methods have also been developed and applied to the resampled series. (Developed by KNMI)

- B2 Stochastic weather generator with change factors derived from RCM output – used to construct PDFs and CDFs of daily temperature and precipitation change for 13 European stations (developed by UEA)
- B3 Predictors from the ERA-40 forced ENSEMBLES RCM runs are used to downscale seasonal indices of temperature and precipitation extremes for four Irish stations using multiple linear regression. A bias correction method is also incorporated. (Developed by C4I)

A third group of approaches developed in ENSEMBLES employs both GCM and RCM output:

- C1 A two step downscaling approach in which output from ENSEMBLES reanalysis forced RCMs is used as "pseudo observations" to investigate North Sea marine surface winds. Monthly mean wind components u and v are obtained from the large-scale information using multi linear regression and then daily wind speeds are obtained using a multi variate autoregression (mvAr) model. (Developed by GKSS)
- C2 An approach in which GCM output is used as predictors and RCM output as predictands for analog and regression-based methods implemented in the ENSEMBLES downscaling portal (see below) (developed by UC)

The ways in which these methods have been applied in the construction of probabilistic projections are outlined in Section 2.5.

All eight approaches listed above provide higher-spatial resolution information than the RCMs. Although only approaches B1 and B3 explicitly incorporate a bias correction step, all approaches which use observations as predictands implicitly encompass some element of bias correction. The 'added value' of downscaling (i.e., better agreement between downscaled values and observations than between raw model outputs and observations) needs, however, to be demonstrated, rather than taken for granted.

The approaches which use GCM output as predictors have the potential to fill gaps in the GCM/RCM matrix. However, those which use point values as predictands (e.g., A1 to A3) do not provide information which is directly comparable with RCM output because of the spatial averaging issue identified above. Directly comparable information can, however, be obtained if gridded data (either observed or simulated) are used as predictands. Both approaches have been demonstrated in ENSEMBLES.

In the first example, E-OBS have been used as predictands in a two-step analogue downscaling model and results for the whole European domain compared with those from the ERA-40-driven ENSEMBLES RCM runs. Results for trends in temperature and precipitation extremes (e.g., Tmax95th percentile) have been compared. The statistical method displays consistently higher correlations with E-OBS than do the RCMs, but the standard deviations of spatial variability across the European domain are consistently low for the statistical method – lower than the worst of the RCMs in the majority of cases. Overall, however, the statistical model results lie somewhat closer to the 'perfect point' on the Taylor Diagram used for validation than those of the RCMs.

In the second example (approach C2), predictors are taken from a GCM run, while output from an RCM forced by this GCM (for present and future periods) is used as pseudo-observations. Three different statistical downscaling models implemented in the

ENSEMBLES downscaling portal (see below) are calibrated using simulated predictors and predictands for the present day (1961-2000). The models are then applied to future periods (2041-2070 and 2071-2100) using GCM predictors, and compared with the RCM output. The statistical methods are generally able to reproduce the RCM simulations of future climate periods – with only small relative increments in the error metric over time. The only exception is the analogue downscaling method, which exhibits large increases in error for summer temperature in the final period. This analysis was undertaken with the aim of exploring the validity of the assumption of stationarity in statistical downscaling. As well as providing insight into this issue, the results (albeit for one region – the northern Iberian Peninsula and one GCM/RCM combination) indicate that statistical downscaling can be an appropriate technique for matrix filling and increasing the ensemble size (i.e., for RCM emulation). It should not, however, be used for untested GCM/RCM pairs – which would require calibration using a reanalysis forced RCM run - without further evaluation such as is being done for the weather-regime matrix filling technique (see Section 2.2). Some theoretical support for the RCM emulation approach is provided by the finding that GCMs and RCMs appear to behave somewhat independently (see Sections 2.1 and 2.4).

Discussions with ENSEMBLES applications users and others during the early stages of the project identified a desire for statistical downscaling software tools (as much as statistically downscaled projections themselves). Thus over the course of the project, UC developed and refined the ENSEMBLES downscaling portal – www.meteo.unican.es/ensembles. The portal provides user-friendly homogeneous access to a subset of ENSEMBLES GCM (both seasonal predictions and climate change projections) and RCM outputs, allowing local interpolation or downscaling to the region/location of interest and bias removal. Users can also upload their own observed grids or networks and interactively downscale the model outputs testing several statistical downscaling methods (including regression, neural networks, analogues and weather typing). Examples of how the portal has been used to explore some specific scientific issues are provided above.

The availability of statistical tools such as the ENSEMBLES downscaling portal makes it easier to:

- Exploit the lower computational demands of statistical downscaling
- Work with larger ensembles
- Work in different regions (subject to availability of calibration/validation data)
- Work with different predictands
- Test the stationarity assumption
- Explore the potential for RCM emulation

There is, however, a potential danger in such tools being used as 'black boxes' and thus a need for user documentation and education. Feedback from ENSEMBLES users has helped to shape both the development of the portal and the accompanying user guidance.

- For further information see:
  - Sections 6.3.2, 6.3.3, 6.3.5, 6.5.2 and 6.6.3 of the ENSEMBLES final report
  - D2B.2, D2B.18, D2B.19, D2B.23, D2B.27, D2B.31, D2B.36, M2B.18

2.4 Weighting of model outputs based on their skill at representing current climate

Many of the techniques for constructing probabilistic projections from multi-model ensembles (see Section 2.5) require or permit the use of model weighting, i.e., rather than treating all models as equally likely, performance-based weighting is used to give more emphasis to those models which perform 'better' and less emphasis to those which perform 'less well'. Performance is defined as the ability of a model to reproduce observed climate – which is generally seen as a 'necessary but not sufficient' condition with respect to future performance.

The development and application of model weighting schemes is one of the cross-cutting issues addressed in many of the ENSEMBLES Research Themes (RTs). At the start of the project, rather little prior experience was available – particularly with respect to weighting in the context of regional projections (most of the experience related to projections of mean global temperature change). Thus considerable time and effort was put into identifying and discussing a number of critical key questions – concerning the rationale and philosophy behind weighting schemes, as well as their technical details.

