



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

Deliverable D2B.14: Recommendations for the modification of statistical downscaling methods for the construction of probabilistic projections

Due date of deliverable: November 2006

Actual submission date: October 2007

Updated: November 2008

Start date of project: 1 September 2004

Duration: 60 Months

Organisation name of lead contractor for this deliverable:
Fundación para la Investigación del Clima (FIC)

Revision [3]

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)	

D2B.14: Recommendations for the modification of statistical downscaling methods for the construction of probabilistic regional projections

Version 1: Luis Torres (FIC), Jaime Ribalaygua (FIC) and Fidel González Rouco (Universidad Complutense de Madrid) (March 2007)

Version 2: Clare Goodess, UEA (October 2007)

Version 3: Clare Goodess, UEA (November 2008)

1. AIM, SCOPE AND ISSUES

The main aim of this deliverable is to provide a focus for discussion of methodological issues that have not been previously or widely addressed and to make recommendations on the modification of Statistical Downscaling Methods (SDSMs) to produce probabilistic regional projections. Given the ‘newness’ of the issues considered, the discussion here is largely theoretical. Existing SDSMs with which the authors are familiar are used to illustrate issues, but quantitative, worked examples are not generally presented. The latter will emerge from discussion and implementation of the recommendations during the last two years of the ENSEMBLES project. All RT2B partners were invited to provide inputs to this deliverable, but it is primarily the work of the FIC authors, with some additional material provided by UEA, who also edited the report.

Probabilistic projections are needed in order to take into account and quantify the uncertainties associated with climate change predictions. The so-called ‘cascade of uncertainties’ arises from different sources. For the purposes of this report, these uncertainties can be classified according to three groups:

- A. Uncertainties ‘previous’ to downscaling: i.e., what will be the low-resolution atmospheric configuration in the future? (not for a specific date in the future, but the frequency of occurrence of each configuration)?
- B. Uncertainties related to downscaling: e.g., if the low-resolution atmospheric configuration for a day is a ‘certain one’, what will be the high-resolution surface effects?
- C. Uncertainties ‘downstream’ of downscaling: i.e., what will be the impacts of the projected changes on human and natural systems?

The first group encompasses a number of uncertainties associated with all the scientific and technical steps taken in the simulation of climate at large scales. They include uncertainties in emission scenarios and the resulting greenhouse gases concentrations and uncertainties related to the design of GCMs (i.e., resolution, time stepping, parameterizations, etc), together with physical processes – both those which are generally not yet widely taken into account in climate simulations, such as carbon cycle effects, and ‘unknowns’ such as the future evolution of natural forcings. All these uncertainties which affect our understanding of future climate at the large scales simulated by GCMs have been the concern of a considerable number of

studies, including ENSEMBLES, and some strategies have already been developed to address and quantify them. We suggest that these can be broadly grouped into frequentist approaches (Dubrovsky et al., 2005; Piani et al., 2005; Tebaldi et al., 2006; Schmidli et al., 2007), i. e. multi-GCM, ensemble strategies, different emission scenarios, and Bayesian methods (Kennedy and O’Hagan, 2001; Murphy et al. 2004; Tebaldi et al. 2004; Benestad, 2005; Greene et al., 2005; Katz, 2005; Tebaldi et al., 2005). These uncertainties and associated methods are a major concern across the ENSEMBLES project, but not exactly the main concern of this particular report.

Less attention has been paid to the second group of uncertainties, although it is also very important, particularly with respect to statistical downscaling and it is, therefore, the main concern of this report.

It is also true that not so much attention has been paid to the third group of uncertainties, although a couple of examples of a probabilistic end-to-end risk-based assessment for the UK water sector have recently been published (Wilby and Harris, 2006; New et al., 2007). These issues are, however, being addressed by other research groups in ENSEMBLES.

There is an “obvious” way of obtaining probabilistic downscaled projections: apply the downscaling to different GCMs-ensemble members-emissions scenarios, obtaining from each input a deterministic scenario, and obtain the probability density function (PDF) from the deterministic scenarios population. This follows the frequentist way (which can be modified to encompass a stochastic element as illustrated in Section 4.2), but there are also quite a number of studies with a Bayesian orientation, i.e., involving comparison with observations in one way or another to move from a *prior* to a *posterior* distribution (Tebaldi et al., 2004; Benestad, 2005; Greene et al., 2005; Katz, 2005; Tebaldi et al., 2005).

The main concern of this report, however, is to think about (and, if possible, set up recommendations for) the modification of SDSMs in order to produce probabilistic projections from a single input, considering and quantifying the uncertainties associated with the second question (B) above.

So, in this report we consider:

1. Uncertainties related to statistical downscaling
2. Ideas about how to address and quantify these uncertainties, and their modification when the SDSM is applied to future climate simulations
3. The need to produce probabilistic downscaled output from a single input (i.e., one ensemble member of one GCM, for one emissions scenario)
4. Recommendations for the modification of SDSMs to produce probabilistic output from a single input
5. (More briefly) how SDSMs can be used to produce probabilistic output from multiple or probabilistic inputs

There are several additional considerations, closely related to this report, but not exactly within its scope, that are raised below but which are not discussed further.

It would be very interesting, for example, to make a detailed comparative analysis of the magnitude of the different uncertainties, all over the cascade. The contribution of each source to the final global uncertainty could be explored using, for example, ANOVA methods

(Winkler et al, 2003). Depending on the contribution of the uncertainties related to downscaling, their consideration could be less or more important. The different sources of uncertainty may not be independent, and the final global uncertainty may not be just the addition of the individual uncertainties. These are the types of issues that will be explored during the last two years of ENSEMBLES as part of RT2B Task 2B.2.13 and reported in journal papers towards the end of the project.

The ultimate goal for many developers of probabilistic projections is to obtain a ‘global’ PDF that considers and quantifies all the uncertainties. This raises some very interesting issues, such as model weighting - in our case with respect to the performance, compared with present-day observations, of different SDSMs. A number of other ENSEMBLES deliverables discuss model weighting [e.g., D1.2: Systematic documentation and intercomparison of ensemble perturbation and weighting methods; D2B.8: Working paper on model weighting for the construction of probabilistic scenarios in ENSEMBLES; D3.2.1: Definition of measures of reliability based on ability to simulate observed climate in hind-cast mode; D3.3.1: Evaluated RCM-system for use in RT2B (choice of RCM-GCM combinations and preliminary RCM weights)], but these focus on how to weight GCM and RCM outputs rather than SDSM outputs. More work is clearly needed on the development of appropriate evaluation metrics and weighting schemes for SDS. Work is also needed on how then to use this information within a probabilistic framework.

The use of weighting schemes is based on the assumption that present-day performance provides a good measure of the credibility of future projections. It would, however, be more appropriate to view good present-day performance as a ‘necessary but not sufficient’ criterion for model credibility. In the case of SDS, specific new problems need to be considered. In particular, there are the stationarity and overfitting problems. Different SDSMs use different predictors and different relationships between predictors and predictand, and the stationarity of these relationships may also be different. The problem is that there is no “metric” to quantify this, and maybe the only approach possible is a theoretical analysis to quantify the stationarity problem for each method, paying attention to the predictors and the predictor/predictand relationships used, and whether they reflect physical linkages that should not change (see for example, the FIC contribution to STARDEX D10 http://www.cru.uea.ac.uk/cru/projects/stardex/deliverables/D10/D10_FIC.pdf), rather than being merely empirical results that could be non-stationary and overfitted. We will discuss later some more ideas about how to "quantify" the stationarity problem. Stationarity and robustness issues will also be explored during the last two years of ENSEMBLES as part of RT2B Task 2B.2.14 and reported in journal papers towards the end of the project.

