Intern rapport ; IR 2009-04

Interpolation methods for climate data

literature review

R. Sluiter

De Bilt, 2009

KNMI intern rapport; IR 2009-04

De Bilt, 2009

PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
De Bilt
The Netherlands
http://www.knmi.nl
Telephone +31(0)30-220 69 11
Telefax +31(0)30-221 04 07

Auteur: Sluiter, R.

Interpolation methods for climate data
Literature review

Dr. R. Sluiter

KNMI, R&D Information and Observation Technology
De Bilt, 19 November 2008
Version 1.0

Table of contents

1 Introduction

Providing climatological and meteorological data products covering the whole country as maps or gridded datasets is an important task for KNMI. To calculate these maps, the observations of meteorological stations in the Netherlands need to be interpolated.

Recent research on interpolation of climatological and meteorological information with the support of GIS has shown that interpolation has a large development potential within climatology and meteorology (e.g. COST-719 (KNMI, 2005)). At the same time the demand for gridded (interpolated) data products is increasing: numerical weather models are working at higher spatial resolutions and may be initiated by gridded data from observations, or may provide "first guess fields" that may be used in the interpolation process. Hydrological models are also improving and need continuous high resolution climatological input data, etc.

To fulfil the internal and external demand for interpolated data, KNMI has started this research project to improve the existing climatological interpolation. This facility is the whole chain from the measurement data in the database to various map products. The project includes the introduction of new interpolation methods, new interpolation applications, and new mechanisms of presentation and downloading of interpolated data. In the existing situation, yearly, monthly and daily maps are generated for a number of meteorological elements like precipitation, temperature, wind and insolation. The maps are based on the measurements of typically 30-36 automatic meteorological stations scattered around the Netherlands. In addition, daily precipitation is measured at around 325 locations by volunteers. For the interpolation of all these maps the same spline technique is used and the results are available only as annotated images on the website. This research project aims at improving the interpolation process by taking into account ancillary data like land-sea gradients, altitude, rain radar and yearly trends of environmental factors like circulation patterns and land-use. As each variable requires a specific interpolation technique, different methods will be considered and documented. In the project we will use a set of GIS applications for interpolation like R for statistical computing and graphics (including GSTAT) and ArcMap.

1.1 Objectives

The first objective of this report (technical report 1) is to acquire knowledge through a literature review about interpolation of climatological and meteorological interpolation, including the use of ancillary data. The focus is on the Dutch situation and is driven by the user requirements of KNMI and external parties.

The second objective of this report is to give an overview of the current products and possible interpolation and data providing applications.

The objectives of technical report 2 are to document the methodology and procedures in detail, based on the pilot study of phase 2.

2 Data

2.1 Source data

The source datasets are the observations of the automatic meteorological stations and the observations of the manual voluntary network. These observations are stored in a database (KIS) and are quality controlled. The observation network for the different meteorological variables is not constant in time and space, which has consequences for the interpolation strategy (section 3). Below a short description of the measurement network is given based on Buishand (Buishand et al., 2008).

On 325 locations cumulative precipitation (including snow) of the last 24 hours is collected at 08:00 UT using manual rain gauges. Since 1946 the number of locations did not change significantly, but many stations have been moved or replaced by other stations. The locations of the stations are shown in figure 1.

Since 1999 the station network is active in its current composition with 19 manned stations and 17 unmanned stations, shown in figure 1 (since 2003 all stations are automatic). These stations acquire all relevant meteorological elements. Since 1956 hourly measurements are available on 15 stations and from 1971 to 1990 3-hourly measurements were available on 30 additional "termijnstations".

Around 1990 automatic stations were introduced which caused a large change in the network. The "termijnstations" were discontinued.

Global radiation is measured since 1957 starting with one station (De Bilt), expanding to 5 (1965), 6 (1984), 13 (1987), 32 (1999) and 33 at present.

Figure 1 - Automatic meteorological station network in the Netherlands (left) and voluntary network for precipitation (right)

2.2 Rasterized data products

Based on the observations of the meteorological stations many rasterized data products are calculated. At present these maps are only available as annotated images, the new production facility shall provide "real data" on which calculations can be made. Table 1 gives an indication of the raster products that are produced at present. This table is based on the products that are distributed on the website of the climatological division (KNMI, 2008a), the website of the weather department (KNMI, 2008b), publications in the climate atlas 1971-2000 (Heijboer and Nellestijn, 2002), "Maandoverzicht van het weer in Nederland (MOW)" and "Maandoverzicht neerslag en verdamping in Nederland" (MONV). The table can be expanded by the user requirements for the climate scenarios and from external parties like the 'Waterdienst". For example, by the "Nationaal Hydrologisch Instrumentarium" (NHI) historical measurements, actual measurements and forecasts (model fields) are needed.