These questions were also of interest to the wider research community. In July 2007, a meeting on 'Regionalizing climate models with skill' was co-hosted by ENSEMBLES and NCAR (Linda Mearns – an ENSEMBLES affiliated partner) and WCRP to discuss six key questions identified by the hosts:

1. What are sensible regional tests for AOGCMs?
2. What do climate models need to get 'right' for a given region?
3. Should there be some combination of regional and global indices of credibility?
4. What are the differences in user vs. public confidence criteria?
5. How can we incorporate expert judgement in weighting schemes?
6. What are the most important/urgent improvements needed in AOGCMs in order to improve regionalization of climate predictions?

Neither the 2007 meeting nor the ENSEMBLES work has been able to address or answer all these rather broad questions. However, some progress has been made in ENSEMBLES in addressing a rather more specific set of questions identified in cross-RT discussions in September 2005 – and first examples of weighting schemes have been produced and tested. The answers to some of these questions from the perspective of regional downscaling are summarised here:

- Is weighting a necessary and appropriate technique?
    - In theory yes – see Section 5.2.5 of the ENSEMBLES final report. In practice, weighting is always used [e.g., equal weighting with all weights = 1; or 1 (for models used) and 0 (for models not used)].
- How should weights be calculated?
    - See Section 5.2.5 of the ENSEMBLES final report for a discussion of the weighting scheme developed for dynamical downscaling and Section 6.2.2 for the scheme for statistical downscaling – see also Table 2.2 and discussion below.
- How should weights be used to construct PDFs and other forms of probabilistic projections?
    - See examples in Section 2.5 of this deliverable, and Section 6.2.3 of the ENSEMBLES final report.
- How can the performance of a weighted prediction be compared with an unweighted one?

    - See Section 6.2.2 of the ENSEMBLES final report (and Figures 6.11 and 5.6) and discussion below.
- Should the weights be assigned to the RCMs and GCMs separately or to the RCM-GCM combination?
    - This question has not been specifically addressed in ENSEMBLES – only the first approach has been followed – but see discussion below.

Table 2. Criteria and evaluation metrics used in the ENSEMBLES weighting schemes for statistical and dynamical downscaling.

Proposed weighting scheme for statistical downscaling	*RT3 weighting scheme for dynamical downscaling*
1. Statistical downscaling model performance (i.e., conventional metrics such as correlation, bias and RMSE calculated over an independent validation period)	f1: large scale circulation based on a weather regime classification
2. Reproduction of trends and climate states	f2: meso-scale signal based on seasonal temperature and precipitation analysis
3. Performance of driving-model predictors (i.e., ability of forcing GCMs to reproduce the predictors used for statistical downscaling)	f3: probability density distribution match of daily and monthly temperature and precipitation analysis
4. Stability of predictor-predictand relationships	f4: extremes in terms of re-occurrence periods for temperature and precipitation
5. Correction for multiplicity of statistical downscaling models	f5: trend analysis for temperature
	f6: representation of the annual cycle in temperature and precipitation

The preliminary weighting schemes developed for dynamical and statistical downscaling in ENSEMBLES (Table 2.2) share a number of common general features:

- They are not tailored to any specific application
- They are not based on a single variable
- The metrics extend beyond simple measures of mean climate (e.g., they include extremes and trends)
- All metrics are treated equally in calculating the final weight (but the possibility exists to give some metrics more 'weight' than others)
- The metrics are combined using multiplication.

There are also some differences between the two schemes. In particular, the statistical downscaling weights are calculated for each individual location using the station series as the observed target or reference. Whereas the dynamical downscaling weights are calculated for each of the eight Rockel regions using ERA-40 as the reference, since this provides the RCM forcing. A European average is then produced and it is recommended that this is used rather than the sub-regional values.

In the case of the weighting scheme for statistical downscaling, only criteria (1) and (2) have been implemented to date. Criterion (3) could be fairly readily implemented using, for example, EOF analysis – the results of which could be quantified as a single metric using anomaly correlation coefficients and root mean square errors. While desirable, it is not evident how a quantitative metric for criterion (5) could be implemented. The issue of independence is also highly relevant to RCM weighting – since many RCM implementations stem from the same roots and share the same or similar physics. There is thus a danger that these models get too large a weight in the final outcome. Ideally, a weighting scheme should compensate for this effect.

While the standard approach is to combine the different metrics using multiplication – two further approaches have been used in combining the six dynamical downscaling metrics (Table 2.2). The first alternative is 'normalisation' - obtained by constraining the ratio between the highest and lowest assigned weights. The second alternative, is to rank all models according to their order of performance in terms of each of the metrics, sum these ranks, and then transform this rank sum into a model weight which is obtained by dividing the sum of the rank sums by the rank sum of each model and then normalising it so the total sum of weights is equal to 1. These two alternatives have the effect of reducing the spread between model weights. It should be noted that they should only be applied when all members of the RCM ensemble used to construct the weights are used.

The weighting schemes developed for both statistical and dynamical downscaling should be viewed as first and partial attempts. A number of issues still need to be resolved or further explored. For example, one of the justifications for developing rather broad and general schemes (rather than schemes focused on particular applications), was the desire to capture the skill of underlying model processes. Both sets of weights include the reproduction of trends, for example, based on the argument that the ability to capture observed temporal trends implies that a model is capturing some of the processes associated with anthropogenic warming. This relates to the views expressed by ENSEMBLES scientists in workshop discussions that a climate model is more likely to be credible if the climate physics and processes are well represented than simply if the simulated mean control climate for the important variables is close to the mean observations. However, while other ENSEMBLES research themes have done a lot of work exploring process issues (see Chapters 9 and 10 of the ENSEMBLES final report), it has not been possible to incorporate this work or qualitative expert knowledge in the quantitative weighting metrics. Similarly, it has not been possible to include information about model performance on seasonal-to-decadal timescales (though the latter is more of an initial value problem – whereas climate change is more of a boundary value problem and may involve feedbacks which are not relevant on shorter timescales).

A number of issues relating to the use of weighting in end-to-end applications still need to be considered or resolved. In particular, it has not yet been possible to combine regional weights with those from global models. For statistical downscaling, one strategy would be to incorporate metric three (see Table 2.2). In the case of dynamical downscaling, one potential approach would be to calculate the RCM weights using the GCM-forced control runs (the current weights are based on ERA-40 forced runs). An analysis of 50 km resolution simulations performed with the RCA RCM indicates that the biases are generally larger when RCA is forced by GCMs compared to ERA-40 and that the biases are very different given different GCMs (Kjellström et al., 2009). In contrast, an analysis of fine-scale skills and large-scale consistency between the ENSEMBLES RCMs and forcing GCMs, indicates the independence of the GCM and RCM contributions (see deliverable D3.4.2). This suggests that it would be legitimate to use a simple weighting scheme based on the product of GCM and RCM weights. Whereas, if the processes are non-linear, the weighting has to be done based on the output of the combined process.

