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

D2B.32 Journal papers on uncertainty in regional climate projections

Due date of deliverable: April 2009
Actual submission date: February 2010

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

Organisation name of lead contractor for this deliverable: IAP/CNRM

Revision [1]

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)	

1. Uncertainty in regional climate projections

The work reported in this deliverable addresses two of the scientific/technical questions posed in the RT2B description of work: (1) which are the most important sources of uncertainty for high-resolution regional climate projections, particularly with respect to extremes? and, (2) which, if any of these sources of uncertainty have been reduced as a result of the ENSEMBLES work and which could be reduced with further work?

Two major analyses have been undertaken (each of which will eventually be reported in detail in a journal paper), one focusing on statistical downscaling, the other on dynamical downscaling.

The team working on statistical downscaling, was led by IAP, and included ARPA-SIM, FIC and UC. One of the novel aspects of this work is the development of a weighting methodology and criteria for statistical downscaling. This complements the weighting scheme for dynamical downscaling developed by ENSEMBLES RT3. The two schemes are compared and the wider progress made on weighting-related issues discussed in Section 6.2.2 of the ENSEMBLES science summary report (Goodess et al., 2009). Here, in Section 2, the rationale for the statistical downscaling work is presented, together with the weighting scheme and the technical design of the study. Plans for completion of the work are also reported.

For the work on dynamical downscaling, CNRM has carried out an assessment of uncertainties in the ENSEMBLES transient RCM runs (see deliverable D2B.22) using the approach applied to the PRUDENCE runs (Déqué et al., 2007) together with a new approach based on weather regimes to estimate changes for missing cells in the GCM-RCM matrix. The implications of these results with respect to the robustness of RCM results and the partitioning of uncertainty in the RCM simulations are discussed in Section 6.2.1 of the ENSEMBLES science summary report (Goodess et al., 2009). A journal paper was submitted to *Climate Dynamics* in summer 2009 (Déqué et al., 2009 – available as an appendix to this deliverable, but with restricted dissemination status) and the referees comments received in September 2009. In order to address the latter comments, it is necessary to maximise the filling of the GCM-RCM matrix with actual RCM runs. Therefore, it was agreed to wait until the largest possible RCM ensemble was available, i.e., until the end of December 2009 (see deliverable D2B.22) and to submit a revised manuscript early in 2010 once the additional model data have been processed.

2. Uncertainty in regional climate projections constructed using statistical downscaling

2.1 Introduction

Statistical downscaling (SDS) is one of the two widely used ways to bridge the spatial scale mismatch between what global climate models (GCMs) are able to provide and what is needed in impact studies. Analogously to other modelling techniques, SDS models are also accompanied by a suite of uncertainties (as also discussed in

ENSEMBLES deliverable D2B.14). These uncertainties are additional to uncertainties related to other parts of the climate modelling chain, including those related to emission scenarios, between-GCM, within-GCM, and impact models.

The uncertainties in SDS models are connected with the selection of: (i) the statistical model (transfer function); (ii) variable(s) used as predictor(s); (iii) geographical area on which the predictor is defined; (iv) geographical area on which the predictand is defined (applicable to some methods only); and, (v) parameters of the models (e.g., number of principal components, canonical correlation pairs, criteria in stepwise regression, etc.). These uncertainties have been neglected in many past studies. For example, results of a single model are sometimes presented as a 'truth'. Also, the 'best' SDS model is often selected according to its performance compared with present climate, and only this model is used to assess future climate change, although for this purpose, the selected model may be suboptimal.

The uncertainties embedded within SDS models can be taken into account by constructing an ensemble from a set of variants of SDS models, differing in the transfer function, choice of predictors, their parameters, etc. The effect of the uncertainties and their quantification can be evaluated if all other options are kept constant: the emission scenario, driving GCM, time horizon, etc.

