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

D5.9 Report on the analysis of possible gridding methods

Due date of deliverable: 01-03-2006
Actual submission date: 03-03-2006

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

University of East Anglia

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)	

Methods for gridding of long-term (1960-present) high-resolution daily European climate data

Nynke Hofstra[1], Malcolm Haylock[2,*], Phil Jones[2], Mark New[1]

[1] Oxford University, Centre for the Environment, South Parks Road, Oxford, OX1 3QY, England
[2] University of East Anglia, School of Environmental Sciences, Climate Research Unit, Norwich, NR4 7TJ, England

Abstract

Daily gridded data are needed for the validation of e.g. climate models. In this study a comparison of available interpolation methods will be undertaken and the best performing one selected for the development of a daily, long-term (1960-present), high-resolution (up to 20 km grid) dataset for Europe of mean, minimum and maximum temperature, precipitation, mean sea level pressure and snow depth. Six methods are to be compared: natural neighbour interpolation (NNI); angular distance weighting (ADW); kriging; thin plate splines (TPS); reduced space optimal interpolation (RSOI); and conditional interpolation (CI). So far, we compared four of the six methods, calculating the root mean squared error (RMSE), linear error in probability space score (LEPS) and the Pearson correlation coefficient (R) in a cross-validation exercise.

Taking into account that two out of the six methods (RSOI and CI) have not yet been automated and that none of the methods have yet been fine tuned, all four working methods perform well. The simple interpolation method NNI gives, generally, the best results. However, the comparison of methods will continue. Both RSOI and CI will be added to the comparison and a validation of area average values for the UK and Switzerland will be introduced. After that the best performing method will be chosen, the dataset developed and validated and the uncertainties of the dataset estimated.

1. Introduction

The main objective of the EU funded project ENSEMBLES is to develop a common ensemble climate forecast system for Europe for use across differing time and spatial scales. Currently, prediction of climate variability is problematic, because of uncertainties in climate models. To reduce the uncertainty, validation of the models, i.e. comparison of the models with observations, will be performed in the project. Following *Osborn and Hulme* [1998], the outputs of climate models are regarded closer to area-average values than point data. This means that area-average observations will be needed and used for validation.

Gridded data are not only of major importance for the validation of regional and global climate models [*Caesar, et al.*, 2006], but are used for validation of models in many other fields of study, such as terrestrial biospheric modelling [*Piper and Stewart*, 1996], hydrological modelling [*Goovaerts*, 2000] and soil quality modelling [*Shen, et al.*, 2001]. Gridded datasets have been developed for different countries, different temporal and spatial resolution and with the use of different interpolation methods (see Table 1 for a reasonably comprehensive but incomplete list). *Piper and Stewart* [1996] were the first to grid daily data.

[*] Corresponding author: M.Haylock@uea.ac.uk

Until then mostly monthly [e.g. *New, et al.*, 2000; *Price, et al.*, 2000] and annual data [e.g. *Hutchinson*, 1995; *Vicente-Serrano, et al.*, 2003] were interpolated.

In the project we will develop a daily dataset for Europe on a high resolution grid, up to about 0.2 degrees. One of the objectives of ENSEMBLES is the development of higher resolution climate models to provide more regionally detailed climate predictions and better information on extreme events. This means that the space aggregation level of the observations must be of a high resolution. To gain information on extreme events, the time aggregation level should be of high resolution. A daily time scale is chosen. The climate variables to be validated are minimum, maximum and mean temperature, precipitation, sea level pressure and snow depth. Before we develop the dataset, it is important to choose the best performing interpolation method. The objective of this study is to compare available interpolation methods and to choose the one/several we will be using for the development of the dataset. It is probable that the best performing method might differ between the climate variables. The period chosen for this study is the first of January 1961 to the 31st of December 1990. When the dataset is developed it will cover the period from at least 1960 to present, depending on the available station data.

Vicente-Serrano et al. [2003] list global interpolators (e.g. Polynomial Equations and Empirical Multiple Regression), local interpolators (e.g. Thiessen Polygons or Natural Neighbour Interpolation, Inverse or Angular Distance Weighting and Splines), geostatical methods (various versions of Kriging) and combinations of the above. According to this study, the best performing interpolation method ‘varies as a function of the area and the spatial scale desired for mapping’. Also the temporal scale and the climate variable interpolated influence the decision about the best performing method. Daily precipitation, for instance, is discontinuous, both in space and time. However, monthly precipitation and daily temperature are more continuous and the best performing interpolation method for these climate variables and time scales may be different.

The methods can be divided into stochastic and deterministic methods. This distinction is a fundamental philosophical difference that has divided the statistical community [*Laslett*, 1994b]. Deterministic methods assume that the observed data are the only possible set of observations that could have occurred. Stochastic methods assume that the observations are just one of many possible “realisations”. The stochastic methods therefore seek to determine the underlying large-scale spatial variation in the data and use this to interpolate. An important consequence is that stochastic methods do not generally reproduce the observations in the interpolated data (“inexact interpolation”) whereas deterministic methods do (“exact interpolation”). We have included two stochastic methods and two deterministic methods.

Several factors may be important in interpolation. In many of the studies [e.g. *Price, et al.*, 2000] elevation is seen as the main determinant of the climate variables and used in interpolations. Also distance to coast, seasonality, latitude-longitude differences and synoptic state are mentioned as being of importance [*Perry and Hollis*, 2005]. A correlation analysis will be performed to investigate the relation between the climate variables and the variables and the specific one under study for Europe (to be included in a journal paper). Some interpolation methods have the opportunity to introduce important factors, which may improve the interpolation methods.

Shen et al. [2001] compared several methods for daily data. They conclude that most interpolation methods give too smooth results and are thus best for fitting mean conditions.