In this report, we focus on PDFs as a way of representing the uncertainties. However, these may not be the most appropriate approach for all users or applications. Alternative ways of presenting probabilistic projections are discussed in ENSEMBLES deliverable D2B.18, together with user requirements.

2. UNCERTAINTIES RELATED TO STATISTICAL DOWNSCALING

Uncertainty analysis should not be viewed as a minor component that can be ‘added on’ once a SDSM has been developed, but should be an integral part of the development of any SDSM (Katz, 2002).

As explained in the previous section, the main focus of this report is those uncertainties directly related to downscaling and to SDS in particular. Thus a key question is, if the low-resolution atmospheric configuration is a ‘certain one’, what will be the high-resolution surface effects?

Whatever the temporal or spatial resolution used by the SDSM, a certain low-resolution atmospheric configuration used as predictor (for example, a geopotential height field at 12UTC of the problem day for SDSMs working on a daily basis, or monthly mean SLP) is not unequivocally or deterministically related to a certain high-resolution surface effect, but to a set of effects, more or less dispersed, due to the uncertainty associated with the downscaling procedure.

Most of the uncertainties involved in the statistical downscaling procedure can be grouped into two types:

1. Related to the spatial and temporal **resolution** of the input (i.e., GCM output):

- 1.1. Uncertainties due to **higher spatial-resolution structures** not resolved in the low-resolution configuration used as predictor: the same low-resolution configuration can have different high-resolution structures “inside”, that are not resolved at that low spatial resolution and which may have different surface effects. For example (for SDSMs working at the daily scale), the same "instantaneous" low-resolution configuration sometimes does and sometimes does not have small convective cells "inside", that can produce very high precipitation amounts.
- 1.2. Uncertainties due to **higher time-resolution phenomena** not resolved in the GCM output. For example, convective cells can last just a few hours, i.e., between two time steps of the GCM output, and will not therefore be evident in that output. We should distinguish here between the time stepping in the GCM and the temporal resolution of its output. The time step is typically 30 minutes or less, which is less than the few hours characteristic of convective events. However, we are talking here about the resolution of the input to the SDSM, i.e., the archived output of the GCM, typically 6 to (more usually) 24 hours.

Both types of resolution uncertainties are related since the spatial dimension of meteorological events is related to their temporal evolution.

2. Related to the statistical downscaling **method**:

- 2.1. Uncertainties due to **forcings not considered** by the method: i.e., is the parameter space of the predictors adequately sampled?
- 2.2. Uncertainties due to the **stationarity** problem: the relationships detected in the past may not remain the same in the future in the context of climate change
- 2.3. Uncertainties due to **overfitting**: the relationships detected in the past may be overfitted, and may give bad results when applied outside the calibration period
- 2.4. Uncertainties due to the **range of applicability** of the SDSM: even if the relationships detected in the past remain the same in the future, they may not be rigorously applicable in the future if the values of predictors fall outside of the population used to define those relationships (Bürger and Cubasch 2005)
- 2.5. Uncertainties due to the **overall underlying skill** of the method

- 2.6. Uncertainties due to the **spatial resolution of predictands**: if the spatial resolution of predictands is sparse, there will be little covariance structure in them. Thus some meteorological and climatological events might be represented at single sites and evolve unregistered in others. Such type of variability, or at least part of it, will likely be filtered out as noise and contribute to uncertainty. This poses the question of whether there should be certain spatial resolutions required of the predictand network in order to represent a given band of temporal scales, i.e., the shorter the timescales, the denser the network should be.

Some of the uncertainties may be related, e.g., with respect to 2.1 and 2.2 and uncertainties stemming from physical processes which are not actually taken into consideration. This has been discussed often in the case of GCMs, i.e., forcings which are not taken into account in some simulations such as aerosol and land use changes; and, mechanisms which are not modelled such as the carbon cycle and changes in the extent of land ice. Similar considerations should apply with respect to downscaling. For instance, the role of some of the variables which are not used to train the SDSM might be more important in the future perturbed climate (e.g., humidity and cloudiness). This example again reflects the need to consider the instability and stationarity of the downscaling relationships since such factors could actually contribute to having somewhat different/modified large-scale to regional/local-scale links in the future.

The issue of natural variability, i.e., interannual variability, also needs to be considered and is also relevant with respect to a number of the issues listed above.

Our aim should be to obtain a probabilistic output from the statistical downscaling of a single input, i.e., a PDF of the predictands which takes into account all the above uncertainties. However, it is recommended to analyse as much as possible each of these uncertainties individually to start with, because the knowledge of each is essential to improve the analysis of the global uncertainty (Katz, 2002). However, the extent to which each of these individual uncertainties can be quantified differs, as we will see in the next section.

3. HOW TO CONSIDER AND QUANTIFY STATISTICAL DOWSCALING UNCERTAINTIES

3.1 Ensembles, frequentist and Bayesian approaches

For the developer (and user) of probabilistic climate projections, one of the first questions should be to consider whether or not it is possible to consider all uncertainties. The likely answer is that it is not possible to consider all, at least not in the first instance. ENSEMBLES, for example, is focusing on the production of conditional probabilistic projections, i.e., conditional on a single emissions scenario with the main emphasis on the SRES A1B scenario.

The goal of ENSEMBLES is to work with multi-model ensembles. In this section, however, we reduce the factors that need to be addressed by considering only those uncertainties associated with the use of a single forcing model. Multi-model issues are discussed in later sections of this report.

For instance, if the approach taken to evaluate uncertainties is a frequentist one, this means that you shuffle a few factors in the production of your regional projections: for instance, you produce an ensemble of regional predictions by randomizing the inputs to the model, or selecting a variety of downscaling parameters (e.g., distance values in the case of analogues, number of modes in the case of EOF, SVD or CCA approaches, parameters in the neural network approach) and so on. This can be viewed as comparable to producing GCM ensembles by considering a range of models, physical parameters, etc. The resulting ensemble should thus illustrate the variability of possible results considering your unknowns.

If the approach taken to evaluate uncertainties is a Bayesian one, then the downscaled outputs are compared to observations, and the behaviour of the errors analysed. The spread of the errors provides an uncertainty assessment, which is not only due to considering one member of a potential statistical ensemble but also to all the factors/physics/mechanisms which are not taken into consideration in the approach by any members of the ensemble, since the errors should represent a 'true-comprehensive' error when compared to reality. Even so, this does not support the idea that one can take into account all uncertainties, since such evaluation provides different results depending on the temporal period considered and ultimately it is the result of comparing to the present and not the future climate.

So we have two different approaches to uncertainties assessment, the frequentist and the Bayesian. Both approaches consider temporal and spatial variations – which are interesting with respect to the deterministic approach to SDS (i.e., one value for each point/time), because they provide information about the predictability of the system. This approach is also interesting from a methodological point of view since SDSM "improvements" usually try to reduce deterministic errors in a validation period - with a probabilistic approach, the reduction of SDSM uncertainties will be an improvement itself, especially if it also contributes to reduce the uncertainties related to future climate change.