In making such an ensemble, one should decide whether all ensemble members are given the same importance and the same chance to affect the resultant PDF of change or whether the ensemble members are weighted. Here, it is proposed that the ensemble members should be weighted, at least because weighting allows 'bad' models (whatever 'bad' may mean) to be eliminated from affecting the PDF of climate change. A question one more step ahead is whether the individual weights (weighting criteria) should be equally important or should be given different 'weights'. This goes beyond our task; here we assume that all the weights are equally relevant. Therefore, the weights are quantified so that their maximum value is +1.

2.2 Weighting

Most of the weighting criteria proposed stem from the assumptions on which the SDS technique is based. Five criteria are formulated below. In order that they can be implemented, each criterion must be expressed in quantifiable terms. There is no unique way of making such a quantification; here we either present several possibilities or provide an arbitrary choice of one criterion. Some weighting criteria are difficult to quantify; in such cases we discuss ways that could potentially lead to formulation of a quantifiable criterion.

Weighting criterion 1. Reproduction of the predictand by the SDS model. This is one of the most important properties of SDS models and the property according to which SDS models have most frequently been evaluated. The better the reproduction, the higher the weight which should be attached to the corresponding realization of the climate change estimate. Clearly, a model unable to describe the observed relationship between predictor and predictand, cannot be used with confidence in estimates of future climate change. Any reasonable measure of correspondence may play the role of this weight. For practical purposes, correlation coefficient is sufficient, having the important properties of being equal to one for a perfect correspondence, to zero for no correspondence, and being negative if the predictor-predictand relationship is reproduced with the wrong sign.

Weighting criterion 2. Reproduction of trends. SDS models are trained so that they optimize the predictor-predictand relationship on short time scales. However, climate change proceeds on time scales much longer than that to which SDS models are fitted. Therefore, it is important in climate change studies to verify that the SDS models correctly reproduce long-term behaviour. Therefore, it is of great importance to check if SDS models are capable of simulating long-term climate variations, whether trends or contrasts between different climate states (cold/warm, dry/wet, etc.). This property of SDS models has frequently been overlooked in SDS studies. The rationale behind this weighting criterion is that the ability of a SDS model to simulate recently observed long-term changes gives credence to its ability to correctly describe a future development of climate (although the latter cannot be verified, and the validity of the former does not imply the validity of the latter, for example in situations when recent and future climate changes are governed by different mechanisms).

This criterion can be quantified in several different ways, two of which are mentioned here:

1. The weight is linear between the observed value of the trend (weight equal to one) and zero trend on one side and double of the observed trend on the other side (weight equal to zero in both cases). This is applicable to variables for which the trends are distant from zero (e.g., temperature). In the case of trends close to zero (e.g., precipitation in some regions), the weighting would give spurious outputs.

2. The weight is the test statistic for the equality of regression lines or correlation coefficients, normalized to one for perfect reproduction of the observed trend.

Weighting criterion 3. Reproduction of predictors in the driving GCM. Even a very good SDS model (e.g., in terms of its performance on observed data) would fail when used for estimates of future climate change if its predictors are simulated incorrectly by the GCM used as input for downscaling. Therefore, the ability of the driving GCM to accurately simulate the predictor variable(s) is important. In contrast to the previous two criteria, it is not *a priori* clear here what property(ies) of the predictor to quantify. The bias, which is used as a criterion in the original REA method applied to GCMs (Giorgi and Mearns 2002), is not applicable here because it is easy to correct in SDS outputs (in fact, most SDS methods are trained so that they reproduce the mean value, and also variance). Since SDS models frequently treat the predictor variables as fields, and many of them use principal component analysis (PCA) or some similar technique to reduce the dimensionality of predictor(s), principal components (PCs) of predictor(s) may serve as a basis for the quantification of this weight. One reasonable option is to calculate pattern similarity between several leading PC loadings of real and GCM-simulated predictor data. The similarity measure should be the congruence coefficient (uncentred correlation) rather than the correlation coefficient since only the former is known to be correctly applicable to PC loading patterns (Richman 1986).