The number of days with precipitation is also often not adequately represented [*Shen, et al.*, 2001]. To test these problems in our methods, we will emphasise the comparison of extreme values and the number of rain days between observed and interpolated values. Another problem with the current interpolation methods is stated by *Hewitson and Crane* [2005]. According to them the current methods estimate values at new point locations, instead of determining area averages. They use interpolation to a very dense grid and average these values to the grid they need. *Feng et al.* [2004] also average area averages from a dense grid. We will use the same method. The dense grid will be a 0.1 degree grid. This will be averaged to a 0.2 and a 0.5 degree grid.

In this study the following four methods are compared. The Natural Neighbour Interpolation (NNI) is, according to *Shen et al.* [2001], one of the simple interpolation methods that can retain the variability of the data. Angular Distance Weighting (ADW) has worked well for monthly data [*New, et al.*, 2000] and has been used for daily data before [*Caesar, et al.*, 2006]. Kriging and Thin Plate Splines (TPS) have been used widely [e.g. *Hutchinson*, 1993; *Hutchinson*, 1995; *Hutchinson and Gessler*, 1994; *Kyriakidis, et al.*, 2001; *Rubel and Hantel*, 2001] in different forms and for different applications and we have, therefore, chosen to use both as well. Unfortunately, so far, there has not been time to include Reduced Space Optimal Interpolation (RSOI), chosen because of its extensive use by the Hadley Centre for gridding monthly sea level pressure (SLP) and sea surface temperature (SST) and daily SLP, and the newly developed Conditional Interpolation (CI) method. So far, CI can only be used for precipitation data, so we will only apply it for precipitation. This method is chosen, because it seems to perform well [*Hewitson and Crane*, 2005] and the use of self organizing maps to define the synoptic state, which is used in the interpolation, seems plausible. RSOI and CI will be compared to the current results of the other four methods later. Except for the CI, we will test all methods for all climate variables. It would be possible to use different methods for the different climate variables. We will, however, try to use the same method for the whole area. This means that the method will have to cope with a heterogeneous surface.

There are several requirements for all interpolation methods: that they are fully objective and readily reproducible; that they are fully automated; and that they are computationally efficient. Objectivity is a strict criterion that ensures that the method chosen for the final gridding is identical to the method that is verified to be the best. Currently none of the methods apart from Natural Neighbour Interpolation are at that stage. The current station selection for the other methods is based on a fixed search radius of 250 km for precipitation and 400 km for the other variables. This step must ultimately become part of the objective methodology. Full automation follows from the criterion of objectivity but is also dictated by the sheer volume of the data required. Interpolating every day for thousands of grid points and seven variables requires complete automation for timely execution. The data volume also dictates the efficiency criterion. An attractive method that could be included is stochastic simulation, a geostatistical method that generates many possible "realisations" of the interpolated data depending on the spatial properties of the input data. This provides potentially valuable estimates of uncertainty but requires too much in the way of computer resources for such a large area and high time resolution.

In this paper the data, the interpolation and the validation methods will be explained in Section 2. In Section 3 the results achieved so far will be given and in Section 4 these results are discussed. Section 5, the conclusions and further plans, is followed by a time table of the work to be done before the delivery of the dataset (in 18 months time).

Table 1: A reasonably comprehensive, but incomplete list of gridded datasets (rr is precipitation, tx ,tn and tg are maximum, minimum and mean temperature, mslp is mean sea level pressure, T is temperature, RH is relative humidity and evap. is evaporation. # means that the value is unknown)

Reference	Time aggregation level	Space aggregation level	Period	Area	Climatic factors	Interpolation method
Ansell et al. 2005	daily	5 degree	1850-2003	European-North Atlantic	mslp	reduced space optimal interpolation
Caesar et al. 2005	daily	2.5 x 3.75 degree	1946-2000	global	tx, tn	angular distance weighting
Feng et al. 2004	daily	1 degree	1951-2000	China	tx, tn, tg, skin surface T, RH, wind speed, wind gust, sunshine duration, rr and pan evap.	a modified Cressman scheme
Hewitson and Crane, 2005	daily	0.1 degree	1950-2000	South Africa	rr	conditional interpolation
Hutchinson, 1995	annual	0.01 degree	1955-1988	part of Australia	mean rr	thin plate splines
Kiktev et al. 2003	daily	2.5 x 3.75 degree	1950-1995	global	climate extremes	angular distance weighting
Kyriakidis et al. 2001	daily	1 km	Nov 81 - Jan 82	California	rr	alternative forms of kriging
New et al. 2000	monthly	0.5 degree	1901-1996	global	tg, diurnal T range, rr	angular distance weighting
Piper and Steward, 1996	daily	1 degree	1987	global	tx, tn, rr	nearest neighbor interpolation scheme
Price et al. 2000	monthly	#	1961-1990	parts of Canada	tx, tn, rr	thin plate smooting splines and GIDS
Rubel and Hantel, 2001	daily	1/6 degree	1996-1998	Baltic sea catchment	rr	ordinary block kriging
Shen et al. 2001	daily	polygons	1961-1997	Alberta	rr,tg	nearest station assignment
Vicente-Serrano et al. 2003	annual	#	1950-2000	part of Spain	rr, tg	several

2. Methods

2.1 Strategy

In this section the data used for the comparison of the methods will be described in section 2.2. Then, in Section 2.3, the six methods (NNI, ADW, Kriging, TPS, RSOI and CI) are briefly introduced with their advantages and disadvantages. Interpolation is undertaken to the location of the stations, so cross-validation (see Section 2.4) is possible. The climate variables mean, maximum and minimum temperature, precipitation, mean sea level pressure and snow depth are interpolated. For the interpolation of temperature, anomalies are used, for precipitation the percentage of the monthly sum. These transformations provide a more stationary series for interpolation, a condition to which the stochastic methods are particularly sensitive. This should reduce the importance of local factors such as elevation, however future work will assess whether added skill could be obtained by incorporating this. The back calculations for both temperature and precipitation are currently performed using the station data. In the future monthly-interpolated grid values will be used, which are derived from many more stations than we have daily data for to encompass the higher availability of digitised monthly series. Because these monthly data don't exist for snow depth, the actual value is interpolated rather than the proportion of the monthly mean. For mean sea level pressure the elevation issue doesn't exist and for this climate variable also the actual value is interpolated.