3.2 Potential approaches for analysing uncertainties

We will now consider each of the uncertainties identified in Section 2 and how they might be explored. As previously said, the quantification of each of these uncertainties is not equally feasible, because there is not always a universal objective metric to be used. In most cases, the effect of worsening the SDSM skill, or at best, the effect of some improvement (for example, an increase of spatial resolution) could be quantified, but there is not an objective way to quantify all the uncertainties (for example, due to the impossibility of working at a sufficiently high spatial resolution to resolve all the predictands forcings).

Regarding the spatial and temporal resolution of the input (GCM)

- Uncertainties due to **higher spatial-resolution structures** not resolved in the low-resolution configuration used as predictor: a first attempt to quantify these uncertainties could be to use the SDSM in hind-cast mode, applying it to Reanalysis output, firstly at its maximum spatial resolution, and then at the (lower) resolution used by the GCM whose "spatial resolution uncertainties" are to be quantified. The (expected) worsening of the validation results of this second application will give an idea of these uncertainties, although not all them will be considered, because the maximum resolution of currently-available Reanalyses is still far below the ideal resolution needed to capture structures that affect very much predictands. (If a GCM has higher resolution than the Reanalysis, a

second application should be done with a lower resolution). This approach will probably give only a first guess of this contribution to uncertainty, due to the limited resolution of the Reanalyses normally used for SDS (i.e., ERA40, NCEP/NCAR). Better assessments could be obtained using this "resolution modification approach" with higher resolution reanalyses (e.g., as now available for North America or obtained from RCMs).

- Uncertainties due to **higher time-resolution phenomena** not resolved in the archived GCM output: similar to the previous point, reducing the time resolution of the Reanalysis could give an idea of these uncertainties - for example, by comparing the validation results obtained applying the SDSM to six-hourly Reanalysis information with those obtained with daily information.

As said before, both types of uncertainties are related since the spatial dimension of meteorological events is related to their temporal evolution. This has to be considered in designing sensitivity studies to quantify and disentangle the two effects.

The limitations in the spatial/temporal resolution of GCMs used to drive SDSMs could be assessed by using both the output of the GCM and an RCM (forced by the same GCM) to drive a SDSM. This would help to understand the extent to which the limitations in the spatial/temporal resolution of GCMs play a role. The availability of ENSEMBLES ERA-40 forced RCM runs at 50 km and 25 km resolution, for example, allows downscaling to observations both using the Reanalysis and some dynamically downscaled versions of them which should include more convection and physics at smaller spatial/temporal scales. At least two RT2B partners (CFI and UC) plan to apply statistical downscaling to RCM outputs in the last two years of the project – work which should be facilitated by the expanding capabilities of the ENSEMBLES web-based downscaling service developed by UC.

Regarding the statistical downscaling method

- Uncertainties due to **forcings not considered** by the method: an objective quantification of the uncertainties due to predictors not considered is difficult (only subjective sensitivity analyses based on theoretical considerations can be done), but the effect of reducing or modifying the set of predictors can be analysed. Here appears again the frequentist/Bayesian discussion. You cannot calculate the uncertainty 'of not using' a given set of alternative predictors unless you consider the error when comparing to observations. Alternatively, if you increase the number of your predictors sequentially you would produce a frequentist ensemble of runs. That uncertainty is informative of all the range of possibilities that your method can produce (as in an ensemble of GCM simulations), but it is bounded by the first decision constraining the analysis, i.e., to use a given set of predictor variables and parameters. It does not contain information about the mechanisms that you did not take into account.
- Uncertainties due to the **stationarity** problem: this is a very important issue that should be carefully addressed. There is, however, no universal metric, so again, no fully objective quantification of these uncertainties is possible. As a first step, a careful theoretical analysis can be undertaken focused on whether the predictors/predictands relationships used are considered to capture physical linkages (that will not change), instead of being empirical results, that could be non-stationary (and/or overfitted).

In addition, some objective sensitivity analyses could be done (such as recommended in STARDEX deliverable D16 (www.cru.uea.ac.uk/projects/stardex) and proposed to be undertaken as part of ENSEMBLES Task 2B.14 during the final two years of the project. For example, validation of the SDSM using independent periods is crucial in order to see the SDSM sensitivity to the change in the climate regime of predictors. The use of different sets of calibration and validation periods should also be tested: for example, training the SDSM with wet or cold years, and validating it in dry or hot periods. This approach could be used in a more severe way: for example, using as the training period only elements belonging to the medium and low part of the predictand PDF in the reference (observational) period (for example, days with low / medium temperatures), and as calibration period the elements belonging to the higher part of it. In this case, the validation results would indicate the SDSM sensitivity to a clear shift and deformation of the predictand PDF, even more than that expected in a climate change context.

This stationarity problem can also be explored using model simulations as surrogate climates where the downscaling relationships can be studied in the instrumental period, and also how they are modified under the future climate change scenario simulations. This can be done using large scale predictors and a subset of grid-point time series at the regional scale within different periods of a climate simulation (González-Rouco et al. 2000, Frías et al. 2006). Within ENSEMBLES Task 2B.14, it is planned to calibrate SDSMs using GCM predictors and RCM predictands for the control period, then apply these relationships to GCM predictors for the future, finally comparing the SDS outputs with the RCM outputs for the same future period. This analysis is, however, based on the assumption that the RCM parameterisations adequately capture any non-stationarities

Another issue that could be discussed is that the downscaling relationships are actually never stationary, they vary with time as natural climate variability does. So, if you perform a downscaling analysis within the instrumental period, the resulting skill of the method varies depending on the validation period which is considered. This actually adds to the uncertainty and as such, it should be discussed. The issue is whether the assumption that the empirical relationships are constant with time is valid under certain statistical bounds in the context of climate change.

- Uncertainties due to **overfitting**: some of the previous considerations and suggestions apply here. And some additional techniques to avoid overfitting can be suggested, e.g., re-sampling over multiple validation/calibration periods or cross-validation with different step-periods.
- Uncertainties due to the **range of applicability** of the SDSM: it should be determined how frequently we are applying the relationships detected in the past, to the future with values of predictors outside of the population used to define those relationships. This could give an idea of the associated uncertainties. The SDSM could be out of its range of applicability not only because the predictor values for the future are out of the population used to define those relationships. We present the example of a two step analog method developed by FIC and used in STARDEX. In this method, the first step is the selection of the most analogous days, and the second step for temperature is a multiple linear regression (with forward and backward stepwise selection of potential predictors) performed using the analogous days population. This SDSM can be out of its range of applicability when the predictor value for the problem day is out of the cloud defined by the analogous days population, but also if the similarity of the analogous days, one to each

other and to the problem day, is clearly different in the future than in the validation period. The more different are the analogous days (one to each other and to the problem day), the bigger is the uncertainty associated with the simulation of that problem day predictand (because there is more variance among the analogous days used in the simulation). We will see later that this variance could give an idea of the increase (or decrease) of the SDSM uncertainties when it is applied to the future, with respect to the uncertainty quantified in the validation period.