Important issues are the definition of the grid size (cell size) the geographical projection/coordinate system, time-steps, processing method and delivery method. User requirements for cell size range from 1 to 10 km. In the Netherlands different projections and coordinate systems are used. Most national datasets are in the Dutch RijksDriehoekstelsel (RD) but radar data has a polar stereographic projection etc. Processing can be done highly automatic (website products) or with manual intervention (climate atlas). Data delivery methods should be in line with national and international data exchange systems. These issues will be discussed in the next sections.

2.3 Ancillary data

This research project aims at improving the interpolation process by taking into account ancillary data like land-sea gradients, altitude, rain radar and yearly trends of environmental factors like circulation patterns and land-use. At present, known gradients like land-sea gradients are not automatically included in the interpolation process and the observation network is not distributed in a way that these patterns are adequately presented in the interpolated maps. Much research has been done to include spatial trends in the interpolation process (Dobesch et al., 2007). Based on that research and knowledge available at KNMI, the following ancillary data is considered important for the Dutch situation:

- Land sea gradients: for example important for sunshine, temperature, duration of precipitation, wind speed and wind gusts.
- Large water masses like IJsselmeer: for example important for sunshine, temperature, wind speed and duration of precipitation.
- Elevation: the "higher" areas in the Netherlands (Veluwe, Heuvelrug and Limburg) receive relatively more precipitation.
- Land use: urban effects on temperature and roughness, dominating land-use near meteorological stations, effects on fog occurrence.
- Latitude: north-south effects in temperature.
- Precipitation radar: when available, cumulative precipitation derived from the precipitation radar can be used to optimize the interpolation of precipitation measured at the stations (Schuurmans et al., 2007). KNMI operates weather radar since 1959. Since 1998 5-minute reflection images with 2.4km resolution are available, since January 2008 with a resolution of 1 km. Since 2003 corrected cumulative 24-hour images are available as well as corrected 3-hourly reflection images.
- Incomplete time-series: information from incomplete measurement series to optimize the prediction at selected locations.
- Input from numerical weather models to provide reliable first guess fields.

For each climatological element the potential ancillary data sources are mentioned in table 1. In phase 2 of the research they will be studied in more detail.

			Meteo	Climate	Climate	Climate	Climate	Climate	1981-2010 / Climate Atlas			Schetsboeken				
Variable	**Description**	**Unit**	**10 min**	**daily**	**decade**	**monthly**	**running monthly**	**yearly**	**monthly**	**season**	**year**	**season**	**year**	**# stations**	**Ancillary spatial data**	**Source**
Airpressure	Air Pressure	hPa							x		x			35	Coastal effects	Climate Atlas
Evap	Makkink Evaporation	mm			x	x					x	x	x	15	Coastal effects, Circulation patterns	KNMI MONV
Fog-days	Mean number of days with fog	days							x		x			15	Fog index (not operational)	Climate Atlas
P	Mean Cumulative precipitation	mm		x	x	x	x		x		x	x	x	325	Radar, Elevation	KNMI climate website
P-deficit	Cumulative precipitation deficit, april 1 - september 30	mm										x	x	325	Elevation	Climate Atlas
P-duration	Mean duration of precipitation	hour								x	x			35	Coastal effects, Circulation patterns	Climate Atlas
P-intensity-days>0.3	Mean number of days with precipitation intensity >= 0.3 mm.	days									x	x	x	35	Radar, Elevation	Climate Atlas
P-intensity-days>1.0	Mean number of days with precipitation intensity >= 1.0 mm.	days									x			325	Radar, Elevation	Climate Atlas
P-intensity-days>10.0	Mean number of days with precipitation intensity >= 10.0 mm.	days									x		x	325	Radar, Elevation	Climate Atlas
P-recurrence10y	Highest 24h cumulative precipitation per 10 years												x	325		Schetsboeken
Rad	Global radiation	joule/cm2				x				x	x			33	Coastal effects	KNMI climate website
RelHum	Relative humidity	%								x				35	Coastal effects	Climate Atlas
RelHum12	Relative humidity at 12.00 UT (+- daily min)	%								x				35	Coastal effects	Climate Atlas
Snow-cover	Snow cover	cm		x										325	x	KNMI climate website
Snow-days	Mean number of days with snow	days									x			325	x	Climate Atlas
Sun	Total hours of sunshine	hour		x					x		x	x	x	35	Coastal effects	KNMI climate website
Sun-days<20	Mean number of days with sunshine duration =<20%	days									x			35	Coastal effects	Climate Atlas
Sun-days>20<50	Mean number of days with sunshine duration >20% and =<50%	days									x			35	Coastal effects	Climate Atlas
Sun-days>50<80	Mean number of days with sunshine duration >50% and <80%	days									x			35	Coastal effects	Climate Atlas
Sun-days>80	Mean number of days with sunshine duration =>80%	days									x			35	Coastal effects	Climate Atlas
T-da	Mean daily amplitude of temperature	deg-C								x				35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
Thunder	Mean number of days with thunderstorms	days								x	x			15	Distance to the coast, Circulation patterns	Climate Atlas
T-max	Maximum temperature	deg-C		x					x		x	x		35	Coastal effects, Land Use, Circulation patterns	KNMI climate website
T-max-days<0	Mean number of days with maximum temperature < 0	days									x			35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-max-days>15	Mean number of days with maximum temperature => 15	days									x			35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-max-days>20	Mean number of days with maximum temperature => 20	days									x		x	35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-max-days>25	Mean number of days with maximum temperature => 30	days									x		x	35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-mean	Mean temperature	deg-C	x	x					x		x			35	Coastal effects, Land Use, Circulation patterns	KNMI climate website / KNMI actual weather website
T-min	Minimum temperature	deg-C		x					x		x	x		35	Coastal effects, Land Use, Circulation patterns	KNMI climate website
T-min-days<0	Mean number of days with minimum temperature < 0	days									x		x	35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-min-days<-10	Mean number of days with minimum temperature < -10	days									x			35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
T-min-days<-5	Mean number of days with minimum temperature < -5	days									x			35	Coastal effects, Land Use, Circulation patterns	Climate Atlas
Visibility	Maximum visibility	m	x											15	x	KNMI actual weather website
W-gust	Max wind-gust	m/s		x										35	Coastal effects, Roughness	KNMI climate website
W-m10m-bft	Mean wind /10 minutes	bft	x											15	Coastal effects, Roughness	KNMI actual weather website
W-m10m-ms	Mean wind /10 minutes	m/s	x											15	Coastal effects, Roughness	KNMI actual weather website
W-mh	Mean wind /hour	m/s		x						x	x			35	Coastal effects, Roughness	KNMI climate website