Except for the NNI, all methods use a search radius to select stations used for gridding. This search radius is defined for ADW to be the correlation decay distance (CDD). This is based on a graph of the distance on the x-axis set out against the correlation between stations on the y-axis, the CDD is the distance at which the mean correlation falls below $1/e$ [*Caesar, et al.*, 2006; *New, et al.*, 2000]. Stations within a search radius as large as the CDD are used for calculation of the grid-point value. For the other methods the search radius can be taken as small or as large as one wants. To be able to compare the methods, for now, for all methods, the value of 250 km for precipitation and 400 km for the other climate variables have been taken. When the methods are fine tuned, real CDDs will be calculated. Also the possibility of using a variable CDD in space will be studied. For ADW, the 10 stations with the highest weight (see Section 2.3.2) were taken for the interpolation. The weight of the other stations is so low, that they have little impact on the result. For the other methods the 30 closest stations are used, because the method improves when more than 10 stations are used and with more than 30 stations the cross validation becomes too slow. Also the number of stations used for the interpolation will be subject to further investigation. For each grid point the stations used for interpolation are determined on each day depending on data availability.

2.2 Data

The data used is collated by KNMI. Details are given in deliverable 5.8 of the ENSEMBLES project. 1445 precipitation stations, 959 minimum temperature stations, 955 maximum temperature stations, 630 mean temperature stations, 231 mean sea level pressure stations and 126 snow depth stations have been used for the interpolation. For some methods, in sparse areas there may be insufficient stations present within the search radius of the method. In those cases interpolation has not been performed. The final gridded product will address this issue by interpolating to sparse regions but providing large uncertainty estimates.

2.3 Interpolation methods

2.3.1 Natural Neighbour Interpolation (NNI)

Two of the most elementary interpolation methods are Thiessen polygons and triangulation. Thiessen polygons divide the interpolation area into regions where all points in a region are closest to the same station. Therefore isolated stations are represented by larger polygons. Interpolation is just a matter of deciding into which polygon a point falls and assigning it the value of the closest station. The interpolated values are therefore a function of just one station. Triangulation is a slight improvement as it uses the nearest three points by dividing the area into triangles devoid of stations. The interpolated values are taken from the planes passing through the triangle of points surrounding the point.

Natural Neighbour Interpolation (NNI) takes the best of both methods and was developed by *Sibson* [1981]. It is included here as a fast and simple baseline method that has been used for many years as a standard part of the library of graphics functions provided by NCAR. NNI is different to the other methods in that it objectively chooses the number of neighbours from which to interpolate based on the geometry. All the other methods in this comparison select a set of neighbours based on either a fixed radius or fixed number of points. The weights for each station are selected based on the proportional area rather than distance. NNI produces an interpolated surface that has a continuous slope at all points, except at the original input points. The resulting interpolated surface can be visualized as producing a snugly fitted sheet stretched over all of the input data. *Sibson* [1981] overcame the problem of discontinuous gradients at the stations locations by calculating their gradients from the natural neighbours. This method, known as Natural Neighbour Nonlinear Interpolation, has not been implemented here.

Advantages of NNI are that the method is computationally efficient and grid-point estimates are not subject to under- or overshooting, because they cannot exceed the magnitude of the highest/lowest value in the contributing data points [*New, et al.*, 2000]. Moreover, NNI can often retain the variance of daily data [*Shen, et al.*, 2001]. Disadvantages of NNI are that, according to *Borrough and McDonnell* [1998], the method is not suitable for variables like precipitation and temperature, that have gradual spatial variation. Furthermore, the method does not take account of station distance [*New, et al.*, 2000] and it does not allow factors like elevation to be considered [*Goovaerts*, 2000].

2.3.2 Angular Distance Weighting (ADW)

The Angular Distance Weighting method used in this study is a modified version of Shepard’s algorithm [*Shepard*, 1968], also used by *New et al.* [2000] for monthly data and by *Caesar et al.* [2006] for daily data. Stations are selected using the search radius explained above. Then weights are assigned to stations, using the distance between the grid point and the stations and the angle between the stations. This way stations that are closest to the grid point and stations that are not surrounded by other stations get a higher weight than stations further away and stations surrounded by other stations. If there are more than 10 stations used for the interpolation, the 10 stations with the highest weights are taken and a grid point value is calculated by multiplication of the weights and the value of the stations. If there are less than 3 stations within the search radius, the value for the grid point cannot be calculated.

An advantage of ADW is that the method is flexible when irregularly spaced station data are gridded. As for NNI, the grid point estimates are not subject to under- and overshooting [*New, et al.*, 2000]. *Hewitson and Crane* [2005], however, question the assumption that, under convective rainfall systems, the spatial representation of a station is proportional to, for example, the inverse distance squared from the station. Another disadvantage is concluded by *Kiktev et al.* [2003], who find that the use of ADW caused significant reduction in frost days, an increase in very warm night-time temperatures and a decrease in the number of consecutive dry days in their results.

3.2.3 Kriging

Kriging is an interpolation method used extensively in the geosciences. The technique has a long history of development since the pioneering mathematical formalisation of Kolmogorov in the 1930s [*Kolmogorov*, 1939]. However, it was not until the published work of *Matheron* [1963], in recognition of the efforts of *Krige* [1951; 1966], and the later work of *Journel and Huijbregts* [1978], that the method became popular in the mining industry for mapping ore reserves.

Kriging is a stochastic method that applies the general methodology known as best linear unbiased estimation (BLUE): the "estimated" (interpolated) value is a linear combination of the predictors (nearby stations) such that the sum of the predictor weights is 1 (unbiased) and the mean squared error of the residuals from the interpolating surface is minimized (best estimate). The interpolating surface is therefore a local function of the neighbouring data, but conditional on the data obeying a particular model of the spatial variability.