Another issue which has not been fully addressed is the extent to which the weighting schemes are considered relevant by applications users. One potential draw-back of using more process-based or general weighting schemes is that what is considered to be the 'best' model overall, may not be very good for the particular variable, season and region of interest to the user. On the other hand, varying the weighting for each impact model may lead to problems across impact models in an integrated impact assessment.

More importantly in the context of end-to-end applications (both on seasonal-to-decadal and climate change timescales), it has not been possible to address the following two general questions in ENSEMBLES:

- Can weights from impacts models also be combined?
- At what stage(s) should the weighting be applied in an integrated (from the coupled model, through the downscaling to the application model) prediction system be carried out?

A number of caveats concerning the construction and use of the ENSEMBLES regional weighting schemes need to be considered by users and it is recognised that there are inevitably subjective choices in constructing these preliminary weights. Nonetheless, they provide first examples for use in developing probabilistic projections (see Section 2.5) and for exploring the sensitivity of the projections to the use (or not) of weighting. Only limited sensitivity tests have been performed so far and indicate that weighting makes rather little difference to the projections (e.g., to the PDFs). However, these findings may be specific to the particular methods and regions considered - and it is not yet possible to draw more general conclusions about sensitivity to weighting of regional projections. It may also be the case that incorporating GCM weights would increase the sensitivity to weighting.

- For further information see:
    - Sections 5.2.5 and 6.2.2 of the ENSEMBLES final report
    - D2B.8, D2B.32, D3.2.2, D3.4.2
    - A special issue of *Climate Research* on 'Regional climate model evaluation and weighting' is being produced by ENSEMBLES RT3 partners.

2.5 Techniques for analysing multi-model ensembles to generate probabilistic projections

In preparation .....

2.6 Interpreting multi-model ensembles of climate projections over Europe

In order to locate the model simulation results from ENSEMBLES in the context of wider uncertainties, Figure 2 has been constructed to compare projections of mean annual temperature and precipitation change over northern Europe (48°N-75°N 10°W-40°E) and the Mediterranean Basin (30°N-48°N 10°W-40°E) for the mid-term (2030-2050) and long-term (2080-2100) future. Using these plots, it is possible to compare GCM-based and RCM-based projections generated specifically in ENSEMBLES for the A1B (medium, non mitigation) emissions scenario with probabilistic projections based on multiple sources of information for the same emissions scenario. In addition, Figure 2 also depicts the ENSEMBLES GCM-based projections of climate change under the E1 mitigation scenario. This is a scenario specially developed using an integrated assessment model in Research Theme 7 of ENSEMBLES, in which atmospheric concentrations of greenhouse gases are stabilised at 450ppm CO_2-equivalent. Comparison of the green (E1) with the blue (A1B) symbols indicates clearly how the effects of mitigation become apparent only towards the end of the 21st century. The depiction of uncertainty in future projections using bivariate PDFs such as those shown as coloured zones in Figure 2 has been carried forward in the next section into impact assessments, where probabilities of future climate outcomes have been translated into estimates of impact risk.