Table 1 - Indication of the raster products that are produced at present.

3 Interpolation methods

3.1 Introduction

Many different interpolation methods exist. In this review, which is mainly based on Dobesch et al. (2007) and Tveito et al. (2006). I classify these methods in three main categories: deterministic, probabilistic and other methods. Deterministic methods create a continuous surface by only using the geometric characteristics of point observations. Probabilistic methods are based on probabilistic theory and use the concept of randomness, the realized interpolated field is one out of many realizations. Probabilistic methods allow to include the variance in the interpolation process and to compute the statistical significance of the predicted values. "Other methods" are applications that are specially developed for meteorological purposes using (a combination of) deterministic and probabilistic methods. As one of the objectives is to take into account ancillary data into the interpolation process, techniques using ancillary data will receive more attention.

3.1.1 Global and local interpolators

A first distinction can be made between global and local interpolators. Global interpolators use a single function on an entire dataset to map a whole area, local interpolators use single or multiple functions on subsets of the dataset within a predetermined window. Global interpolators generally produce smoother surfaces.

3.1.2 Exact and approximate interpolators

A second distinction can be made between exact and approximate interpolators. Exact interpolators reproduce the original values at the data points on which the interpolation is based. Approximate interpolators do not reproduce the original values, assuming uncertainty in these values and will reduce errors by the effect of smoothing.

3.1.3 Temporal scales and interpolation strategy

The temporal scales determine the interpolation strategy. According to Tveito (2007) climate reference maps (normals) should be based on the interpolation of absolute filtered, but biased values. Monthly or seasonal maps should be based on the interpolation of the anomalies or normalized values. Daily and/or synoptic maps (10-minutes) should be based on the interpolation of absolute values which are often noisy and biased.

In practice, a two step procedure is followed for interpolation of climate elements: 1) normalization of the in-situ data to obtain the spatial stationarity assumption (section 3.1.5) 2) interpolation of the normalized fields using any applicable method.

To calculate anomalies, the gridded long term mean is needed. Ideally, the long term mean should be calculated over a time-span of 30 years which is a difficult task in the Dutch situation because of the changes in the measurement network (section 2.1). Alternatively a shorter period can be selected for elements with less spatial variability. In general, for temperature absolute anomalies are used (difference between measurement and mean), for precipitation relative anomalies are used (quotient of measurement and mean). Disadvantage of the use of anomalies is that the entire dataset should be reprocessed as soon as new long term means are available (Buishand et al., 2008).

3.1.4 Measures of success

The success of the interpolation method can be assessed in three ways: data splitting, cross validation and calculation of the Kriging variance.