- Uncertainties due to the **overall underlying skill** of the method: we could try to quantify these uncertainties in the validation phase, by comparing the downscaled predictands with the observed ones. Nevertheless, some of the previous uncertainties will be introduced all together, and it is not likely to be feasible to quantify and disentangle these uncertainties, especially as some of them may not be independent, in order to isolate the overall underlying skill, i.e., the skill we might expect from a perfect model and input information.
- Uncertainties due to the **spatial resolution of predictands**: predictands could be used at different spatial resolutions, with the validation results giving an idea of the resulting uncertainties. In ENSEMBLES deliverable D2B.12, for example, Valentina Pavan and co-authors demonstrate that statistical downscaling performance can indeed be very sensitive to the station density.

Though natural variability contributes to some of the uncertainties discussed above it is of a somewhat different nature. It also relates very much to the detection/attribution discussion and to the issue of discriminating a clear response/change signal of the predictand variable in the future climate. One way to quantify natural variability is to use output from very-long unforced GCM control runs. While this approach has been used in a number of climate change assessment studies, the further step of downscaling from such simulations has not, to our knowledge, been undertaken. Whether or not such an approach is feasible and/or worthwhile maybe worth some consideration. Certainly the whole issue of how to assess and represent natural variability in probabilistic SDS approaches needs to be more explicitly addressed by ENSEMBLES partners.

3.3 Considerations with respect to an analogue SDSM

We will now present an example of downscaling uncertainties identification, and the importance of their consideration (although in this case, they were not quantified). It is related to an analogue methodology, that was developed by FIC and modified within the STARDEX project, trying to capture the downscaling uncertainties in order to improve its skill for precipitation extreme indices (Goodess et al, 2007a).

This analogue SDSM selects the most "similar" days to the problem day, using a similarity measure optimised to detect as most similar low-resolution atmospheric configurations those with most similar high-resolution precipitation effects. The daily precipitation simulated for the problem day is obtained from the precipitation observed at that site, on the "n" most analogous days.

The downscaling uncertainty is easily detected, observing the "n" most analogous days to a certain problem day: those "n" days, having very similar low-resolution atmospheric

configuration (the reference data-set is quite long, so very similar days can be found), sometimes have very different surface effects, due to the uncertainties identified in Section 2. And this downscaling uncertainty is different depending on the problem day. For example, if the problem situation may have or have not "inside" convective cells, in some of the analogous days those cells appeared (producing heavy precipitation), and in some of them not (without precipitation). If the problem configuration is not compatible with lower resolution structures (frontal precipitation configurations, or dry anticyclonic situations...), the analogous days effects are very similar to each other.

In a first attempt, the problem day precipitation was obtained by averaging the analogous days precipitation. This gave large biases, especially for extreme indices (for example, underestimating pq90, 90th percentile of rainday amounts, mm/day), because the downscaling uncertainty was not considered, as all the analogous days were averaged thus smoothing, in particular, the extremes.

The method was then modified in order to consider all the analogous days. It was assumed that the probability of occurrence of each analogous day precipitation was $1/n$. The pq90 index was defined for each season (e.g., 92 days). The modified SDSM simply determined the pq90 value for that season as the value that was exceeded by the 10% of the $n * 92$ analogous days. Now, all the extreme events were considered, and no averaging was performed: i.e., the downscaling uncertainty was considered. Nevertheless, the predictand (pq90 in this case) was simulated as a categorical value, and no PDF of the predictand was constructed. So, the downscaling uncertainties were considered, although they were not quantified.

This simple modification produced a significant improvement in the extreme indices simulation, reducing very much the underestimating bias.

This general approach could be applied to other SDSMs: i.e., working at its own maximum time resolution, instead of providing one categorical value of the predictand, a set of different possible predictand values associated with the predictors could be provided. For example, in regression methodologies, the set of predictand values could be obtained considering the regression errors distribution (i.e., the variance inside the population used to calculate the regression). Some SDS approaches may imply the assumption of certain hypotheses about the statistical distribution of the predictands associated with a set of predictors (for example, of the regression errors). The analogue methodologies don't imply these assumptions, which might be considered as one advantage of such an approach.

3.4 Modification of SDSM uncertainties when applied to future climate simulations

We now try to present some ideas about how uncertainties may be modified, with respect to the assessments performed for the validation period, when the SDSM is applied to future climate simulations. Again, it is not easy to provide recommendations that could be applied across the very different types of SDSMs, and so we again consider the FIC analogue method that we know best.

In this analogue method, for precipitation, the variance of the n analogous days observed precipitation is a reflection of the SDSM uncertainty (due to the different sources described before: spatial and temporal resolution, forcings not considered...). From these observed precipitation values a PDF for each problem day and point can be obtained.

Regarding temperature, this uncertainty assessment is provided by the dispersion of the cloud of points of the multiple regression (of the selected predictors) performed using the most analogous days population: from this cloud of points, a PDF can be obtained for each problem day and point.

These uncertainty assessments are frequentist, but Bayesian assessments could also be done comparing observations to the downscaled PDFs obtained by driving the SDSM with Reanalysis in the validation period. A comparison of the two approaches would give an idea about how the frequentist approach described above is really assessing properly the uncertainties. This particular SDSM is, however, considered to assess uncertainties quite well, especially for temperature and for the validation period.

When the SDSM is applied to climate simulations (not associated with observations and therefore limiting the direct analyses possible), one can expect (at least for future climate simulations) that the clouds of points would be in general more disperse, producing "more flat" daily PDFs, that would imply greater uncertainty, compared to the application of the SDSM to Reanalysis outputs.

It would also be possible to analyse if the similarity of the analogous days (one to each other and to the problem day) decreases when the SDSM is applied to future climate simulations, with respect to its application to Reanalysis outputs. Such a decrease could be related to more "rare" problem days in the future, with analogous days less similar one to each other. Furthermore, the relationships between the similarity and validation errors could be analysed for the validation period. This is also related to some of the ideas pointed out before, because one can expect that more different analogous days should have more different predictand values and more disperse clouds, and therefore "more flat" PDFs (i.e., more uncertainty).

So, with an SDSM like the FIC two-step analogue method, one can in some sense assess the evolution of the uncertainties when applied to climate simulations, with respect to the uncertainties quantified in the validation period.

4. NEED OF PRODUCING PROBABILISTIC OUTPUT

4.1 Need of producing probabilistic output from a single input

While it is important that downscaling studies using only one driving GCM are eventually placed in a comparison framework with results of other assessments using different simulations, having PDFs associated with each single downscaling assessment would be more informative about the limitations of each particular study since different studies potentially arriving at very similar deterministic evaluations could be imbedded in significantly different ranges of uncertainty. This approach would not preclude eventually using the uncertainties associated with each assessment for a number of GCM simulations to integrate them into a single probabilistic evaluation. Such a 'disaggregated' approach has methodological implications, since advances in the downscaling application could stem not only from obtaining less errors in a certain validation period but also from reducing the uncertainty associated with certain parameter configurations.