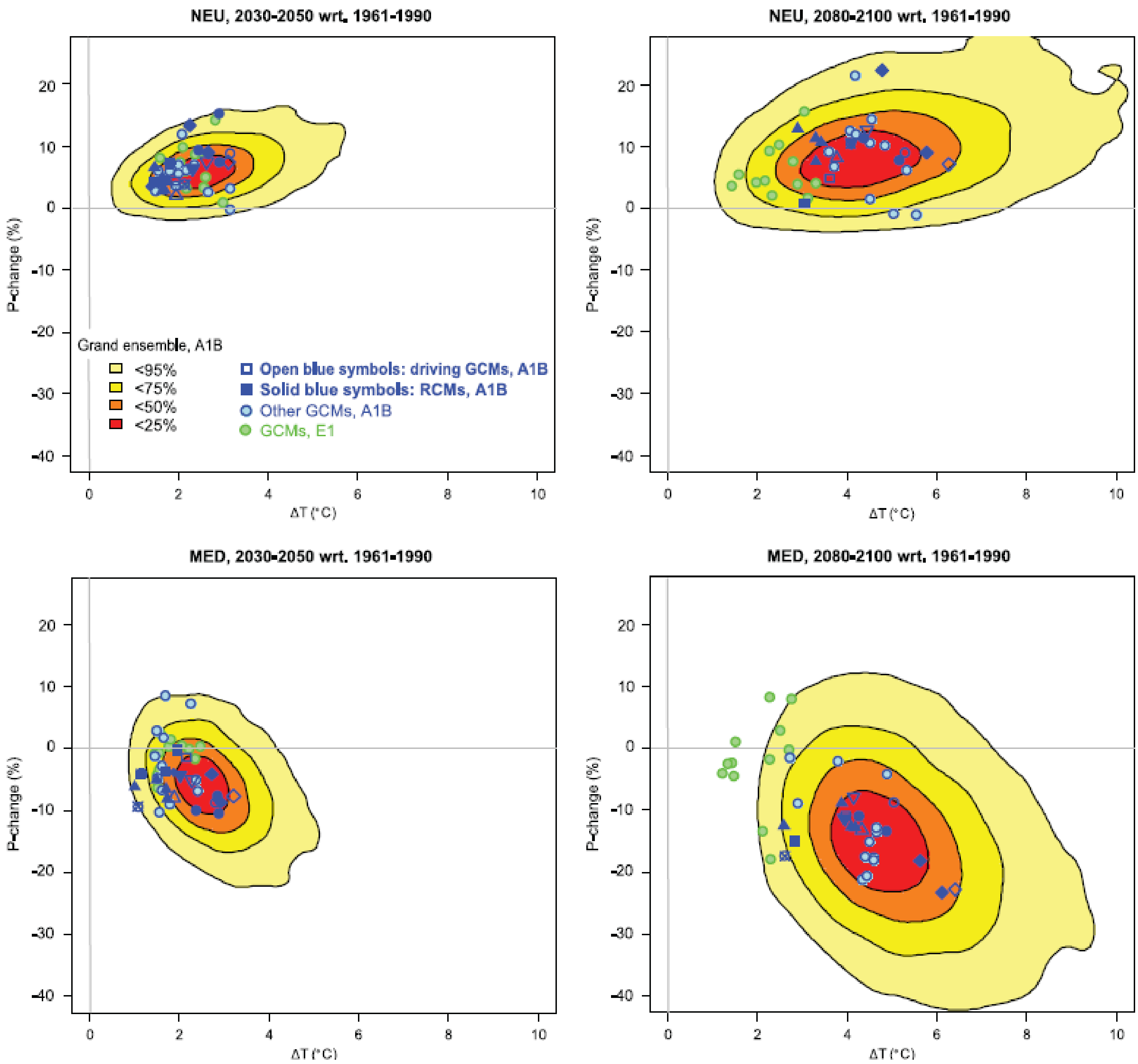

Figure 2. Annual changes in temperature (T) and precipitation (P) in northern Europe (NEU, top panels) and the Mediterranean Basin (MED, bottom panels) by the periods 2030-2050 (left) and 2080-2100 (right) relative to 1961-1990. Coloured areas depict probabilistic projection percentiles based on a statistical emulation of various sources of uncertainty from models and observations for the A1B scenario. Dark blue symbols are projections from RCM (closed) and their driving GCM (open) simulations for the A1B scenario (atmospheric concentration of ~700ppm CO_2 only in 2100); light blue symbols are other GCM simulations for A1B; green symbols are GCM simulations for the E1 mitigation scenario (stabilisation at 450 ppm CO_2-equivalent after 2100). Source: van der Linden and Mitchell (2009).

Other sub-sections are in preparation......

3. Examples of applying multi-model ensemble climate projections in impact assessments

This section describes some of the approaches examined in ENSEMBLES for assessing impacts of climate change from the multi-model ensemble climate projections described in Section 2. Figure 3 illustrates the alternative approaches explored in different impact studies.

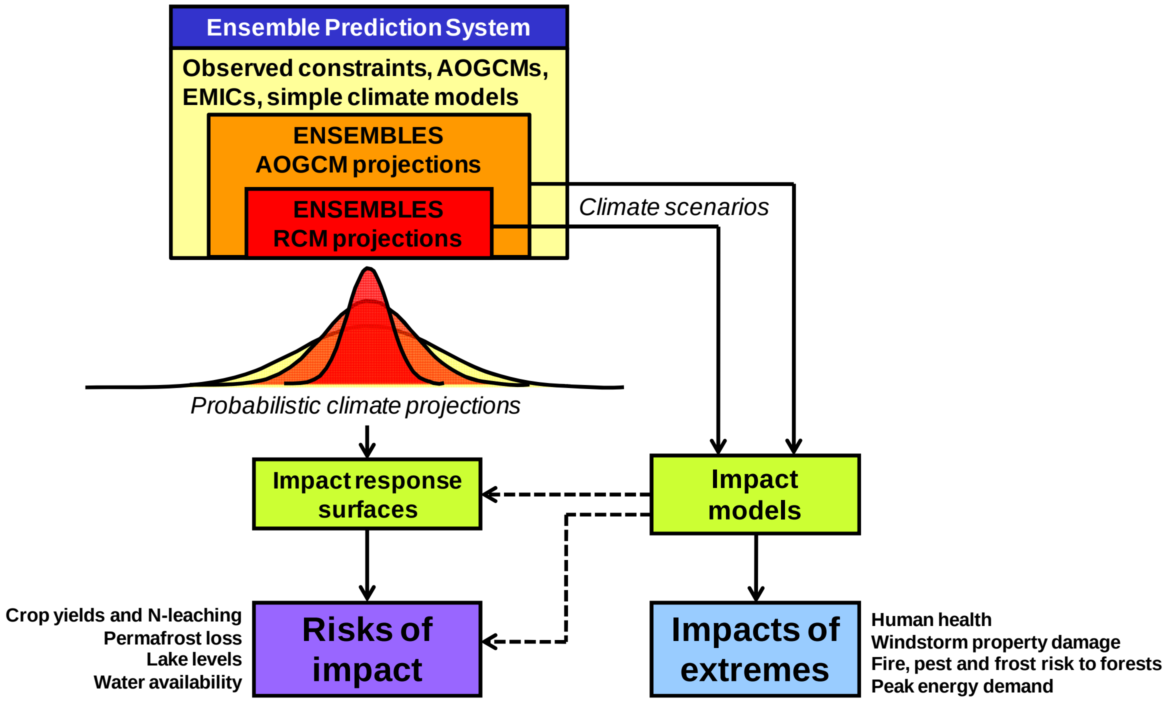

Figure 3. Approaches to impact assessment adopted in ENSEMBLES using outputs from the Ensemble Prediction System. One approach applied individual climate scenarios from the multi-model ensemble of climate projections as input to impact models. This approach was used in several studies of the impacts of extreme events (right hand pathway). It can also be used to evaluate the probability of exceeding impact thresholds (dashed arrow leading to Risks of impact). The second approach also seeks to evaluate impact risks, using probabilistic climate projections combined with impact response surfaces (left hand pathway). Impacts studied in Work Package 6.2 are listed alongside the respective approach applied. Source: Morse et al. (2009).

Two distinct approaches to assessment can be identified: multiple scenarios and impact response surfaces. The multiple senarios approach is used directly to assess impacts in several ENSEMBLES studies, and is illustrated in the next section. Examples of the response surface approach are included in section 3.2. Section 3.3 then describes how the multiple scenarios approach can also be used as a quality check of the response surface approach. Finally, some other issues relating to impact uncertainties are raised in section 3.4.

3.1. Multiple scenarios approach

Most of the impact studies undertaken in ENSEMBLES followed a conventional *multiple scenarios* approach involving the application of a representative set of climate scenarios drawn from the ensemble of model outputs provided in the project. In Work Package 6.2, a several studies focused on the impact of extreme events, which are expected to be responsible for some of the most damaging and costly impacts of future climate change.

One of the drawbacks of the multiple scenarios approach is the arbitrariness of scenario selection and application (commonly due to limitations on availability of scenarios and on resources to undertake multiple simulations with impact models), which may compromise the representativeness of the scenarios. In this context, the study by Donat et al. (2010) is of interest because not only does it seek to evaluate wind damage potential in Europe using a scenario-based approach, but it also treats the uncertainty in the multi-model ensemble of scenario-based results probabilistically.