The key to kriging is deciding which statistical model best describes the spatial variation of the data. This is determined by fitting a theoretical function to the experimental "variogram": the absolute difference between stations as a function of their distance. For this comparison we have selected the best of five models: Gaussian, Exponential, Spherical, Hole-effect and Power. These are the most common functions used for variogram modelling and their mathematical description can be found in most geostatistical texts [e.g. *Kitandis*, 1997; *Webster and Oliver*, 2001]. The best function was determined by fitting each of these non-linear functions to the experimental variogram using the method of *Marquardt* [1963], and selecting the model with the lowest Chi-squared statistic. The variogram modelling code was custom written and the Kriging code was adapted from GSLIB [*Deutsch and Journel*, 1998]. All variogram models were three parameters, one of which was the nugget variance to allow for spatial variation at a scale not resolved by the station network.

One of the preferred methods for the creation of climatological maps is residual Kriging. It is widely used within climatology [*Tveito, et al.*, 2005]. A disadvantage is that for Kriging questionable assumptions are needed. Also the interpretation of the variograms, prior to fitting the surface, is subjective [*Price, et al.*, 2000]. Kriging would perform very well in fields that are relatively stationary in time and homogeneous in space. That is not the case with daily climate data [*Shen, et al.*, 2001].

3.2.4 Thin Plate Splines (TPS)

Thin-plate spines share many of the same characteristics as kriging. Indeed there have been several comparisons of the two methods [*Hutchinson*, 1993; *Hutchinson and Gessler*, 1994; *Laslett*, 1994a]. Both BLUE methods, these two methods differ in the function used to model spatial variability. The five common statistical models, or "covariance functions", used in kriging derive from the geological background of kriging but are not based on any theory of the distribution of mineral deposits [*Handcock, et al.*, 1994]. These functions are chosen merely from the requirements that the spatial covariance be positive definite. Splines use a different covariance function and one that is rarely used in kriging [*Hutchinson and Gessler*, 1994]. Although there have been some attempts to unify the two approaches, such as the method of Matern Splines [*Handcock, et al.*, 1994], the two methods are usually treated as independent.

TPS is accurate, operationally straightforward and computationally efficient for the spatial interpolation of annual mean rainfall. All records can be used, no matter how short. By minimizing the generalized cross validation, the amount of data smoothing can easily be optimized and TPSs are appropriate for use across large heterogeneous areas [*Hutchinson*, 1995]. Compared to Kriging, for TPS there is no need for prior estimation of spatial auto-covariance structure, which might be a difficult process [*Hewitson and Crane*, 2005; *Hutchinson*, 1995]. Disadvantages are that the method is generally unsuitable for interpolation of anomaly fields, because in these fields, mainly for precipitation, sharp spatial discontinuities occur. The method can be adapted for that, but then considerable under- and overshooting takes place [*New, et al.*, 2000]. The amount of data required is very high [*Hutchinson*, 1995] and in data-sparse regions results may be misleading (e.g. negative precipitation) [*Price, et al.*, 2000].

3.2.5 Reduced Space Optimal Interpolation (RSOI)

RSOI is a stochastic method that has been used at the Hadley Centre for developing their gridded datasets. It has been used for gridding monthly sea surface temperatures (SST – [*Kaplan, et al.*, 1997]) and sea level pressure (SLP – [*Allan and Ansell*, 2006]), and more recently for daily SLP [*Ansell, et al.*, 2006]. The technique is a combination of data reduction and least squares optimal estimation. The data reduction involves using the most recent data-rich sub period to compute leading empirical orthogonal functions (EOFs), which are used as a first order model for the time variation in the data. Optimal interpolation techniques are then employed in the EOF ("reduced") space, greatly enhancing computational efficiency. The optimal interpolation shares many similarities with kriging and splines in that it seeks to find a least squares unbiased solution to the interpolated surface. Unlike the other methods which treat each day individually, it has the added advantage of using time correlations in interpolation.

An advantage of RSOI is that it takes the heterogeneous patterns of spatial covariability, which is expected in complex topography, objectively into account. Within the model it is also possible to calculate error estimates for the reconstructed data [*Schmidli, et al.*, 2001]. However, for daily mean sea level pressure, the gridded values are smoother than the observed values [*Ansell, et al.*, 2006].

2.3.6 Conditional Interpolation (CI)

Conditional Interpolation is a method developed and used by *Hewitson and Crane* [2005]. So far, the method has only been used for the interpolation of precipitation and we will use it accordingly. After development of a target grid and selection of observational stations (all stations within a search radius), the synoptic states are defined. The spatial distribution of rainfall is a reflection of the synoptic state, so to define the synoptic state, self organizing maps (SOMs) of the precipitation data are used (see *Hewitson and Crane* [2002] for an explanation of SOMs). The technique using SOMs is a clustering technique. For each station 24 (a user defined number) maps are created of archetype precipitation fields. These precipitation fields are sorted, for example from heavy to no rain and from rain in the south east to rain in the north west. The daily patterns will then be associated with the closest precipitation field. The synoptic state is used for the definition of whether there is rain or no rain at a location and is used when weights are defined.

After definition of the synoptic state, the interpolation is performed. First the phase (wet/dry), as determined with help of the SOMs, is interpolated. The weights used are calculated in about the same way as for ADW. For the wet grid points an interpolation of the magnitude of the precipitation is performed. To do this the distance weights used earlier are adjusted with the synoptic state. Stations with a synoptic state similar to the grid point will get a higher weight than stations with a very different synoptic state. Interpolation to grid points has been performed, but the results have been averaged to obtain area averages.