While the ‘previous’ and ‘downstream’ uncertainties (see Section 1), must be addressed at some stage, advantages of producing probabilistic output from a single input include:

- It allows, not only consideration of downscaling uncertainties, but also allows their quantification.
- It allows, eventually, a more robust combination of different outputs (from different SDSMs, or from the same SDSM for different inputs (GCMs/ensemble members/emissions...)). The reason is that all the PDFs of each output can be combined, instead of combining single deterministic values. For example, imagine two outputs obtained from applying one SDSM to two GCMs for the same emissions scenario. One provides a certain increase and the other one a couple of degrees more. If the uncertainties of the second one are clearly larger (i.e., the PDF is wider), due, for example, to a lower spatial resolution (i.e., higher "resolution uncertainties"), the combination of both PDFs will not give the same result as the combination of the deterministic values.
- Quantifying uncertainties based on resampling all the parameter space (although this may be difficult to fully determine) of a given SDSM would be beneficial from the point of view of understanding the different ranges of variability/error associated with different SDSMs and their underlying assumptions. This type of uncertainty assessment would contribute to methodological progress in as much as the target of some studies in the future might not be a better prediction of mean values but a reduction of associated uncertainties which is perhaps more relevant in the climate change context: SDSM improvements could be obtained in terms of reducing uncertainty, even without diminishing deterministic errors.
- Quantifying uncertainties and their variation in time and space will be informative about the predictability of the system and its space-time variability, thus establishing new targets to improve the methodology.
- Quantifying uncertainties based on simulated-observed predictand comparisons will be informative about the optimal range of SDSM parameters to be chosen for certain variables/regions/seasons and hence help to reduce concerns about stationarity.

Some inconveniences are:

- It increases the work to be undertaken, so it has to be justified.
- It could happen that the procedures needed to obtain probabilistic output may give worse deterministic (i.e., the mean of the PDF) simulations for a certain downscaled index than using other strategies that optimise the deterministic output. It might, however, be possible to combine both strategies, e.g., forcing the PDF to represent the deterministic values obtained directly.
- The combination of outputs will be a bit more difficult (although eventually more robust, as discussed above).

Comparative analysis of the ‘previous’ and ‘downstream’ uncertainties (see Section 1) with the downscaling uncertainties considered here, may finally indicate that the latter are of less concern or even inconsequential compared to the other two at least for some parameters, seasons and locations. Nevertheless, the analyses proposed here are clearly very interesting from an analytical point of view, and for methodological and application purposes. And indeed such a comparative analysis cannot be done properly without a more exhaustive and quantitative analysis of the downscaling uncertainties. While quantification is desirable, it is not, however, always possible with respect to all the sources of uncertainties discussed here. In such cases, the experience of the STARDEX project (www.cru.uea.ac.uk/projects/stardex)

for example, indicates that advances can still be made e.g., (see Section 3.3) and interesting results obtained from theoretical and qualitative explorations of uncertainty.

4.2 Need of producing probabilistic output from multiple inputs

Given the overall aims of ENSEMBLES, the need to work with multiple inputs, i.e., multi-model ensembles, does not really need to be restated or defended here. Rather, it is a matter of developing and implementing appropriate approaches for doing this at the regional or local scale. Relatively few examples exist in the literature of probabilistic regional projections (Allen et al., 2000; Benestad, 2004; Tebaldi et al., 2004; 2005; Ekström et al., 2007; Fowler et al., 2007) and even fewer examples of their use in impacts studies (Luo et al., 2005; Wilby and Harris, 2006).

Here, we use the example of work undertaken by UEA in the CRANIUM (www.cru.uea.ac.uk/projects/cranium) and ENSEMBLES projects to illustrate some of the issues that need to be considered. In this work (Goodess et al., 2007b), UEA explored inter-model uncertainties in climate change projections using output from 10 different European RCMs and a stochastic weather generator (Kilsby et al., 2007) linked in a probabilistic framework. The RCM runs came from the PRUDENCE project – most were forced by HadAM3, but three were also forced by a different global model (ECHAM4 or ARPEGE), giving a total of 13 RCM runs, all for the A2 scenario. Changes in mean temperature and precipitation, together with changes in their variability, were taken from each RCM run for the grid square nearest to the location of interest and used to perturb the parameters of the weather generator. For each of the 13 RCM runs, the weather generator was run 100 runs – paired differences between control (1961-1990) and future (2071-2100) time periods were then used to construct PDFs and CDFs from all the output. For ENSEMBLES, this approach was applied to seven mainland European stations: Linköping, Karlstad, Saentis, Basel, Beograd, Kaliningrad and Timsoara. All results are available from this website: <http://www.cru.uea.ac.uk/projects/ensembles/crupdfs>.

While these PDFs and CDFs provide useful first examples of what station-scale probabilistic outputs may look like - and hence are very useful for communication purposes (see SKCC Briefing Note 2 available from www.k4cc.org; Goodess et al., 2007b), it is important to note the following features:

- They are conditional projections, i.e., conditional on a particular greenhouse gas emissions scenario (the A2 scenario) [ENSEMBLES projections will be conditional on the A1B scenario]
- They are based on a small sample of GCMs (most RCMs are based on the same model, with only two other GCMs considered)
- Unless indicated otherwise, the ensemble averages presented are unweighted (i.e., they do not take any account of differences in model performance for the present day)
- When a preliminary weighting scheme is applied, it is difficult to assess its influence, since so many RCMs are driven by the same GCM.

During the last two years of ENSEMBLES, this approach will be repeated using:

- The new ENSEMBLES RCM simulations from WP2B.1 which provide a much more evenly spread matrix of different GCMs and RCMs
- New weighting schemes produced by RT3 and others

Some methodological issues will also be explored, notably a consideration of the extent to which the stochastic weather generator variability can be considered as a ‘true’ representation of natural variability and how this variability interacts/compares with the variability from the driving GCMs/RCMs. The potential danger of double-counting sources of uncertainty in this probabilistic framework, and in the weighting scheme, will also be considered.

4.3 Need of producing probabilistic output from probabilistic inputs

An extension of using multiple inputs to produce probabilistic outputs would be to use PDFs as direct inputs to SDSMs and is a more challenging issue. The original ENSEMBLES description of work identified how “to generate scenarios based on the ‘grand probability’ distributions which will be constructed in RT1 and RT2A” as one issue to be addressed in the modification of SDSMs within a probabilistic framework.

The preliminary RT1 outputs produced by the Hadley Centre are for annual temperature and rainfall for aggregated regions or countries and are being used by WP6.2 to explore the development of response surface approaches to impacts assessment. At an earlier stage of the project, the Hadley Centre told downstream users that they were open to discussion on time periods, regions, variables etc and also indicated that ultimately they could look at the possibility of producing PDFs for circulation-related parameters/indices, e.g., NAO. However, it is not clear how the SDSMs being used in RT2B (see Table 1) could be adapted to handle such probabilistic circulation information. The most viable approaches are likely to be some sort of conditional weather generator (Palutikof et al., 2002) or conditional regression approach.

While using probabilistic circulation changes is problematic, the approach with better practical potential is to use a weather generator (e.g., Kilsby et al., 2007) to sample from a PDF of changes in the surface variable(s) of interest (e.g., daily temperature and precipitation). This would be an extension of the work undertaken by UEA in the CRANIUM and ENSEMBLES project (see Section 4.2), i.e., the weather generator would be calibrated on observations and the statistical parameters then perturbed using change factors derived from sampling a PDF, rather than from individual RCM runs.