Weighting criterion 4. Stationarity of the predictor-predictand relationship. The stationarity assumption is one of the most important assumptions of SDS. If the relationship between the predictor and predictand changes in time under current climate, it is unlikely that the relationships will hold unchanged under future climate. On the other hand, perfect stationarity under current climate is not proof of future stationarity, it is only an indication. Here, the weight may be based on running correlations between the predictor and predictand: if they do not change in time, the weight is equal to one; if they change significantly in time, a penalty based on a test statistic for the equality of correlation coefficients is introduced.

Weighting criterion 5. Correction for multiplicity. Some SDS models are very similar, e.g. if 500 hPa heights plus 850 hPa temperature, or 500 hPa heights plus 1000/500 hPa thickness are used as predictors in the same SDS model. One model thus provides very little additional information over the other. The inclusion of similar models in the ensemble may bias the final PDF of climate change response towards them. However, how to evaluate the similarity of models? Clearly the similarity of the climate change response itself cannot be used as a similarity measure since even entirely dissimilar models can result in identical climate change response. A more suitable candidate is perhaps the similarity of time series because day-to-day variability is likely to differ between different models.

This is illustrated in Table 1, which is based on a selection of SDS models described in Huth (2004) and shows correlation coefficients between statistically downscaled time series. Clearly, time series simulated by similar methods are highly correlated whereas dissimilar models result in correlations not so high (but not low either). This is evidence in favour of using the similarity between time series as a proxy for multiplicity of SDS models. But, it is necessary to bear in mind that even different SDS models may produce very similar time series simply because they are ‘correct’. Therefore, the quantification of this weight is open for further discussion. The only feasible way of correction for multiplicity, available at present, is thus an *a priori* removal of one or more highly similar methods. This, however, allows quite a subjective decision to enter the whole procedure and reduces the number of members in the ensemble.

The final weight with which each SDS model enters the production of the final PDF of climate change response, is the product of all individual weights.

2.3 Implementation

The work flow in SDS-ensemble making is outlined in Figure 1. The following decisions relating to the working procedure were made in order to ensure a consistent approach by all ENSEMBLES groups contributing to the analysis:

- The analyzed period covers the whole period for which ERA-40 reanalysis is available, that is, September 1957 to August 2002.

- The SDS methods included in the study are:
    - multiple linear regression (MLR) of gridded data,
    - MLR of PCs,
    - MLR on data stratified by classification of circulation patterns,
    - neural networks (multilayer perceptron, local neural networks),
    - analogues coupled with regression,

- weather typing by self organizing maps (SOMs)
- two-step analogue method (FIC)
- canonical correlation analysis (ARPA-SIM)
- methods implemented in the ENSEMBLES downscaling portal (www.meteo.un-ican-es/ensembles).

- Wherever possible, several predictors (or sets of predictors) and meaningful options (number of PCs, classification methods, numbers of classes, etc.) are used.

- Predictors are taken from ERA-40 reanalysis. 500 and 1000 hPa heights and 850 hPa temperature are used preferentially. The window on which the predictors are defined is 25º to 80ºN, 50ºW to 55ºE. The grid resolution is 2.5º x 2.5º.

- The predictands (that is, the downscaled variables) are the daily maximum and daily minimum temperature. 10 stations across Europe were selected, representing as much varied climatic and geographical settings as possible. The station temperature data are taken from the ECA&D database; the criterion for the selection of stations was their completeness in the analyzed period. The selected stations are:
  - Sodankylä (FI),
  - Valentia (EI),
  - Salamanca (ES),
  - Bologna (IT),
  - Tirgu Jiu (RO),
  - Poltava (UA),
  - Praha (CZ),
  - Bamberg (DE),
  - Hohenpeissenberg (DE),
  - Zugspitze (DE).

- For training and validation, four divisions of the dataset into training and validation periods were produced in the following manner: Every fourth year was withdrawn for validation, starting in turn with 1957/8, 1958/9, 1959/60, and 1960/1. Thus 34 (11) or 33 (12) years form the training (validation) subsample. The reason why cross-validation is not conducted is that some methods cannot be run in the cross-validation framework at all (analogues) or their calculations are time consuming so that cross-validation can not be realized within a reasonable time frame (e.g., for neural networks).