GCM and RCM scenario simulations produced in RT2A and RT2B were analysed with respect to future changes in wind-storm risk and related loss potentials. In most simulations, as well as in the ensemble mean of multi-model simulations, increased extreme wind speeds were found over northern parts of central and western Europe under increased greenhouse gas forcing. Decreased values of extreme wind speeds were projected for southern Europe. Storm loss potentials were calculated by applying a storm loss regression model.

Consistent with the changes in extreme wind speeds, higher storm losses were estimated for western and particularly for central Europe, assuming that no adaptation to the changed wind climate takes place. Uncertainties in the range of changes in loss potential are accounted for by using two different measures. First, the standard deviation of the change signals across the different climate model simulations have been computed, revealing values of the same order as the mean changes for most regions considered. However, as an uncertainty measure, the standard deviation is strongly influenced by outliers.

An alternative measure of uncertainty considers the arbitrariness of the multi-model ensembles used in the study. There are numerous combinations of model outputs that might have been selected as ensembles among the nine GCM and eight RCM simulations, in addition to the nine member and eight-member ensembles. For example, if only eight of the nine GCMs had been chosen as ensembles, there are nine possible combinations of these; selecting ensembles of seven GCMs from nine gives 36 possible combinations, and so on. Overall there are 511 possible ensemble combinations comprising between one and nine GCM members. Consideration of the signals from all of these combinations results in a relatively symmetrical distribution of possible change signals around the ensemble mean and further allows for the construction of probabilistic information about the range of expected changes (Figure 4). For example, extending this analysis to include additionally the eight RCM simulations produces 131,071 different sub-ensembles based on seventeen individual model simulations. Using this distribution, as calculated for grid boxes over Germany for the end of the 21st century, a mean increase in loss potential of 25% can be estimated, with a 90% confidence interval of between +13% and +37%.

Note, however, that these probabilities are conditional on the range of projections provided in the ENSEMBLES project. This range is not as wide, for example, as the range sampled to produce the "grand ensemble" PDFs displayed in Figure 2.

In principle, the multiple scenarios approach can be applied to any impact model, regardless of its complexity, as long as climate projections are available at the resolution and time step required to adjust the input climate variables. Indeed, additional flexibility can be offered by attaching tools such as a stochastic weather generator, which samples from the multi-model ensemble (or a PDF representing this) to represent scenarios of changes in daily or sub-daily variability (e.g. see Jones et al. 2009).

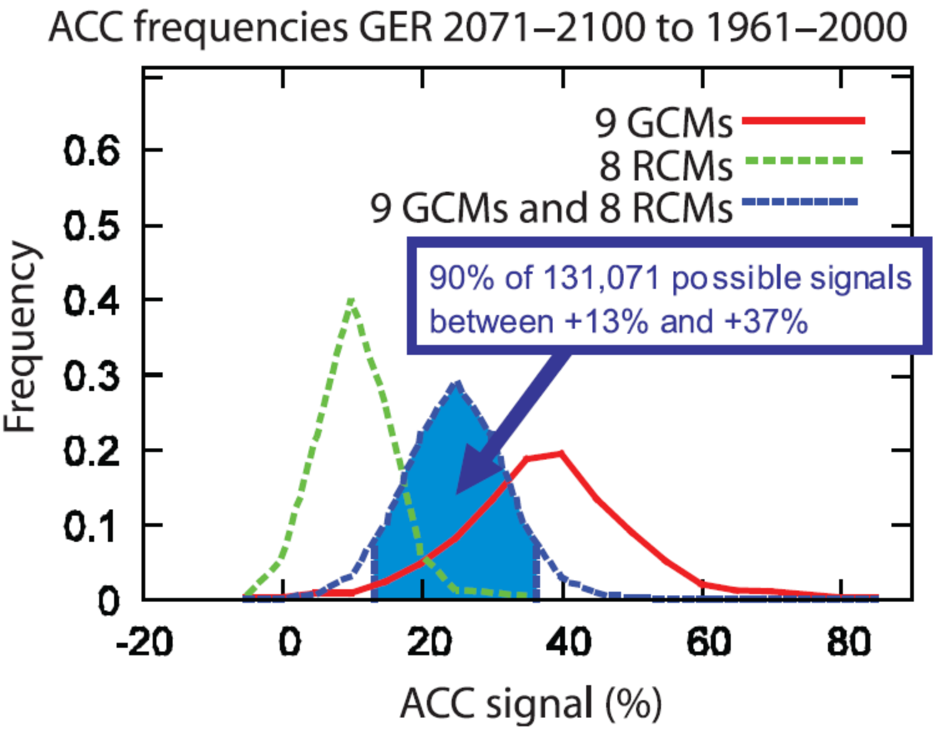

Figure 4. Anthropogenic climate change (ACC) signal in storm loss potential over Germany by 2071–2100 relative to 1961–2000 under an A1B emissions scenario, expressed probabilistically based on ensembles comprising all possible combinations of available ENSEMBLES simulations by GCMs (red curve), RCMs (green dotted curve) and all models (blue dotted curve). Source: Morse et al. (2009)

With a large number of impact model simulations for multiple scenarios, PDFs of the results can be generated to depict the probability of exceeding a pre-specified impact threshold. However, this is also disadvantage of the approach, as all climate projections have to be run through the impact model, which may be impractical for many complex models. This is the rationale for evaluating an alternative *response surface* approach.

3.2. Response surface approach

A number of techniques were described in Section 2 for representing the ENSEMBLES multi-model projections probabilistically as PDFs. The technique that was selected for use by impact modellers in ENSEMBLES was developed at the Met Office Hadley Centre (MOHC) and is illustrated in Section 2.6, above. Joint probability distributions (PDFs) of temperature and precipitation changes for the SRES A1B emissions scenario[3] were provided for regions of Europe specified by the impact analysts for 20-year periods relative to 1961-1990 at decadal intervals throughout the 21st century. These were derived from a large ensemble of GCM projections designed to sample uncertainties in key Earth system processes through a perturbed parameter approach, combined with results from alternative climate models and a multivariate set of observational constraints measuring the relative performance of alternative model variants in simulating a variety of aspects of present day climate and historical climate change (Murphy et al. 2009). These PDFs were sampled using a Monte Carlo procedure to produce a large set of individual estimates (10,000) of seasonal or annual mean temperature and precipitation change for individual GCM grid boxes across Europe, and also for a set of

[3] A1B is a moderate emissions scenario in which atmospheric greenhouse gas concentrations reach about 850 ppm CO_2-equivalent by 2100 (IPCC, 2007).

aggregated regions. These are referred to as "grand ensembles" in the remainder of this report, acknowledging the many sources of uncertainty that these PDFs attempt to capture.