This is the approach being taken as part of work on the development of the next national UK climate scenarios (UKCIP08 – www.ukcip.org – now likely to be released in Spring 2009 rather than November 2008 as previously planned). UEA has modified its stochastic weather generator (Kilsby et al., 2007) to sample from joint probability distributions, for 25 grid boxes constructed by the Hadley Centre, in order to provide daily and hourly time series output for individual 5 km grid squares across the UK. As of November 2008, the first set of final PDFs have just been provided to UEA.

Two key issues arise in using such an approach:

- Weather generators typically produce self-consistent series of multiple variables (required for applications such as agriculture and built environment impact studies), thus physically consistent change factors in all variables are needed, i.e., there is a need to sample from joint probability distributions. This is relatively simple where only two variables are used, but increases in complexity with the number of variables. In earlier versions of the UEA weather generator, for example, only temperature and precipitation were perturbed

(changes in secondary variables such as sunshine and relative humidity were driven only by the generated changes in temperature and precipitation). In more recent versions, change factors are also used for some secondary variables (including sunshine) to make the output more consistent with the underlying RCM changes. The approach that will be used for the UKCIP scenarios is described in a draft technical report that is currently undergoing expert review, and is confidential until release of the scenarios in Spring 2009.

- The second issue concerns how many times the PDF should be sampled. Weather generators, including the UEA one (Kilsby et al., 2007), are typically stochastic. Sensitivity experiments have indicated that it is optimal to run the UEA weather generator 100 times for each set of change factors. Further work is needed to determine how this stochastic spread relates to the uncertainty range expressed by the PDF. A minimum recommendation might be to sample the 10th, 50th and 90th percentiles of the PDF. But even then, if the user has a complex impacts model, it may not be possible to use all 300 series (i.e., 100 runs for each set of change factors). In this case, the user could sample randomly from the 100 runs, or use the approach taken in CRANIUM (www.cru.uea.ac.uk/projects/cranium). Here, a single series was selected by ranking all 100 series on the basis of the annual number of rain days (precipitation is the primary weather generator variable on which all others depend) and then taking the modal series, i.e., the series ranked 50. Ideally, the PDF should be more intensively sampled, but this would generate extremely large volumes of daily time series output. This output could however be used to construct PDFs for the station of interest

Such an approach could be applied in ENSEMBLES to the PDFs constructed from RCM output from WP2B.1 using the methodologies (such as Reliability Ensemble Averaging and kernel algorithms) being developed in RT3. These PDFs are likely to be for the Rockel sub-European regions and for seasonal temperature and precipitation. This is rather coarse in terms of both spatial and temporal scale (for the UKCIP work, UEA is working with monthly PDFs for 25 km grid boxes) for generating station series. It would be more appropriate to apply the approach to the PDFs for European cities which will be produced by Michel Dequé, CNRS. However, this would be dependent on the availability of appropriate daily station data for calibrating the weather generator and joint probabilities (temperature and precipitation) for perturbing the weather generator. Without the joint probabilities, only precipitation could be generated, since in this particular weather generator, temperature is dependent on precipitation.

5. CLOSING REMARKS AND RECOMMENDATIONS

This deliverable is intended as a discussion document to set the context for and prepare the way for detailed work that will be undertaken in the last two years of ENSEMBLES (see Table 1 which describes partners' plans for SDS work in WP2B.2 as of late summer 2007). It does not address in detail all the issues that arise in applying statistical downscaling within a probabilistic framework, but focuses on issues concerning the production of probabilistic output from a single input (using the authors' experience with particular methods as illustration). Given the complexity of the issues that arise in an end-to-end approach, breaking down the problem into its separate components seems a sensible strategy – i.e., 'walking before attempting to run'.

Sections 3.2 and 3.4 detail a number of specific recommendations with respect to obtaining probabilistic output from a single input. This sensitivity analysis type approach can also be considered as analogous to the ‘perturbed physics/parameter’ approach taken elsewhere in ENSEMBLES. FIC plan to implement many of the recommendations from Section 3.2 and 3.4 over the coming months (Table 1), and other partners will also explore some of these issues (e.g., NMA and ARPA-SMR, Table 1).

Other partners will explore stochastic approaches, which also produce probabilistic output from a single input – e.g., UEA, GKSS, NMA, NIHWM (Table 1). Here, one of the outstanding issues is to explore how the stochastic variability of weather generators, for example, compares with observed natural variability.

The above analyses focus on the application of single SDSMs - an important extension of this work will be that by IAP exploring the uncertainties arising from the use of multiple SDSMs (Table 1). This will be one of the contributions to ongoing Task 2B.2.13 work on assessment of uncertainty in regional projections.

Further consideration needs to be given to how to combine information from multiple inputs and models to produce PDFs, and other probabilistic outputs (see deliverable D2B.18), i.e., work on ensemble averaging techniques and Bayesian approaches including Monte Carlo sampling. This work will benefit from discussions with ICTP who are working on modification of the REA method (see deliverable D2B.6) and Hayley Fowler (who has recently joined RT2B as an affiliate partner – Fowler et al., 2007). Other specific issues that need to be addressed in more detail are how to calculate and apply weighting schemes for SDS and how to incorporate natural variability. These issues will be the main focus of Task 2B.2.12 work in the coming months which will produce recommendations and guidance on methods for constructing probabilistic regional projections.

The original ENSEMBLES description of work identified three further aspects regarding the modification of SDSMs in order:

- To generate scenarios based on the ‘grand probability’ distributions which will be constructed in RT1 and RT2A
- To generate scenarios for GCM/emissions forcing scenarios for which RCM output is not available, i.e., to extend the RCM ensembles developed in WP2B.1 by the synergistic use of RCMS and SDS
- To generate long stable time series that have the required characteristics of a common parent population for extreme value and other statistical analyses

The first issue has been discussed in Section 4.3. The more promising approach would be to use weather generators to sample from the regional or local PDFs generated from the WP2B.1 transient runs using the methodologies developed in RT3. However, constrained resources and the delivery of such PDFs rather late in the project, mean that it will likely not be possible to explore this approach further during ENSEMBLES.

The second issue will be addressed as part of Task 2B.2.14 work on assessment of the robustness of statistical downscaling and synergistic use of statistical and dynamical downscaling, principally by UC using the ENSEMBLES web-based downscaling service (see deliverables D2B.4 and D2B.19). The third issue is being addressed by KNMI (see Table 1).

It is evident that applying SDSMs within a probabilistic framework raises many challenges. While not all of these are dealt with in detail in this discussion document, it nevertheless provides a sound starting point for detailed analytical and numerical work in the last two years of the ENSEMBLES project.