CI is a new method, and therefore not many advantages and disadvantages are known. An advantage of CI is that area averages are calculated, instead of grid point values [*Hewitson and Crane*, 2005]. However, we can easily apply this area averaging technique to the other interpolation methods. It also gives the opportunity to use the synoptic state in interpolation, a unique approach in all the interpolation methods.

2.4 Cross validation

The aim of all interpolation methods is to predict point values at defined locations. Later on the point values will be averaged over an area, but for now it is useful to have point values, because it facilitates comparison between the methods. The method used in this study to compare the methods and to define the best performing method is cross-validation. Instead of interpolation to a regular grid, the stations are one by one deleted from the dataset and interpolation is performed for the location of the deleted station. The interpolated values can then be compared to the observed values for each station. Three skill scores were calculated to compare the four interpolation methods: the root-mean-square error (RMSE), the Pearson correlation coefficient (R) and the Linear Error in Probability Space score (LEPS).

RMSE is a measure of deviation from the observed value, caused by either biases in the interpolated data, poor modelling of the daily variability, or incorrectly capturing the variance of the station. Its dependence on the squared error means that it is not resistant to outliers deviating from a Gaussian distribution. The units of RMSE are the same as the data and we would expect it to be highest in those months and regions with higher values of the data (e.g. higher rainfall). RMSE is bounded below by 0 (best case) but is unbounded above.

R also depends on squared deviations and so is similarly not a resistant measure. However this statistic removes the effect of any bias in the interpolated data and highlights just problems with modelling the daily variability. Problems with correctly capturing the variance will not be highlighted as the measure normalises the observed and modelled values by their standard deviations. The statistic is standardised and, therefore, can be compared across regions and months. R takes values between -1 (worst case) to 1 (best case).

LEPS is a skill measure used extensively in seasonal forecasting [*Ward and Folland*, 1991]. It measures the error in probability space with reference to climatology. Like RMSE it indicates any deviation from the observed value. Unlike RMSE it is resistant to outliers as it assumes no particular statistical distribution. LEPS takes the values -100 (worst case) to 100 (best case). A value of 0 indicates the same skill as predicting the median for every case.

According to *Shen et al.* [2001] cross-validation is both commonly used and the most effective method to assess the error of climate data estimation. On the other hand, *Hewitson and Crane* [2005] state that the objective of gridding data is to obtain area averaged values. In this sense, acquired low errors for stations may not give equally good area averaged values for the grids. Also comparison of methods in this way does not necessarily result in the real best performing method being revealed [*Hewitson*, private communication]. A solution is to compare averaged interpolated values to averaged observed values. For the United Kingdom and Switzerland dense station networks exist. Unfortunately, so far, we haven't been able to use the dense station networks. Therefore, for now, only cross-validation has been performed for the whole of Europe using single stations.

3. Results and discussion

Tables 2, 3 and 4 show the three skill measures for all methods and variables. For each line in the table, the method with the best skill measure is highlighted in bold. The RMSE is a single measure calculated across all stations and months. This will therefore be biased to the stations with higher rainfall and snowfall, however, it should weight each station more equally for the other variables. R and LEPS are calculated for each station individually and then averaged across all stations.

All models are doing very well, with low RMSE and high LEPS and R scores. The results indicate that the two deterministic methods NNI and ADW are performing better than the two stochastic methods TPS and Kriging, with NNI doing the best in the most cases. The largest differences in the skill measure between models occur for the RMSE statistic whereas the LEPS and R measures are more similar. This perhaps indicates that the differences in RMSE between methods are more indicative of the models' ability to capture those stations with higher rainfall and snowfall. This is also indicated by comparing the rr variable (daily rainfall) with r% (daily rainfall as a proportion of monthly total). While ADW has the second lowest (best) RMSE for rr it has the highest (worst) RMSE for r%. The LEPS score does not show this problem as it weights each station equally. The LEPS and R scores (Tables 3 and 4) are generally very similar across the models.

Tables 2, 3 and 4 also indicate which of the variables are better interpolated. Although RMSE cannot be compared across variables, the LEPS and R measures can. When averaged across all models, the temperature variables have the highest R values, followed by pressure, rainfall and snow depth. This is the same for LEPS with the notable exception that snow depth has the highest LEPS scores. This may indicate problems with calculating the skill of the snow depth interpolation across all months when snow is present at many sites only during winter.

Table 2: RMSE for each method for the seven variables rr (rainfall mm), pp (sea level pressure hPa), sd (snow depth mm), tg (mean temperature ^oC), tn (minimum temperature ^oC), tx (maximum temperature ^oC) and r% (rainfall as a proportion of the monthly total). Best performing method highlighted in bold.

	NNI	ADW	Kriging	TPS
rr	**3.17**	3.32	3.83	3.37
pp	3.55	**3.08**	4.13	3.61
sd	64.4	**45.3**	64.2	62.1
tg	**1.35**	1.41	1.69	1.37
tn	**1.79**	1.82	2.20	1.81
tx	**1.65**	1.68	2.01	1.71
r%	**0.0514**	0.0653	0.0624	0.0523

Table 3: As for Table 2 but for LEPS.

	NNI	ADW	Kriging	TPS
rr	**71.6**	69.3	69.8	69.6
pp	70.4	**73.4**	65.3	70.6
sd	88.5	**91.6**	90.0	88.4
tg	**86.3**	85.5	83.3	86.1
tn	**80.2**	79.1	76.1	79.9
tx	**84.9**	83.7	81.8	84.4
r%	**69.5**	66.5	68.3	67.4

Table 4: As for Table 2 but for R.