With data splitting the dataset is splitted up in one part for estimation and one part for validation. This method is only applicable when the number of observations is large and regularly sampled.

The cross validation approach considers all data in the validation process: the interpolation procedure is executed several times, each time one observation is left out and calculated on the basis of the neighbouring observations. The procedure is repeated until for each point the difference between the calculated and observed value is estimated.

Calculation of the Kriging variance is possible with Kriging methods. Kriging variance is also estimated on points where no observations exist and therefore provides a spatial view on the measure of success.

3.1.5 Assumptions

Some interpolation methods, especially geostatistical methods, bring their own assumptions: stationarity, intrinsic hypothesis, isotropy and being unbiased.

According to Tveito et al. (2006) stationarity is described as "*the condition that the probability distribution of the variable is constant in time and/or space, meaning that the same probability distribution function should be expected anywhere/anytime. This is usually a too stricter criterion bearing on natural physical processes, and in geo-sciences second order (2.O) stationarity is often used*". In other words: "stationarity" implies that every location in the study area has a potential statistical distribution with the same mean and standard deviation.

According to Tveito et al. (2006) the intrinsic hypothesis is described as "*An assumption used in geostatistics assuring that a semivariogram requires second order stationarity of the increments of a random function. This is a weaker demand than second order stationarity of the covariance normally used in stochastic theory. For practical purposes this distinction makes no difference*". The "intrinsic hypothesis" implies that the difference between two values taken at two different locations come from a distribution which depends only on the distance (and possibly relative direction) of the two locations

Isotropy is the property of being directionally independent. The opposite is anisotropy: something which is anisotropic, may appear differently, or have changing characteristics in different directions.

Unbiased means that within the observation network no values are more likely to be included than other values within the population. An example of a biased network is a meteorological station network in a mountainous region where more meteorological stations are located in the lower valleys than on the higher mountain plateaus and tops.

3.2 Deterministic Methods

3.2.1 Nearest Neighbourhood (NN) and triangulation

The nearest neighbourhood method assigns the value from the nearest observation to a certain grid cell. NN is also known as Thiessen or Voronoï method. With triangulation Triangular Irregular Networks (TINs) are created from the observation points. The method gives the slope of the surface based on the three neighbouring points; this slope can be used to calculate the value at given grid-cells.

Both methods are fast and simple, but the interpolated fields do not look realistic in all cases. Disadvantage of both methods is the lack of success measures. The methods work best when many data points are available. Ancillary data cannot be incorporated.

The application of both methods in meteorology is limited, but they may be successfully used for dense measurement networks. Tveito (2007) uses NN to visualize the irregularity of station networks within the NORDGRID project. Meteo Norway also uses triangulation as a quick solution for gridding daily precipitation.

3.2.2 Inverse Distance Weighting (IDW)

IDW is an advanced nearest neighbour approach that allows including more observations than only the nearest points. The value at a certain grid cell is obtained from a linear combination of the surrounding locations. The weight of each observation is determined by the distance, the distance function is non-linear. IDW is an exact interpolator.

The method is fast, easy to implement and easily "tailored" for specific needs. The method allows anisotropy in the source data. Ancillary data cannot be incorporated. Measure of success is through cross validation. The method tends to generate "bull's eye patterns". There is no extrapolation: all interpolated values are within the range of the data points (De Smith et al., 2007).

IDW is widely used in meteorology. Examples are interpolation of anomalies within REGNIE (REGionalisierung räumlicher NIEDerschlagsverteilungen) by Deutscher Wetterdienst (Buishand et al., 2008) and the generation of monthly gridded datasets for 36 parameters at 5 km resolution by UK Metoffice (Perry and Hollis, 2005) using a combination of linear regression models and IDW. In Portugal IDW was compared to several other methods for calculating mean monthly temperature using 84 meteorological stations; IDW results were worse than Kriging and linear regression model results. Spain uses IDW to calculate monthly climate anomaly maps of temperature, precipitation and insolation.

3.2.3 Polynomial functions (splines)

Polynomial functions are methods that fit trend functions through the observations by x-order polynomials. Most popular are spline functions. In general they are global interpolators and fulfil the criteria of exact interpolators by fitting many polynomials in regions with overlapping neighbourhood. To ensure that results do not show strongly oscillating patterns in between the observation points, algorithms are used to smooth the resulting surfaces. These algorithms mainly describe the differences between the available spline methods.

It is assumed that the measurements are without error, because errors are not suppressed in the algorithm. Ancillary data can be included when using advanced methods like ANUSPLIN (Hutchinson, 2008) . Measure of success is through cross validation.