- As a common GCM run, the third ensemble member of the ECHAM5 transient simulation was chosen. This choice was made to facilitate comparisons with ENSEMBLES and other RCM runs - many of which also use this simulation. For future climate, the time slice 2070-2099 was chosen.

- The PDF of future temperature change is constructed using the Gaussian kernel algorithm, which allows weights to be introduced. The kernel width is optimized so that a compromise is achieved between the PDF being too smooth (with too wide a window) and too bumpy (with too narrow a window).

However, the choice of width is only a technical matter, which is not crucial for the final output.

2.4 Example using older data

The example presented here is based on SDS models used in Huth (2004). 29 SDS models were applied at 39 central and western European stations to estimate daily mean temperature in winter. The database contained eight winter seasons. The SDS models were applied to the doubled CO_2 run of the CCCM2 GCM. The Gaussian kernel algorithm was used with a kernel width of 0.25ºC. Three stations are chosen for display: Štrbské Pleso (SK) where the SDS models achieved the best performance (in terms of lowest root mean square error), Klagenfurt (AT) where the SDS models performed worst, and Hamburg (DE).

Figure 2 displays results for unweighted SDS models and SDS models weighted by the variance explained (weighting criterion number 1 was applied – reproduction of the predictand by the SDS model). The weighting reduces the effect of SDS models producing low temperature change, especially at Štrbské Pleso and Klagenfurt. At Klagenfurt, the weighting also increases the probability of temperature change slightly over 5ºC.

2.5 Plans for completion of the analysis and publication

As described above, a considerable amount of theoretical and exploratory work on weighting criteria for statistical downscaling has been completed, together with detailed technical work on setting up a consistent analytical framework and preparing predictor/predictand datasets and several SDS models. Some of the groups have already constructed PDFs based on ECHAM5 output (see, for example, PDFs 4 for Prague in Section 6.6.2 of the ENSEMBLES summary report (Goodess et al., 2009)). Unfortunately, due to a variety of institutional and personal reasons, IAP was not able to complete the co-ordination of this work. All groups involved are, however, still committed to completing this work after the end of ENSEMBLES and submitting a journal paper. UC has agreed to take over the co-ordination role.

References

Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., Castro, M. and van den Hurk, B., 2007: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change* **81,** 53-70

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C.M., Jacob, D., Lenderink, G. and Christensen, O.B., 2009: The spread amongst ENSEMBLES regional scenarios: Regional Climate Models, driving General Circulation Models and interannual variability, *Climate Dynamics*, submitted.

Giorgi, F., Mearns, L.O., 2002: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. *Journal of Climate*, **15**, 1141-1158.

Goodess, C.M., Jacob, D., Déqué, M., Guttiérrez, J.M., Huth, R., Kendon, E., Leckebusch, G.C., Lorenz, P. and Pavan, V., 2009: Downscaling methods, data and tools for input to impacts assessments, in van der Linden, P. and

J.F.B. Mitchell (eds): ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK. 160pp.

Huth, R., 2004: Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate,* **17**, 640-652.

Richman, M.B., 1986: Rotation of principal components. *Journal of Climatology,* **6**, 293-335.

Table 1. Correlation coefficients (x100) between time series of daily mean temperature produced by 13 SDS models from Huth (2004). Predictors: Z5 – 500 hPa heights, Z0 – 1000 hPa heights, T8 – 850 hPa temperature, TH – 1000/500 hPa thickness. Methods: PtStepR – pointwise stepwise regression, PCFullR – regression of PCs without screening, PCStepR – stepwise regression of PCs, CCA – canonical correlation analysis. Setting: pts – number of points entering regression models, PCs – number of principal components entering regression models, CCs – number of canonical pairs.