The *response surface approach* that was employed to make use of these PDFs, comprises three steps:

1. The creation of a reduced-form impact model driven by a minimum number of climatic variables. Some future impacts may be sensitive to a range of different climatic factors as well as other non-climate factors (e.g. environmental or socio-economic factors that may mediate the effects of climate) and complex impact models often account for these multiple input variables. This step is an attempt to simplify the effects of climate by considering only a few (typically two) key variables that are of most importance in determining impacts and for which projections are available probabilistically. A variant of this approach is to retain the original (complex) impact model but to depict its responses to climate change in relation only to those climate variables considered to be most influential. Step 1 may also require that simplified information about future climate be used as input to the impact model, for example by making assumptions about the seasonal pattern of changing climate rather than by applying detailed monthly adjustments according to a climate scenario.
2. A sensitivity analysis is conducted of the impact model with respect to the key climatic variables chosen in Step 1. This commonly involves adjusting the observed baseline climatic inputs according to systematic increments that span the foreseeable range of future climate change. The increments of the adjustments should be chosen to be small enough to represent possible non-linearities in response to a changing climate. Impact studies in ENSEMBLES undertook sensitivity analyses for two climate variables (temperature and precipitation), though averaged over different time periods depending on the application. Impacts are computed for each combination of changes in the two climatic variables and plotted as a two-dimensional impact response surface (e.g. of crop yield, runoff, nitrogen leaching).
3. The information about the sensitivity of impacts to climate change contained in the impact response surface is combined with projections of the likelihood of such climate changes represented by the joint PDF. By superimposing the grand ensemble joint PDF onto the impact response surface, the risk of exceeding a given level of impact can then be evaluated. Threshold levels of impact are sector-, system- and region-specific, and selected to illustrate possible levels of tolerance to climate change, the exceedance of which may be regarded as unacceptable by decision makers. Thresholds were defined in ENSEMBLES either based on historical impacts, or according to established operational conditions (e.g. minimum stream discharge for hydroelectric production).

As a test of the response surface approach, estimates of the risk of exceeding an impact threshold can be compared with the equivalent estimates obtained using the multiple scenarios approach for those situations where application of the latter approach is feasible (see Section 3.2.3). One example of an application of the response surface approach is illustrated here, for estimating the risk of disappearance of permafrost features from northern Fennoscandia.

Palsas are mounds with a permanently frozen core that occur at the edge of the permafrost zone across the circumpolar north. They are ecologically valuable for birds, representing a priority habitat listed by the EU Habitats Directive. There is evidence that they are already disappearing and this is presumed to be due to observed regional warming.

Figure 5a is an impact response surface showing the area climatically suitable for palsas with respect to changes in two key climate variables, mean annual temperature and mean annual precipitation, relative to the 1961-1990 observed climate. A critical threshold of impact was selected as the total disappearance of suitable area for palsas (white area in Figure 5a).

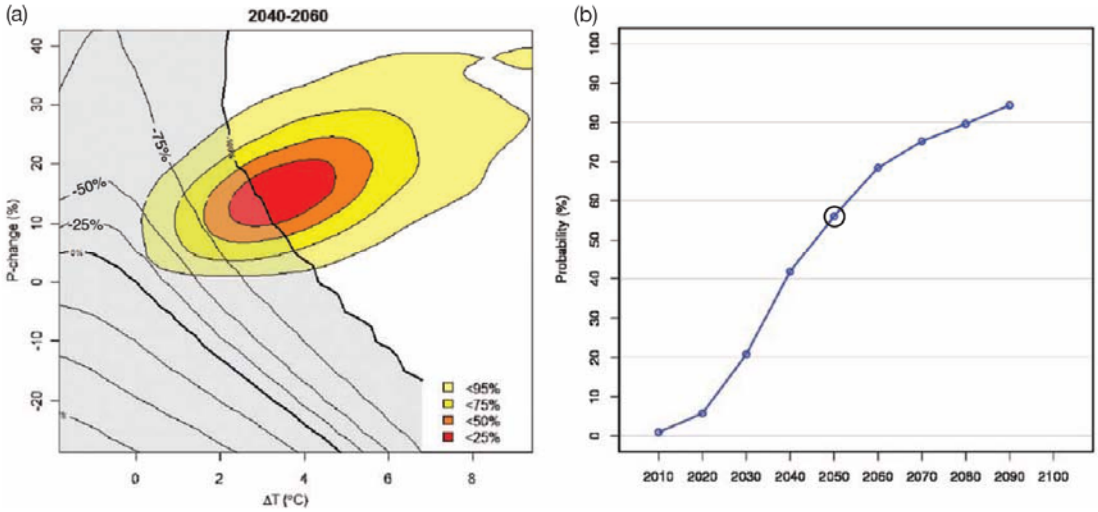

Figure 5. (a) Impact response surface showing modelled percentage change in area of suitability for palsa mires in Fennoscandia as a function of changes in mean annual temperature and precipitation relative to 1961–1990 (isolines). A joint PDF of future climate by the period 2040–2060 (coloured areas) is superimposed, about half of which lies in the area exceeding the −100% isoline (i.e., complete loss of palsa mire suitability). (b) Probability of complete loss of palsa suitability at different future time periods. The value for 2040-2060 taken from (a) is circled. Source: Morse et al. (2009).

Scatter plots of projected mean annual temperature and precipitation change over Fennoscandia for the A1B emissions scenario were obtained from the MOHC and presented as climate surfaces for 20-year periods at decadal intervals throughout the 21st century. These were then superimposed on the impact response surface (Figure 5a). The climate surface falling in the area beyond the -100% isoline represents the risk of total disappearance of palsas at different time periods in the future. Figure 5b indicates that it is *likely* (>66%) that all suitable areas will disappear by the end of the 21st century under the A1B emissions scenario.

3.3. Comparing the multiple scenarios and response surface approaches

In order to test the accuracy of the response surface approach, it is necessary to compare its estimates of impact risk with corresponding estimates made on the basis of simulations with the original impact model using the multiple scenarios approach. The palsa mire study was the only one that attempted this, using an earlier, preliminary version of the MOHC probabilistic climate projections. This is based on a perturbed-physics experiment with a single GCM and has a narrower distribution than the final "grand ensemble" presented in Figure 5.