Polynomial functions are regarded as a good method for interpolation of monthly and yearly climate elements but are less suitable on higher temporal resolutions like days and hours. At present, tension splines are used for the interpolation of all climatological and meteorological elements at KNMI. In Portugal splines were compared to several other methods for calculating mean monthly temperature using 84 meteorological stations, splines performed worse than Kriging, linear regression models and IDW (Tveito et al., 2006).

3.2.4 Linear regression

Linear regression expresses the relation between a predicted variable and one or more explanatory variables. In its simples form a straight line is fitted through the data points. Linear regression models are most often global interpolators. Linear regression models are deterministic, but by considering some statistical assumptions about the probability distribution of the predicted variable the method becomes stochastic. In that case the standard error can be calculated, the inference about the regression parameter and the predicted values can be assessed and the prediction accuracy can be calculated.

For deterministic linear regression models the assumption is that the regression model could be interpreted on the basis of physical reasons, for stochastic linear regression models a normal distribution and spatial independence is also assumed. No extrapolations are allowed from the theoretical perspective. Ancillary data can be included using multiple regression. For deterministic linear regression models the measure of success is through cross validation. For stochastic linear regression models it can be measured by the explained variance and the regression standard error.

Linear regression has to be executed by a standard statistical program; maps can be calculated using calculator functions in a GIS environment. Application examples are given by UK Metoffice that uses linear regression for the generation of monthly gridded datasets for 36 parameters at 5 km resolution in combination with IDW (Perry and Hollis, 2005). The PRISM method is also based on linear regression models (Szentimrey et al., 2007). In

Portugal Linear regression is used in combination with residual Kriging to calculate mean monthly temperature (see section 3.3.6).

3.2.5 Artificial Neural Networks (ANN)

Neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data (Tveito et al., 2006). A feature of ANNs is their adaptive nature through "learning during" the classification/prediction process which makes them powerful and popular for solving complex, non-linear processes. The learning process may be supervised or unsupervised. Because many types of ANNs exist and the theory is quite complex I will not discuss the method in detail, a more detailed description of ANNs can be found in (Tveito et al., 2006).

Most ANNs are "black-box" models (little feedback is given about what is happening with the data in the model) and need much computational power.

In Portugal ANNs were compared to several other methods for calculating mean monthly temperature using 84 meteorological stations; ANN performed slightly less than residual Kriging.

3.3 Probabilistic methods

Probabilistic methods incorporate the concept of randomness and include methods like linear regression models (section 3.2.4), geostatistics and optimum interpolation (section 3.3.1). In this review I will focus on geostatistical methods called Kriging.

McDonnell & Burrough (1998) have shown that for applications in geosciences, when data is sparse (but not *too* sparse) Kriging is the best interpolation technique available.

Kriging interpolation starts with the recognition that the spatial variation of a continuous attribute is often too irregular to be modelled by a simple function. The variation can be better described by a stochastic surface with an attribute known as a regionalized variable.

The regionalized variable theory assumes that the value of a random variable Z at (x) is given by:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''$$

Where: m(x) = a deterministic function describing a structural component of Z at x.
ε'(x) = a random spatially correlated component.
ε''(x) = a residual non-spatially correlated term, or noise (Nugget variance).

When structural effects have been accounted for and the variation is homogenous in its variation, the semivariance γ (h) can be estimated by:

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}\{z(x_i) - z(x_i + h)\}^2$$

Where: n = number of pairs of sample points of observations of the values of attribute z separated by distance h.

A plot of γ (h) against h is called a semivariogram and gives a quantitative description of the regionalized variation (see figure 2). An important factor of the variogram is the range, which describes the distance when the data points become spatially independent. The variogram can be used to estimate the optimal weights λ_I needed for interpolation. The value z(x) for an unsampled point is then calculated with:

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i * z(x_i)$$

The principle is shown in figure 3 (Burrough and McDonnell, 1998).

In general Kriging is a relatively fast interpolator that can be exact or smoothed depending on the method. The method is flexible with input and output data: many outputs can be generated besides the prediction maps like predictions errors and probabilities. Drawback of the flexibility is that it may require a lot of decision making. Measure of success is through the prediction errors or through cross validation. In section 3.3 an overview is given of the different Kriging methods, the overview is mainly based on Tveito et al. (2006; 2007). The review is limited to the essential properties of the methods, mathematical descriptions will be given in technical report 2 for the methods that will be used at KNMI.