6. REFERENCES

- Allen, M.R., Stott, P.A., Mitchell, J.F.B., Schnur, R. and Delworth, T.L., 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change, *Nature*, **407**, 617-620.
- Benestad, R.E., 2004: Tentative probabilistic temperature scenarios for northern Europe, *Tellus*, **56A**, 89-101.
- Benestad, R.E., 2005: Climate change scenarios for northern Europe from multi-model IPCC AR4 climate simulations. *Geophysical Research Letters*, **32**, L17704, doi:10.1029/2005GL023401.
- Bürger, G. and Cubasch, U., 2005: Are multiproxy climate reconstructions robust? *Geophysical Research Letters*, **32**, L23711, doi:10.1029/2005GL024155.
- Dubrovsky M., Nemesova I., and Kalvova J., 2005: Uncertainties in climate change scenarios for the Czech Republic. *Climate Research*, **29**, 139–156.
- Ekström, M., Hingray, B., Mezghani, A. and Jones, P.D., 2007. Regional climate model data used with the SWURVE project 2: addressing uncertainty in regional climate model data for five European case study areas. *Hydrological and Earth Systems Science*, in press.
- Fowler, H.J., Blenkinsop S. and Tebaldi, C., 2007: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27**, 1547-1578.
- Frias, M.D., Zorita, E., Fernández, J. and Rodriguez-Puebla, C., 2006: Testing statistical downscaling methods in simulated climates, *Geophysical Research Letters*, **33**, L19807, doi:10.1029/2006GL027453.
- González-Rouco, J.F., Heyen, H., Zorita, E. and Valero, F., 2000: Agreement between observed rainfall trends and climate change simulations in the Southwest of Europe. *Journal of Climate*, **13**, 976-985.
- Goodess, C.M., Anagnostopoulou, C., Bárdossy, A., Frei, C., Harpham, C., Haylock, M.R., Hurrell, J., Maheras, P., Ribalaya, J., Schmidli, J., Schmith, T., Tolika, K., Tomozeiu, R. and Wilby, R.L., 2007a: An intercomparison of statistical downscaling methods for Europe and European regions – assessing their performance with respect to extreme temperature and precipitation events. *Climatic Change*, submitted.

- Goodess, C.M., Hall, J., Best, M., Betts, R., Cabantous, L., Jones, P.D., Kilsby, C.G., Pearman, A. and Wallace, C., 2007: Climate scenarios and decision making under uncertainty. *Built Environment*, **33**, 10-30.
- Greene, A.M., Goddard, L. and Lall, U., 2005: Probabilistic multimodel regional temperature change projections. *Journal of Climate*, **19**, 4326-4343.
- Katz, R.W., 2002: Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, **20**, 167-185.
- Katz, R.W., 2005: Bayesian approach to decision making using ensemble weather forecasts. *Weather and Forecasting*, **21**, 220-231.
- Kennedy, M.C. and O'Hagan, A., 2001: Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63**, 425-464.
- Kilsby, C.G., Jones, P.D., Harpham, C., Burton, A., Ford, A.C., Fowler, H.J., Smith A. and Wilby, R.L., 2007: A daily weather generator for use in climate change studies. *Environmental Modelling and Software*, **22**, 1705-1719.
- Luo, Q., Jones, R.N., Williams, M., Bryan, B. and Bellotti, W., 2005: Probabilistic distributions of regional climate change and their application in risk analysis of wheat production. *Climate Research*, **29**, 41-52.
- Murphy, J. M., Sexton, D.M., Barnett, D.N., Jones, G.S., Webb, M.I., Collins, M., Allen, M.R. and Stainforth, D., 2004: Quantifying uncertainties in climate change using a large ensemble of general circulation model predictions. *Nature*, **430**, 768 – 772.
- Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J., 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A*. **365** 1993-2028.
- Palutikof, J.P., Goodess, C.M., Watkins, S.J. and Holt, T., 2002: Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean, *Journal of Climate*, **15**, 3529-3548.
- Piani, C., Frame, D.J., Stainforth, D.A. and Allen, M.R., 2005: Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophysics Research Letters*, **32**, L23825, doi:10.1029/2005GL024452.
- Schmidli J., Goodess, C.M., Frei, C., Haylock, M.R., Hundecha, Y., Ribalaygua, J. and Schmith, T., 2007: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research*, **112**, D04105, doi:10.1029/2005JD007026.
- Tebaldi, C., Mearns, L.O., Nychka, D. and Smith, R.L., 2004: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophysical Research Letters*, **31**, art. No. L24213.

- Tebaldi, C., Smith, R., Nychka, D. and Mearns, L.O., 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach for the analysis of multi-model ensembles, *Journal of Climate*, **18**, 1524-1540.
- Tebaldi, C., Hayhoe, K., Arblaster, J.M. and Meehl, G.A., 2006: Going to the extremes. An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change*, DOI: 10.1007/s10584-006-9051-4.
- Wilby, R.L. and Harris, I., 2006: A framework for assessing uncertainties in climate change impacts: low flow scenarios for the River Thames, UK. *Water Resources Research*, **42**, W02419, doi:10.1029/2005WR004065.
- Winkler, J.A., Andresen, J.A., Guentchev, G., Waller, E.A. and Brown, J.T., 2003; Using ANOVA to estimate the relative magnitude of uncertainty in a suite of climate change scenarios. 14th Symposium on Global Change and Climate Variations, Feb. 2003, Long Beach, California.

Table 1: Summary of statistical downscaling methods to be used in WP2B.2. NB This does not list methods implemented in the ENSEMBLES web-based downscaling service (www.meteo.unican.es/ensembles).

Group/Method	Proposed predictands	Proposed predictors	Brief description of method and references	Source of predictors – Reanalysis, multi-model GCM/RCM ensembles	Region(s)/predictand datasets which it is proposed to downscale	Brief outline of how uncertainties will be addressed and/or probabilistic projections derived
ARPA-SIM: Regression, conditioned by circulation	Prec, Tmin, Tmax (mean values and extreme event frequency)	Z500, T850, MSLP, RH850 (monthly means)	CCA for scenarios: Barnett and Preisendorfer, 1987; von Storch <i>et al.</i> , 1993 The CCA technique finds pairs of patterns e.g., correlation between two corresponding pattern coefficient is maximized. In order to reduce noise, before the CCA, the data sets are projected on EOFs (empirical orthogonal functions) and only those explaining the most of the total observed variance are retained. The most important CCA pairs are then used in a multivariate linear model to estimate the predictand anomalies from the predictor anomaly field.	ERA40 Multi-model ensembles of CTL/Scenario CGCM experiments	Region: N-Italy Data-set: Aeronatica Militare, daily data	Production of ensembles of downscaled predictions.
Regression,	Prec, Tmin,		BLUE+MLR/MOS for	ERA40	Region: Italy	Production of