	NNI	ADW	Kriging	TPS
rr	**0.774**	0.757	0.703	0.770
pp	0.934	**0.951**	0.916	0.937
sd	0.575	0.556	**0.629**	0.532
tg	**0.982**	0.981	0.974	**0.982**
tn	**0.965**	0.964	0.951	0.964
tx	**0.977**	0.975	0.968	0.975
r%	**0.749**	0.688	0.673	0.745

The spatial distribution of skill was examined by plotting the maps of LEPS and R averaged across all models (Figures 1 and 2). All variables show regional variation in both skill measures, apart from the temperature variables that show consistently high R. The most distinct regional variation occurs for the LEPS score for rainfall, with an obvious meridional gradient in skill, showing higher skill in the south than in the north. This is most likely indicative of the higher number of rainy days in the north. The long summer Mediterranean drought would be expected to greatly increase the skill in this region for any interpolation method. The maps also show that increased skill comes with an increase in station density. The LEPS scores are consistently highest over the Netherlands, although a less complex orography may also aid interpolation in this region. While the R scores for the three temperature variables show little difference between the variables, the LEPS scores show greater skill for mean and maximum temperature than minimum.

The results described above are results so far and the comparisons will continue. The two methods RSOI and CI will be automated and compared to the four existing methods. Moreover all methods need fine tuning. Part of the fine tuning may be the adaptation of parameters, like for example the search radius for neighbouring stations, which is currently based on a preliminary analysis of the correlation decay distance averaged across the entire region. An important addition that affects all methods to improve skill is the transformation of the input data to reduce skewness. The stochastic methods in particular require that the station data be close to normally distributed. Currently this is not the case for precipitation and perhaps snow depth.

All the methods weight to a lesser extent those stations further from the grid point. This implies that the data show similar spatial correlation in all directions, a case known as “isotropy”. The direction of the prevailing weather and complex topography means this is not generally the case, but the problem is compounded by the fact that the direction and degree of anisotropy varies between regions, variables and time of the year. Therefore there is a need to automatically determine anisotropy and compensate. We are not aware of any studies that have comprehensively dealt with this problem in the climate literature.

As already described in Section 2.4, cross-validation may not be a good method to compare interpolated and station values, because we want to make sure the dataset captures the spatial patterns instead of calculating the exact values for the stations. As soon as we receive the dense station data from Switzerland and the UK, we will also compare area averages of interpolated grid point values with area averages of the dense station data and base the choice of the best performing method also on that validation method.

4. Conclusion and further plans

In this study, so far, four interpolation methods (NNI, ADW, Kriging and TPS) have been compared in a cross-validation exercise, using the statistics RMSE, R and LEPS. The interpolation methods have not been fine tuned and two interpolation methods (RSOI and CI) have not yet been automated. Cross-validation is only one of several possible techniques to compare different interpolation methods, but was for now the only implemented method.

Despite all these problems, first results have been shown. Considering the improvements that will still be possible, the results are good. So far, it seems that the two simpler, deterministic methods NNI and ADW provide the best results. Averaged over all interpolation methods, the results are best for temperature. Some spatial differences can be observed (e.g. higher LEPS skills for precipitation in the south than in the north) as well as some density or orographic differences (higher LEPS and R skills in the Netherlands, with many stations and a flat surface). This will be studied further.

In the coming 18 months the dataset will be developed and delivered to the climate modellers of the ENSEMBLES project. First of all the two remaining methods CI and RSOI will be automated. Then all interpolation methods will be fine tuned. The comparison will continue, using also a comparison of area-average values for the UK and Switzerland, and the best performing method will be chosen. Then the dataset will be developed, validated and, finally, uncertainties will be estimated (see Table 5).