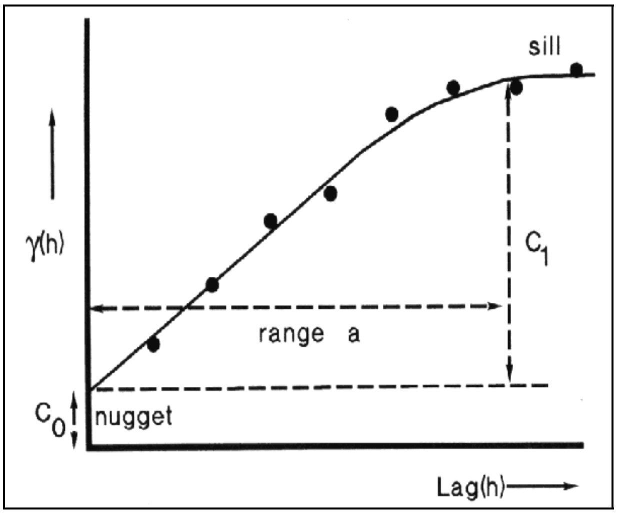

Figure 2 - Variogram example

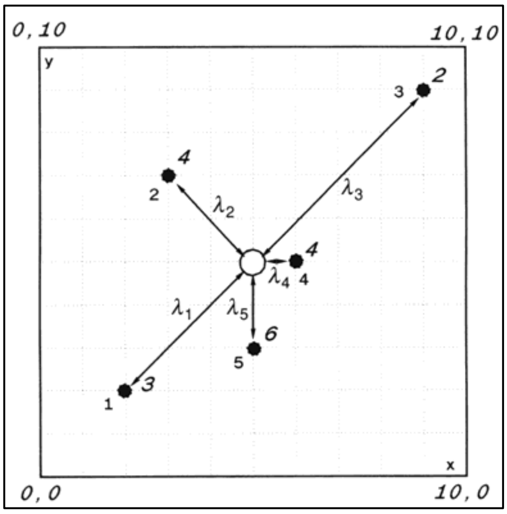

Figure 3 - The principle of Kriging

3.3.1 Optimum interpolation

Optimum interpolation was developed by Gandin (Szentimrey et al., 2007) for meteorology at the same time that Kriging was developed in the geosciences. It is based on a spatial correlation function, assumes second order stationarity and requires a first guess field like model output from numerical weather prediction models. In that combination it is suitable for interpolations at short time resolutions. It is used in Sweden, for example within the ERAMESAN study (Tveito et al., 2006).

3.3.2 Ordinary Kriging

Ordinary Kriging is the basic form of Kriging as described in section 3.3. The prediction by ordinary Kriging is a linear combination of the measured values. The spatial correlation between the data, as described by the variogram, determines the weights.

As the mean is unknown, fewer assumptions are made. The method assumes intrinsic stationarity (section 3.1.5), unfortunately meteorological variables are often not stationary. In some case this problem can be eliminated by using different sizes and shapes of the search neighbourhood. Ordinary Kriging is frequently applied in meteorology, often as part of residual Kriging (section 3.3.6) or indicator Kriging (section 3.3.7).

3.3.3 Simple Kriging

Simple Kriging is Ordinary Kriging with a known mean. Therefore it is slightly more powerful than ordinary Kriging; however the mean is often difficult to derive.

3.3.4 Cokriging

Cokriging is an extension of standard Kriging using a multivariate variogram or covariance model and multivariate (ancillary) data. With cokriging the estimations on a location are based on a linear weighted sum of all examined variables. When more than one co-variable is considered the method may become highly complex.
In Meteorology cokriging is often applied. Schuurmans (2007) used cokriging to combine station data with precipitation radar data (as well as Kriging with an external drift).

In general, cokriging gives better results when the number of covariables are (much) higher that the variable of interest and when the spatial correlation between variables and co-variables is high. The assumptions are the same as for ordinary Kriging plus assumptions with respect to cross-variogram model estimation.

3.3.5 Universal Kriging

Universal Kriging is also known as "Kriging with a trend/external drift". It uses a regression model as part of the Kriging process to model the mean value expressed as a linear or quadratic trend.

Kriging with an external drift is a very common method in meteorology. It assumes intrinsic stationarity and the number of "drift" variables should be much higher that the variable of interest. Examples of applications are given by ENSEMBLES (Mitchell, 2008) that uses external drift Kriging for incorporating elevation into the interpolation of temperature on European scale and 2-D universal Kriging (in combination with indicator Kriging) to model precipitation amounts. Schuurmans (2007) used Kriging with an external drift to combine station data with precipitation radar data (as well as cokriging). In Slovenia universal Kriging is used to calculate precipitation normals.

3.3.6 Residual Kriging

Residual Kriging is also known as detrended Kriging. With residual Kriging the residuals from a previously fitted regression (for example base on linear regression, section 3.2.4) are interpolated using ordinary Kriging. Assumptions are the same as for universal Kriging (section 3.3.5). For measure of success both the methods related to linear regression and related to Kriging can be used.