predictors	method	setting		correlation coefficient (x100)											
Z5T8	PtStepR	~25pts		99	98	80	81	80	81	78	78	73	67	14	11
Z5T8	PtStepR	12 pts	99		97	81	82	81	82	79	78	74	67	14	11
Z5T8	PtStepR	40 pts	98	97		79	81	79	81	77	76	71	65	12	9
Z5T8	PCFullR	7 PCs	80	81	79		95	100	94	99	98	94	93	30	27
Z5T8	PCFullR	11PCs	81	82	81	95		95	100	93	92	89	85	37	33
Z5T8	PCStepR	7 PCs	80	81	79	100	95		95	99	98	94	92	31	27
Z5T8	PCStepR	11PCs	81	82	81	97	100	95		93	91	89	84	38	33
Z5T8	CCA	5 CCs	78	79	77	99	93	99	93		99	95	94	26	23
Z5T8	CCA	6 CCs	78	78	76	98	92	98	91	99		93	92	22	18
Z5TH	PCFullR	7 PCs	73	74	71	94	89	94	89	95	93		95	29	26
Z5	PCFullR	7 PCs	67	67	65	93	85	92	84	94	92	95		36	32
Z0	PCFullR	7 PCs	14	14	12	30	37	31	38	26	22	29	36		97
Z0	PCFullR	10PCs	11	11	9	27	33	27	33	23	18	26	32	97	