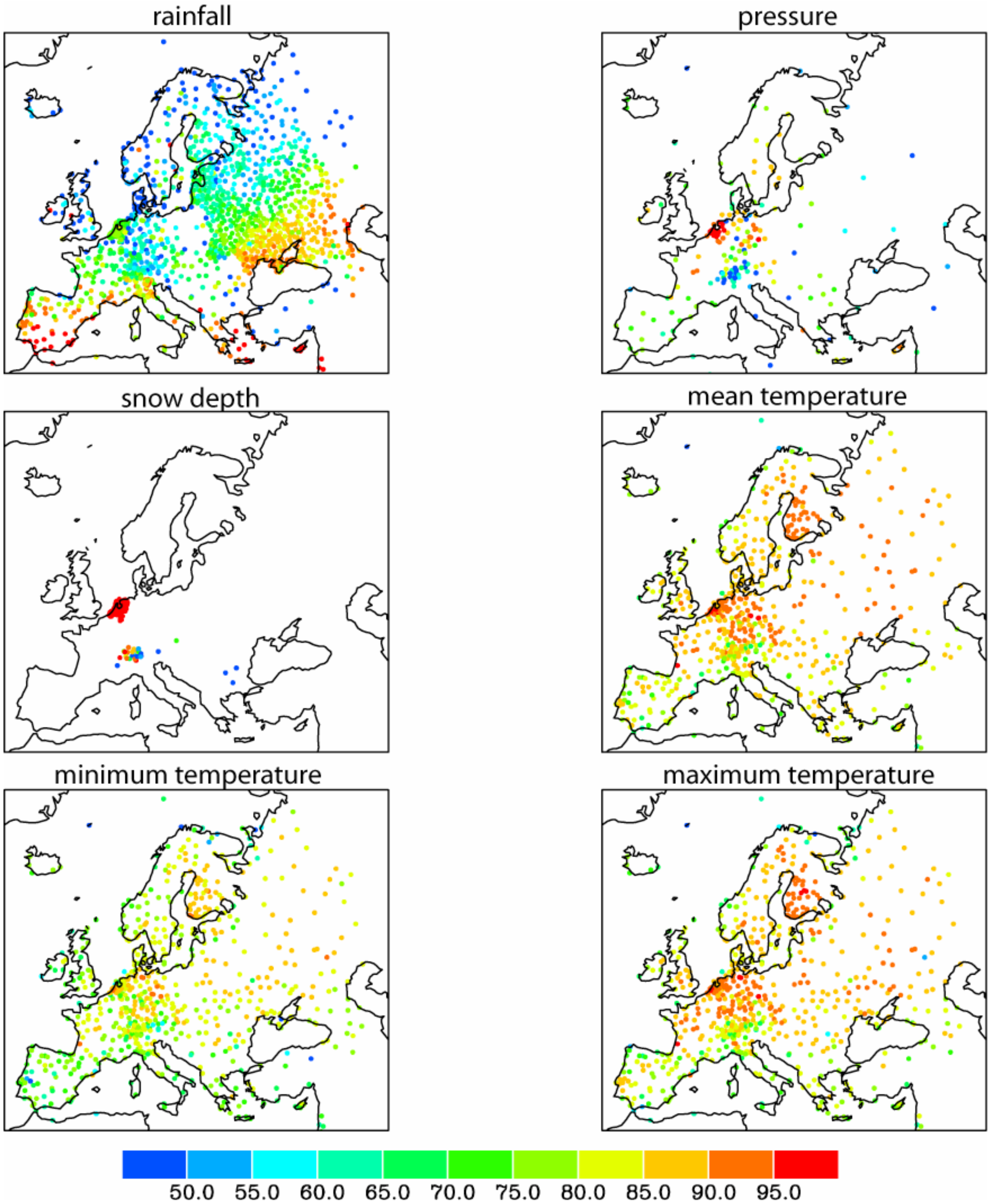

Figure 1: LEPS skill scores averaged across all methods

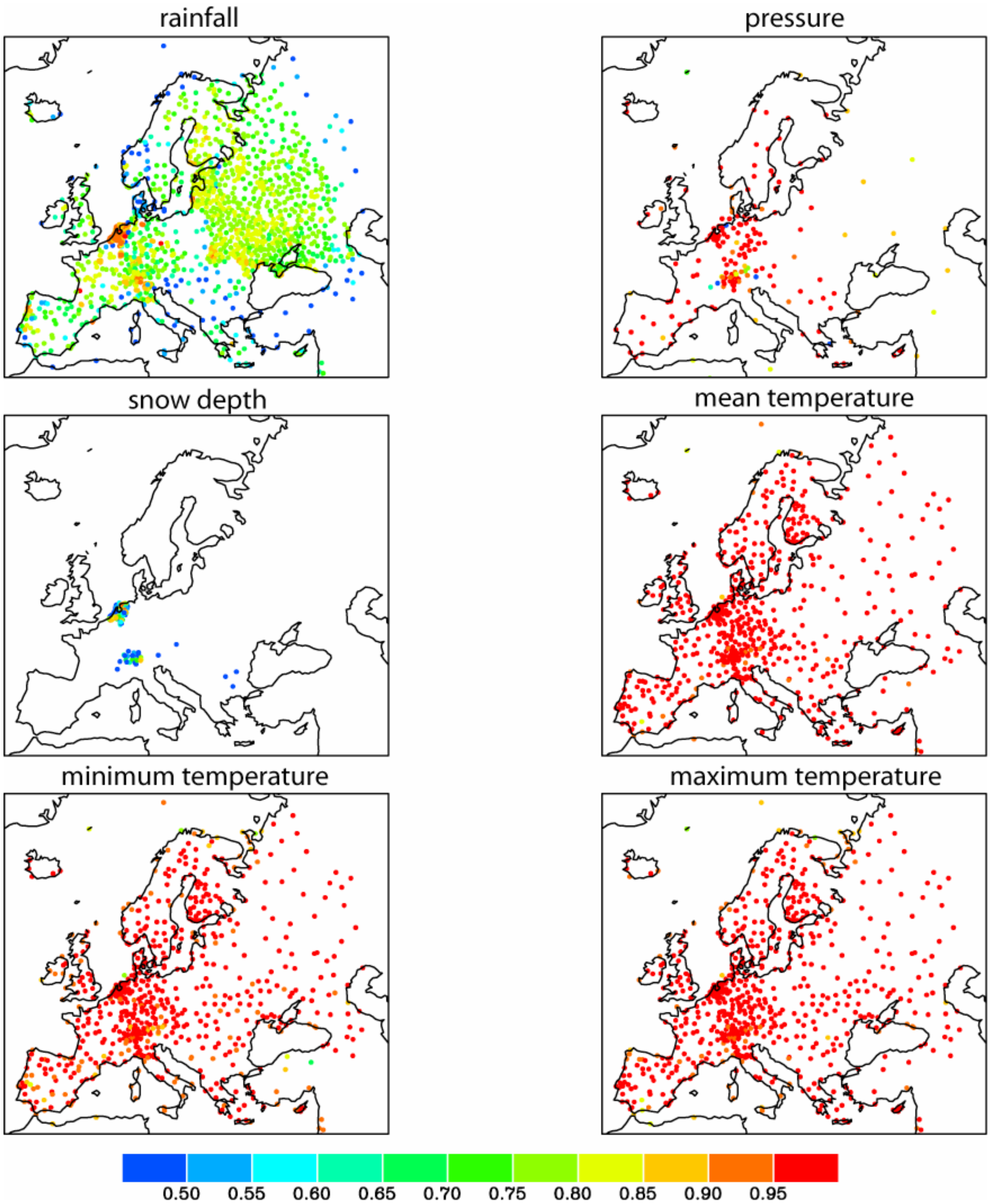

Figure 2: R skill scores averaged across all methods

Table 5: Timetable

Action	From month	To month
Application of Conditional Interpolation and Reduced Space Optimal Interpolation to available station data	18	20
Final comparison and choice of 'best performing' method	21	22
Improve and fine tune the chosen method(s)	23	25
Development of dataset	26	26
Validation of dataset	27	31
Estimation of the uncertainty of the dataset	32	36

Acknowledgements

This study is part of the EU-funded project ENSEMBLES. N. Hofstra is also funded by the Dutch talentenbeurs and the Dutch Prins Bernhard Cultuurfonds beurs. We would like to thank all institutes (see Table 1 of Deliverable D5.8 of ENSEMBLES) that made data available for the study.