Residual Kriging is widely used in meteorology (Tveito et al., 2006). In Poland it was found that residual Kriging performed best of all methods for monthly and seasonal means of air temperature and precipitation totals (Dryas and Ustrnul, 2007). In Portugal residual Kriging was compared to several other methods for calculating mean monthly temperature using 84 meteorological stations. Multiple regression with altitude and distance to the sea in combination with residual Kriging proved to be the best method. The same approach showed good results in Spain, which has a very dense measurement network for temperature (1445 stations, one station every 18 km^2). Residual Kriging in combination with linear regression was also successful in Slovenia, the Nordic countries as a whole Norway and France for monthly and daily temperature (Tveito et al., 2006).

3.3.7 Indicator Kriging

Indicator Kriging is interpolation of a categorical variable like precipitation occurrence. It uses thresholds to create binary data (indicator values, 0/1) and then uses ordinary Kriging for interpolation. The results indicate the probability that a certain threshold is exceeded. Indicator Kriging can be seen as special version of disjunctive Kriging (see section 3.3.9). According to Tveito (Tveito et al., 2006) the method is not suitable for data having a trend.

Examples of applications are given by ENSEMBLES (Mitchell, 2008) that uses indicator Kriging in combination with universal Kriging for the interpolation of precipitation on European scale.

3.3.8 Probability Kriging

Probability Kriging is a non-linear method using indicator variables. The method can be seen as a form of cokriging where the first variable is the indicator and the second variable is the original untransformed data (Tveito et al., 2006). Probability Kriging is more powerful than indicator Kriging but it requires more calculations because cross variances have to be fitted. Like indicator Kriging, the method is not suitable for data having a trend.

3.3.9 Disjunctive Kriging

Disjunctive Kriging is a non-linear procedure in which the dataset is transformed using a series of additive functions. Disjunctive Kriging assumes that all data pairs originate from a bivariate normal distribution.

3.3.10 Stratified Kriging

In general Kriging is applied on complete datasets and is therefore a global method. When there is enough information to classify the area into meaningful sub-areas and there are enough data tot compute variograms for each sub-area it is possible to carry out the interpolation in each separate area (Burrough and McDonnell, 1998). In the Netherlands, the only dataset that is potentially suitable for this method is the daily precipitation network, although the regional differences in the Netherlands are small (Buishand et al., 2008).

3.4 Other methods

3.4.1 MISH

When using geostatistical methods the predictors refer to a single realization in time. When interpolating meteorological data more information about the data is available in the form of long climate data series. Information from these data-series can be used to estimate the spatial trend differences and the covariances statistically. Consequently, when these parameters are known they provide much more information than when using only the predictors of a single realization. The MISH method (Meteorological interpolation based on Surface Homogenized Data Basis) (Szentimrey et al., 2007) is especially developed to incorporate information from time series in the interpolation procedure. It consists of two modules: MISH for interpolation and MASH (Multiple Analysis of Series for Homogenization Software) to obtain homogenized data series. The method is capable to include ancillary data. Within COST-719 the method was intensely "advertised", but use of the method outside Hungary is sparse.

3.4.2 PRISM

PRISM is a knowledge-based system that uses point measurements of precipitation, temperature, and other climate elements to produce continuous, digital coverages. PRISM incorporates expert knowledge of rain shadows, temperature inversions, coastal effects, and more. PRISM coverages are used with Geographic Information Systems (GIS) to construct maps and perform many types of analysis. The PRISM Group (formerly SCAS) was established at Oregon State University (OSU). Applications of PRISM Group products are wide-ranging, the PRISM Group is responsible for nearly all major climate mapping efforts at the federal level in the United States. It is also engaged in international modelling and analysis projects." (Citation taken from (PRISM, 2008)).:

In Europe, PRISM was applied to calculate the long-term precipitation climatology for the Alpine region by Schwarb (Schwarb, 2001). PRISM showed to be very powerful in areas where the station network is unrepresentative for the variation in topography.

Traditional methods showed better results in areas with dense networks. At the time of writing it is unknown if the PRISM software is available for external parties.

4 Data processing and distribution

Many software packages exist for interpolation of point data. A distinction can be made between automatic processing and manual processing (= processing with user intervention). The real-time products of the KNMI weather website and the products at the KNMI climate website shall be made automatic while products like scenarios and the climate atlas probably need manual processing, especially to produce customized visualizations. Therefore the software may vary, but all products should be produced exactly in the same way on both the automatic and the manual platform. Based on this literature review and the software already available at KNMI the following software may be considered for processing.

4.1 Generic Mapping Tools (GMT)

GMT is an open source collection of ~60 tools for manipulating geographic and Cartesian data sets (including filtering, trend fitting, gridding, projecting, etc.) and for producing Encapsulated PostScript File (EPS) illustrations (Wessel, 2008). The existing climatological interpolation facility is based on GMT. GMT does only support trend surfaces (splines) and nearest neighbour interpolation. GMT is available for UNIX, Windows and Mac OS platforms.