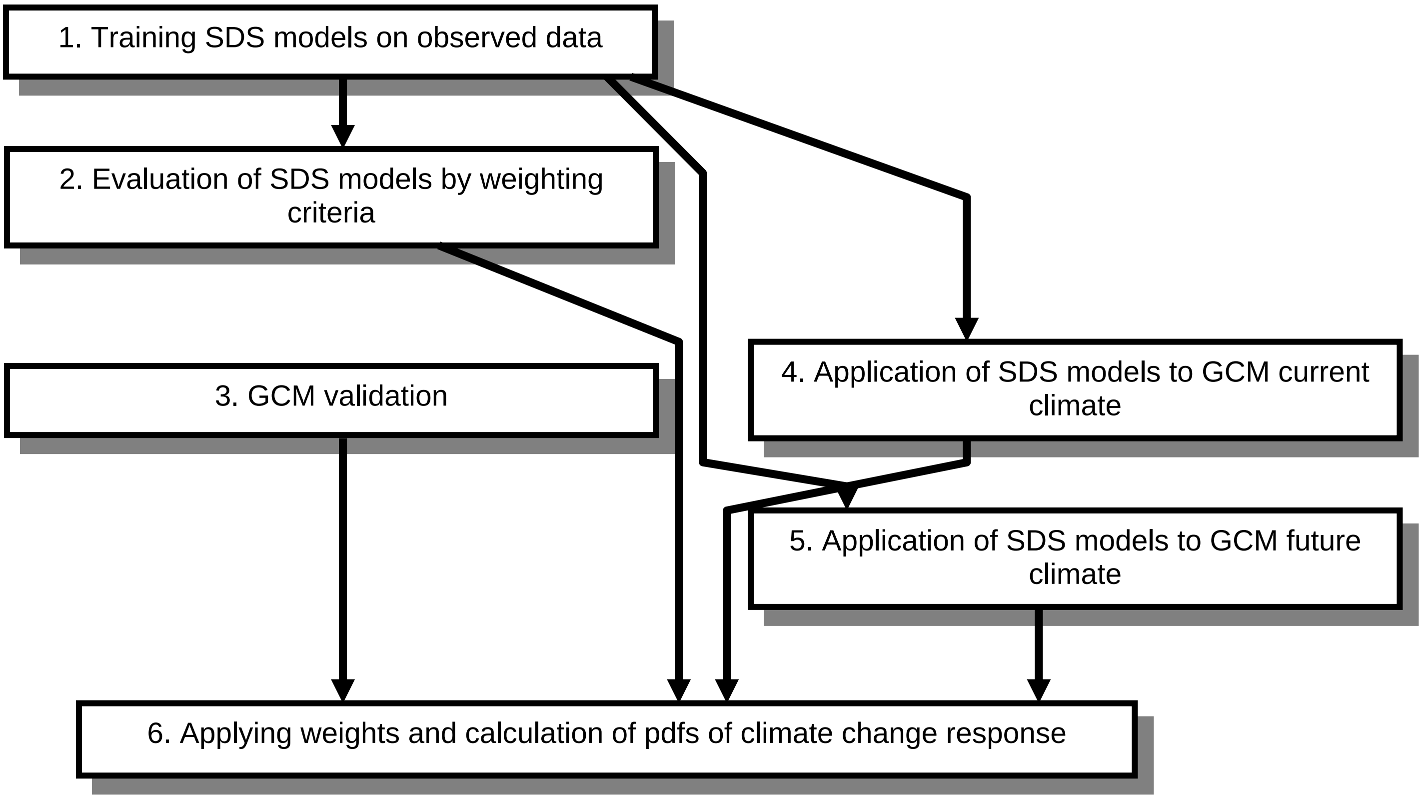

Figure 1. Flow diagram of work in SDS-ensemble making.

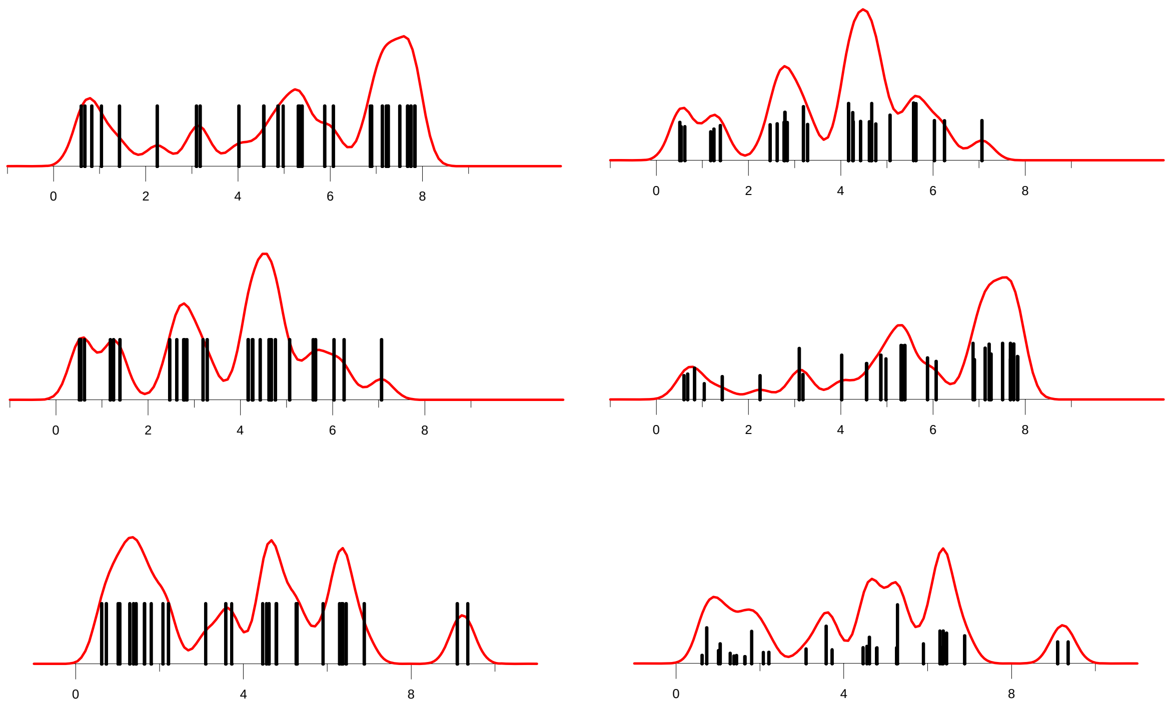

Figure 2. PDFs of temperature change (red) and temperature changes corresponding to individual SDS models (black bars): the position of a bar determines the temperature change (in °C) while its length displays the weight with which the model enters the PDF calculation. No weighting (left) and weighting by the variance explained (right); Štrbské Pleso (top), Hamburg (center), and Klagenfurt (bottom).