The original palsa mire impact model was used to simulate all 12900 members of the climate change ensemble for the period 2010-2030. For each ensemble member, the presence/absence of palsas was estimated on a 10'x10' grid north of the Arctic Circle. The distribution of these impact outcomes identified areas of increased risk of palsa mire disappearance in northern Fennoscandia

The change in area suitable for palsa mires in relation to the suitable area of the baseline period was calculated and compared to estimates derived using the response surface approach. This gives estimates of the change in suitable area, but provides no information about the exact location of changes within the study area. Two versions of the response surface were constructed with the same palsa mire impact model. One version assumed the same temperature change in all months, whereas the second version weighted monthly temperature changes according to the seasonal pattern of temperature changes estimated from a 6-GCM ensemble average.

Probabilities of changes in the total northern Fennoscandian area suitable for palsa mires were higher for the response surface approach than the multiple scenarios approach. However, the discrepancy was insignificant when a seasonal pattern of temperature adjustments was applied, whereas the overestimation was substantial when adjustments were uniform throughout the year (Figure 6). According to these calculations, the risk of total loss of suitability for palsa mires (change of -100%) is estimated to be approximately 10% by 2010-2030.

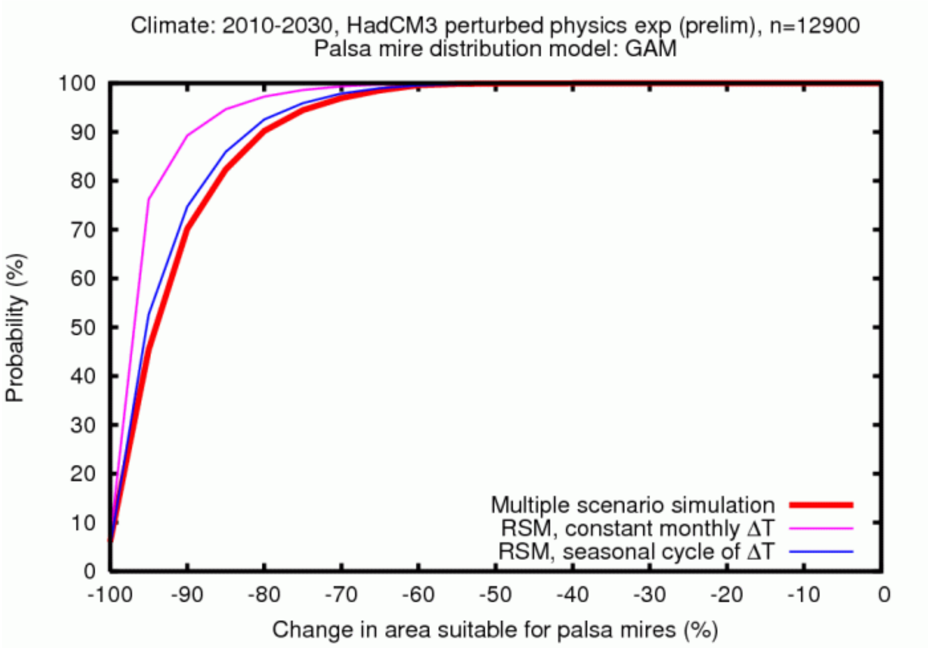

Figure 6. Cumulative probability of change in northern Fennoscandian area suitable for palsa mires for 2010-2030 relative to 1961-1990 using the preliminary MOHC perturbed physics experiment dataset (n=12,900) with the GAM-version of the palsa model simulating multiple scenarios and with two versions of the impact response surface, constructed using constant monthly and seasonally varying temperature adjustments, respectively.

3.4. Further developments of the response surface approach

The ENSEMBLES project has attempted to evaluate the utility of a response surface approach for applying probabilistic climate projections directly for assessing impact risks in Europe. Overall, the results indicate that an impact response surface, once it has been constructed and found to give reliable estimates of the impact variable, can be used in conjunction with a joint frequency distribution of projected climate variables to provide estimates of the risk of a given impact that replicate satisfactorily estimates obtained using simulations with multiple scenarios. No further simulations with the original impact model are necessary, which can significantly reduce the number of simulations needed to conduct an impact assessment compared to simulating all sample members. This can be of particular importance when applying more complex impact models, for which multiple simulations may be precluded on practical grounds. The sample size of the climate change ensemble also directly affects the feasibility of the multiple scenario simulations approach. For instance, the method adopted by the MOHC to construct probabilistic climate change projections for ENSEMBLES, produced sample sizes of thousands.

Overlaying a frequency distribution of future climate onto an impact response surface is also an effective visualisation technique for illustrating the relative likelihood of a given impact. For instance, by animating plots of the joint temperature and precipitation change PDFs superimposed on an impact response surface through time (e.g. at decadal time steps through the 21st century), a powerful impression can be conveyed of how the risk of impact changes through time.

Nevertheless, the impact response surface approach may not be appropriate in all cases. Estimates by some impact models may be critically dependent on several climatic variables, requiring the construction of response surfaces in multi-dimensional space. Conducting a sensitivity analysis of an impact model with respect to combinations of changes in more than two climatic variables would increase the number of required simulations and greatly complicate the visualisation and interpretation of the resulting response surface. On the other hand, constructing two-dimensional impact response surfaces with an impact model driven by more than these two variables requires a simplification of the model or an assumption of how the two selected variables can be related to the full set of required input variables. For example, estimates of impact probabilities for palsa mire disappearance were significantly affected (and improved) by introducing seasonality into the adjustment of annual temperature changes (cf. Figure 6).

It is important to emphasise that no study in the ENSEMBLES project considered the uncertainties of the impact models being applied. Logically, parameter uncertainties in an impact model might be be represented on an impact response surface as an envelope around each of the plotted isolines. In the case of structural uncertainties, it would probably be necessary to construct different response surfaces for each impact model being considered. The calculation of impact probabilities should then account for these additional sources of uncertainty.

A further development of the approach would be to make use of mathematical descriptions of both the impact response surface and the climate projection PDF. With these two surfaces expressed as equations, the overlap area could be solved analytically to give the probability of exceeding a given impact threshold. The method adopted in impact studies, to date, has been to count the points sampled from the MOHC joint PDFs in the overlap area.

4. Discussion and recommendations

Draft, in preparation......