References

Allan, R. J., and T. J. Ansell (2006), A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850-2004 *Journal of Climate*, *Accepted*.
Ansell, T. J., et al. (2006), Daily mean sea level pressure reconstructions for the European - North Atlantic region for the period 1850-2003, *Journal of Climate*, *Accepted*.
Borrough, P. A., and R. A. McDonnell (1998), *Principles of geographical information systems*, Oxford University Press, Oxford.
Caesar, J., et al. (2006), Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded dataset, *Journal of Geophysical Research*, *111*, D05101 05101-05110.
Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 369 pp., Oxford University Press, Oxford.
Feng, S., et al. (2004), Quality control of daily meteorological data in China, 1951-2001: a new dataset, *International Journal of Climatology*, *24*, 853-870.
Goovaerts, P. (2000), Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *Journal of Hydrology*, *228*, 113-129.
Handcock, M. S., et al. (1994), Kriging and Splines - an Empirical-Comparison of Their Predictive Performance in Some Applications - Comment, *Journal of the American Statistical Association*, *89*, 401-403.
Hewitson, B. C., and R. G. Crane (2002), Self-organizing maps: applications to synoptic climatology, *Climate Research*, *22*, 13-26.
Hewitson, B. C., and R. G. Crane (2005), Gridded area-averaged daily precipitation via conditional interpolation, *Journal of Climate*, *18*, 41-57.
Hutchinson, M. F. (1993), On thin plate splines and kriging, in *Computing and Science in Statistics*, edited by M. E. Tarter and M. D. Lock, pp. 55-62, Interface Foundation of North America, University of California, Berkeley.
Hutchinson, M. F. (1995), Interpolating mean rainfall using thin plate smoothing splines, *International Journal of Geographical Information Systems*, *9*, 385-403.
Hutchinson, M. F., and P. E. Gessler (1994), Splines - More Than Just a Smooth Interpolator, *Geoderma*, *62*, 45-67.
Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, Academic Press, London.
Kaplan, A., et al. (1997), Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures, *Journal of Geophysical Research*, *102*, 27835-27860.
Kiktev, D., et al. (2003), Comparison of modeled and observed trends in indices of daily climate extremes, *Journal of Climate*, *16*, 3560-3571.
Kitandis, P. K. (1997), *Introduction to Geostatistics*, Press Syndicate of the University of Cambridge, Cambridge, UK.
Kolmogorov, A. N. (1939), Sur l'interpolation et extrapolation des suites stationnaires, *Comptes Rendus de l''academie des Sciences de Paris*, *208*, 2043-2045.
Krige, D. G. (1951), A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand, Masters thesis, University of Witwatersrand.

Krige, D. G. (1966), Two-dimensional weighted moving average trend surfaces for ore evaluation., *Journal of the South African Institute of Mining and Metallurgy*, *66*, 13-38.
Kyriakidis, P. C., et al. (2001), Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics, *Journal of Applied Meteorology*, *40*, 1855-1877.
Laslett, G. M. (1994a), Kriging and Splines - an Empirical-Comparison of Their Predictive Performance in Some Applications, *Journal of the American Statistical Association*, *89*, 391-400.
Laslett, G. M. (1994b), Kriging and Splines - an Empirical-Comparison of Their Predictive Performance in Some Applications - Rejoinder, *Journal of the American Statistical Association*, *89*, 406-409.
Marquardt, D. W. (1963), An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, *11*, 431-441.
Matheron, G. (1963), Principles of geostatistics, *Economic Geology*, *58*, 1246-1266.
New, M., et al. (2000), Representing twentieth-century space–time climate variability. Part II: development of 1901–96 monthly grids of terrestrial surface climate, *Journal of Climate*, *13*, 2217-2238.
Osborn, T. J., and M. Hulme (1998), Evaluation of the European daily precipitation characteristics from the Atmospheric Model Intercomparison Project, *International Journal of Climatology*, *18*, 505-522.
Perry, M., and D. Hollis (2005), The generation of monthly gridded datasets for a range of climate variables over the UK, *International Journal of Climatology*, *25*, 1041-1054.
Piper, S. C., and E. F. Stewart (1996), A gridded global data set of daily temperature and precipitation for terrestrial biospheric modeling, *Global Biogeochemical Cycles*, *10*, 757-782.
Price, D. T., et al. (2000), A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data, *Agricultural and Forest Meteorology*, *10*, 81-94.
Rubel, F., and M. Hantel (2001), BALTEX 1/6-degree daily precipitation climatology 1996-1998, *Meteorology and Atmospheric Physics*, *77*, 155-166.
Schmidli, J., et al. (2001), Reconstruction of mesoscale precipitation fields from sparse observations in complex terrain, *Journal of Climate*, *14*, 3289-3306.
Shen, S. S. P., et al. (2001), Interpolation of 1961-97 daily temperature and precipitation data onto Alberta polygons of ecodistrict and soil landscapes of Canada, *Journal of Applied Meteorology*, *40*, 2162-2177.
Shepard, D. (1968), A two-dimensional interpolation function for irregularly spaced data, paper presented at 23rd Natl. Conf. ACM, Association for Computing Machinery, New York.
Sibson, R. (1981), A brief description of natural neighbour interpolation, in *Interpreting Multivariate Data*, edited by V. Barnett, pp. 21-36, John Wiley and Sons, Chichester.
Tveito, O. E., et al. (2005), A GIS-based agro-ecological decision system based on gridded climatology, *Meteorological Applications*, *12*, 57-68.
Vicente-Serrano, S. M., et al. (2003), Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature, *Climate Research*, *24*, 161-180.
Ward, M. N., and C. K. Folland (1991), Prediction of seasonal rainfall in the North Nordeste of Brazil using eigenvectors of sea surface temperature, *International Journal of Climatology*, *11*, 711-743.
Webster, R., and M. A. Oliver (2001), *Geostatistics for Environmental Scientists*, Wiley, Chichester, UK.