conditioned by circulation	Tmax (mean values and extreme event frequency)	Z500, T850 (monthly means)	seasonal: Thompson, 1977; Pavan <i>et al.</i> , 2005	Multi-model seasonal ensemble CGCM hindcasts	Data set: UCEA daily analysis	calibrated ensemble of downscaled predictions.
FIC: two-step analogue method	Daily precipitation and temperatures. Wind and humidity are planned to be tested.	Z1000, Z850, Z500; Low tropospheric humidity and thickness (1000 to 500 hPa); Temperature of the previous days (the predictand is used latter as predictor). Instability indexes and snow cover related predictors are planned to be tested, and some others (real wind instead of geostrophic...)	Two-step analogue method, in which (1) the 'n' most similar days to the day being simulated are selected from a reference data set and (2) predictands / predictors relationships are obtained from the 'n' days data set (performing different analyses, including multiple regressions), and applied to the problem day	Reanalysis (ERA40) and multi-model GCM ensembles. If we have time, we would like to apply the method to RCM output	Predictands: T max, Tmin and daily precipitation, both for gridded datasets and site observations (very important for extremes). If we have time, we want to try other predictands, at least wind (very important for some end-users, like wind-power companies). Regions: Europe. We would also like to work in other ENSEMBLES regions, at least in Africa. We are also interested in South America. In this case, we would need ERA40, GCM output and observations for these areas	The method already addresses some uncertainties (developed within Stardex). We plan to work on uncertainties consideration and quantification, as described in D2B14 (relaxing resolutions, analysing the range of applicability of the method...). The method can produce daily probabilistic output (from a single input), and from that daily probabilistic output for multi-model GCM ensembles, we plan to obtain final PDFs
GKSS: conditional stochastic weather generator	Marine surface wind		Monte Carlo simulations and extreme values analysis. Busuioc and von Storch, 2003.	ERA-40 as predictor and RCM winds as predictands.		Will use 1000 year simulations (e.g., ECHO-G simulations) to derive natural variability of wind.
IAP: regression,	Daily	500, 1000 hPa	Days are stratified by	Reanalysis for	Europe, data probably	Different methods

conditioned by circulation	temperature (possibly also daily precipitation)	heights (or SLP), 850 hPa temperature, 1000/500 hPa thickness, for precipitation, also some humidity-related variable	classification based on circulation patterns, within each class multiple linear regression is performed; Huth <i>et al.</i> , 2007 (Huth, R., Kliegrová, S., Metelka, L. (2007): Nonlinearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? <i>Int. J. Climatol.</i> doi: 10.1002/joc.1545	training, GCM control + perturbed ensemble for producing scenarios	from ECA&D project (unless something better has been produced in the meantime) – for all IAP methods	produced by IAP with different predictor sets and different parameters (e.g., no. of PCs, CCA pairs) are taken for a single GCM output, weighted by several characteristics of their performance; uncertainty due to SDS model selection and parameters is compared with other sources of uncertainty
IAP: neural network	Daily temperature	500, 1000 hPa heights (or SLP), 850 hPa temperature, 1000/850 hPa thickness, for precipitation, also some humidity-related variable	Multilayer perceptron with one hidden layer, inputs are either PCs of predictor(s) or their gridpoint values; Huth <i>et al.</i> , 2007	As above	As above	As above
IAP: conditional stochastic weather generator	Precipitation, min and max temperature, solar radiation	N/A	Precipitation occurrence simulated by two-state Markov chain, precip. amount by gamma distribution, other variables by normal distribution; all is conditioned on variability on a monthly scale;	As above	As above	As above

			Dubrovsky <i>et al.</i> , 2004			
IAP: multiple linear regression	Daily temperature (possibly also daily precipitation)	500, 1000 hPa heights (or SLP), 850 hPa temperature, 1000/500 hPa thickness, for precipitation, also some humidity-related variable	Multiple linear regression with stepwise screening of gridpoint values; Huth, 2002	As above	As above	As above
KNMI: nearest-neighbour resampling	Multi-site (sub)daily RCM precipitation (and temperature)	Same as predictands	KNMI will concentrate on the use of nearest-neighbour resampling to generate long stable time series which can be used to determine the exceedance probabilities of very rare multi-day extreme events (1 in 1000 year extremes). Depending on its availability the use of sub-daily data may be considered. (see Leander and Buishand, 2007, J. Hydrol., 332, 487-496)	GCM/RCM ensembles	River Rhine catchment (on RCM grids or possibly transformation to a common regular grid or hydrological sub-catchments)	Previous studies have shown that the uncertainty related to the driving GCMs is generally larger than that of the RCMs (various GCMs in the ensemble is more important than various RCMs). Model uncertainties should be distinguished from e.g. uncertainties in greenhouse gas emissions (are we able/willing to say something about the probability of future emissions?). Probabilistic projections in terms of return periods or extreme quantiles are obtained from the

						GCM/RCM ensemble (i.e. nearest-neighbour resampling applied to each ensemble member).
NIHWM: conditional stochastic weather generator	Temperature, precipitation, drought indices, discharge level of the Danube basin	Low frequency PCs of MEOF of the geopotential at 500 hPa, 500-1000 hPa and SLP	<p>Step 1: Filtering by MEOF (Multivariate Empirical Orthogonal Function) of the predictors, for Atlantic-European region. Markov Models applied to MEOF, see Xue <i>et al.</i>, 2000 and Chen and Yuan, 2004.</p> <p>Step 2: Classification of the atmospheric circulation patterns (CPs) by means of the first PC of MEOF decomposition.</p> <p>Step 3: Construction of Markov chain for CP transformation; estimation of the transition probability matrix, limiting matrix, ergodicity coefficients and other characteristics of Markov modelling.</p> <p>Step 4: Results obtained for large scale circulation are associated with occurrence of extremes for Balkans and Danube basin.</p>			

<p>NMA: conditional stochastic weather generator</p>	<p>Daily precipitation</p>	<p>Monthly means: -SLP (sea level pressure); -specific humidity at 1000, 950, 850, 700 hPa. I am not sure if all these levels are available from the GCM outputs. As I remember well, only 1000, 850 and 500 levels will be available. -instability index (using specific humidity, temperature and potential temperature at 850 and 500hPa)</p>	<p>Mixture between a two-state first order Markov chain and a SDSM based on CCA (Busuioc and von Storch, 2003). Precipitation occurrence is described by a two-state, first order Markov chain and the variation of precipitation amount on wet days is described by two gamma distribution parameters. The four parameters (two transition probabilities and two gamma distribution parameters) are linked by the large scale predictors through the CCA model. Other linear models will be also tested (e.g., CCA for seasonal precipitation).</p>	<p>-Reanalysis for calibration; multi-model GCM ensembles to produce local probabilistic climate change scenarios; the possibility to use multi-model RCM ensembles/only some RCMs is also considered if the model parameters could be calibrated for the current climate RCM outputs</p>	<p>Daily precipitation (including computation of some extreme events) at stations for southern Romania, which are then cumulated on monthly/seasonal scale</p>	<p>-considering ensembles of multi-SDMs obtained by various combinations of predictors giving similar skill over various independent observed data set (validation intervals); -calculating 90% confidence intervals for downscaled values from multi-runs (e.g.1000 runs); -comparison with RCM output climate change scenarios for some appropriate downscaled parameters</p>
<p>UEA: stochastic weather generator</p>	<p>Daily precipitation, Tmax, Tmin, vapour pressure, wind speed, sunshine duration, relative humidity, reference PET</p>	<p>Grid-point change fields (mean and std. dev.) for daily precipitation, Tmax, Tmin (and possibly other variables).</p>	<p>First-order, infinite-state Markov chain model. Secondary variables are all dependent on precipitation. Model parameters (e.g., precipitation gamma distribution) are perturbed using 'predictors'. Kilsby et al., 2007.</p>	<p>Change factors will be taken from the WP2B.1 RCM runs</p>	<p>7 mainland European stations, plus 3-4 UK stations. Daily timescale – but presentation of results will focus on seasonal indices of extremes.</p>	<p>PDFs, CDFs etc., will be constructed – following the CRANIUM approach (Goodess et al., 2007b). Weighting schemes will be tested – including weights from RT3.</p>