- What have we learnt from these exercises?
    - Developing and using probabilistic projections is not easy
    - Time series are still needed
    - Communication is vital
- What guidance can we offer to climate researchers wishing to represent future uncertainties in climate projections probabilistically?

- What specific assessment criteria can we offer for testing the credibility of probabilistic climate projections?
    - First need to define 'credible'!
    - Difficult (impossible?) to define quantitative criteria
    - Important to get people to ask right questions, e.g., when they're presented with a PDF (how was it constructed, what uncertainties are sampled, what are assumptions, what is it conditional on?, etc.)

- What guidance can we offer IAV researchers wishing to express uncertainties in future impacts in terms of risk?
    - Recognise that these are subjective and conditional probabilities
    - Understanding uncertainty and increasing reliability/confidence doesn't necessarily result in narrower PDFs
    - In cases where impact models display a strong response to a few key climate variables, a response surface approach can be used to apply probabilistic climate projections directly for assessing risks of exceeding thresholds of impact
    - May need to consider other uncertainties/sources of information which can't be incorporated in PDFs – e.g., feedbacks, large-scale discontinuities - if taking a cautious approach to risk (Box 9 in Figure 1)

- What new research is needed to improve the representation of future climate uncertainties, to deliver useful products to the IAV community and to refine approaches for applying this information in IAV assessments?
    - More systematic application and testing of techniques developed in ENSEMBLES
    - Still a number of outstanding issues with respect to weighting – e.g., more extensive testing of sensitivity to using weighting is needed, add in GCM weights (and impacts models?)
    - Fill the matrix – using scaling and/or more RCM simulations
    - Try to answer some of the 'difficult' unanswered and unasked questions, e.g., was ENSEMBLES right to focus on spatial resolution rather than uncertainty, i.e., would it have been better to fill the matrix at 50km resolution rather than only partially fill it at 25 km?
    - Say something about opportunities provided by new scenario exercises being conducted ahead of the IPCC AR5 (including CORDEX).

Some of the recommendations from D2B.18 may also be relevant here (with some re-wording):

- More illustrative, working examples of probabilistic outputs should be made available to allow users to explore the utility of different output formats. (e.g., http://www.cru.uea.ac.uk/projects/ensembles/crupdfs/).

- The potential for integrating the facilities offered by existing tools for processing climate data (such as Climate Explorer, the ENSEMBLES R software for extremes and the STARDEX extremes indices software) with ENSEMBLES regional scenario outputs should be explored.

- Wherever possible, ease of access to data archives should be facilitated by, for example, using standard and consistent output formats (in terms of variables, domains, grids etc.). Use of OpenDAP and other methods for allowing remote and transparent use of large data archives should be encouraged.

- The regional scenarios web portal should continue to be updated and expanded as a resource for developers and the more technical users of regional projections.

- Further development of the web-based statistical downscaling service should be supported and, if possible, additional funding sought.

- Concerted effort is needed to address outstanding methodological issues such as weighting and PDF construction. The questions posed for the Prague cross-cutting workshop on Weighting, credibility, reliability, together with those discussed at a joint WCRP/ENSEMBLES/NCAR working lunch in July 2007, are very relevant here.

- RT2B cannot be expected to directly meet the needs of all users, particularly those outside ENSEMBLES, but can provide the building blocks, i.e., access to datasets together with appropriate guidance on methodologies and their use and, where possible, practical tools. User-driven guidance, including underlying assumptions and caveats, is essential to prevent any such tool being used as a black box.

- Thus there should be as much emphasis on the development of recommendations and guidance (e.g., deliverable D2B.26) as on the projections themselves (e.g., deliverable D2B.33).

5. References

To prepare.....

Appendix 1: Special Issue of *Natural Hazards and Earth System Science* (Status, May 2010)

Title: Applying ensemble climate change projections for assessing risks of impacts in Europe

Guest Editors: Timothy Carter (SYKE), Gregor Leckebusch (FUB), Jørgen E. Olesen (AU)

Papers

1. *Carter, Leckebusch, Olesen* Editorial **[In preparation]**

Section I: Models, methods and climate data

2. *Carter, Leckebusch, Olesen, Bindi, et al.* Applying multi-model climate projections in impact and adaptation assessment: investigations in the ENSEMBLES project **[In preparation, using D6.14]**
3. *Goodess, Carter, Klein Tank, Royer, Cubasch, Hoeschel, Voldoire, Christensen, Rummukainen, Jacob, Lorenz,* et al. Preparation, analysis and delivery of the ENSEMBLES regional climate projections and observations for use in impact and adaptation assessments **[In preparation, using D2B.26/D6.13]**
4. *Harris, Collins, Sexton, Murphy* Probabilistic predictions for 21st century European climate. **[Submitted]**

Section II: Estimating impacts of extreme weather events based on ensemble climate projections and their implications for adaptation

5. *Giannakopoulos, Psiloglou, et al.* Multi-model projections of changes in heat wave mortality risk in Athens during the 21st century **[Submitted]**
6. *Jönsson, Bärring* Using regional climate projections for assessing the risk of forest damage in a changing climate (a short review article) **[Submitted]**
7. *Donat, Leckebusch, Wild, Ulbrich* European property damage potentials: development and application of a simple storm regression model to global and regional multi-model projections **[Submitted]**
8. *Jylhä et al.* Aspects of climate extremes in Finland under recent and projected climate **[Submitted]**
9. *Szwed et al.* Applying agricultural, hydrological and human health indices to evaluate the changing risk of weather extremes in Poland based on multi-model ensemble climate projections **[Submitted]**

Section III: Techniques for evaluating probabilistic impacts of climate change using ensemble climate projections

10. *Børgesen, Olesen* A probabilistic assessment of climate change impacts on winter wheat yield and nitrogen leaching in Europe **[Submitted]**
11. *Fronzek, Carter* A comparison of methods for the probabilistic evaluation of climate change impacts on subarctic palsa mires in northern Europe **[In preparation]**
12. *Weiß, Alcamo* A probabilistic assessment of climate change impacts on monthly water availability in Europe **[Submitted]**
13. *Graham, Wetterhall, Andréasson, Rosberg, Yang* Using ensemble climate projections to assess probabilistic hydrological change in the Nordic region **[Submitted]**
14. *Ferrise, Moriondo, Bindi* Probabilistic assessments of climate change impacts on agricultural crops in the Mediterranean region **[Submitted]**

15. *Carter, Leckebusch, Olesen.* Postface **[In preparation, using D2b.26/D6.13 and D6.14]**