4.2 ESRI ArcGIS

ESRI ArcGIS (ESRI, 2008) is the most widely used commercial GIS in the world and the most important module, ArcMap, supports most interpolation methods except linear regression modelling which should be done using an external statistical package. ESRI ArcMap is very powerful for visualization and processing with user intervention. KNMI has ArcGIS licenses through VenW.

4.3 R/GSTAT

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms Windows and MacOS (R-Project, 2008). Included in R is the module Gstat. Gstat is an open source computer code for multivariable geostatistical modelling, prediction and simulation. Gstat can calculate sample variograms, fit valid models, show variograms, calculate (pseudo) cross variograms, fit valid linear models of coregionalization (S extension only), and calculate and fit directional variograms and variogram models. The R/GSTAT combination includes all necessary functions for statistical analysis and interpolation and has proven to be useful in an operational environment (INTAMAP, 2008). It is the preferred solution for automatic processing and it may be useful to provide the input data for manual processing.

4.4 MISH/MASH

The MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) method was developed at the Hungarian Meteorological Service for the spatial interpolation of surface meteorological elements. It is a purely meteorological system in aim and in respect of the tools used. It requires all meteorological information combined with climate and supplementary model or background information (cited from (Szentimrey et al., 2007). MISH uses climatological information from time-series to optimize the statistical parameters in geostatistical models. Within COST-719 the method was intensely “advertised”, but use of the method outside Hungary is sparse.

4.5 Data distribution

To provide the climate datasets, we could develop an INSPIRE compliant infrastructure based (EU, 2008) on OGC web services: Web Mapping Services (WMS) for online visualization and Web Coverage Services (WCS) for downloading raster data (OGC, 2008). Integration of the interpolation techniques with the data provision services has to be done in an early stage; therefore in this research will build further on the knowledge and products of the COST-719, ADAGUC and INTAMAP projects.

5 Discussion

In Tveito (2007) a recommendation framework for the development of an interpolation approach is presented. This framework consists of four steps: 1) choose the right interpolation method, 2) correct use of the interpolation method, 3) test several methods, 4) validation, 5) future research. I'll use this framework to discuss the results of this literature review.

5.1 Choose the right interpolation method

As shown there is no general method that is suitable for all problems: it depends on the nature of the variable and on the time-scale on which the variable is represented. According to Tveito (2007), climate reference maps (normals) should be based on the interpolation of absolute values. Monthly or seasonal maps should be based on the interpolation of the anomalies or normalized values. Daily and/or synoptic maps (10-minutes) should be based on the interpolation of absolute values.

The most used and promising techniques are universal Kriging and linear regression models in combination with Kriging (residual Kriging) or IDW. The choice of IDW or Kriging should be based on the accuracy obtained during the phase 2 pilot study. Also the number of observations plays a role: the 36 measurements of the meteorological station network are at the lower range for successful application of Kriging. This may be solved by combining more realizations in time or variogram pooling, but IDW or splines may be a better choice too. IDW and splines are easier to implement as no variograms are needed for the calculation. To calculate anomalies the long-term mean should be calculated over a period of 30 years, which may be difficult because of the changes in the measurement network in the past. The use of Artificial Neural Networks (ANN) is not preferred because this is a relative new methodology for interpolation, is highly "black box" and requires excessive computing power. At present, it is unknown how useful the MISH method is for KNMI, but as the MISH approach looks promising, it could be tested in a small pilot. The same applies for PRISM, at the time of writing it is unknown if the software is available for external parties. For the special case of optimizing the interpolation of precipitation by precipitation radar data we could continue the work of Schuurmans.

Choosing the right method also includes the selection of coordinate systems and cell size. The coordinate system ideally should match the system that is used inside and outside KNMI. Rijksdriehoekstelsel (RD) is a good candidate as well as the European system ERTS89 that may be adopted by INSPIRE. Note that the choice of a coordinate system becomes less crucial because the proposed delivery systems based on OGC services support transformations between major coordinate systems and projections.

Cell size is not only determined by the feasibility (i.e. how representative is a cell size of 25*25 m when the source data points for interpolation are located 30 km away) but also by the representation of (quasi) details in annoted images and the requirements of external parties, for example for use in hydrological models. Based on the current user requirements, a cell size of 1*1 km may be chosen because it offers enough (quasi) detail for modelling, fits the cell size of the precipitation radar an still does not produce too big data volumes. Another possibility is to use a different cell size for the interpolation process (eg. 500 m.) and for the downloadable products (eg. 5000 m.).

The effects of smoothing should be considered during the pilot study. However, smoothing is inherent to all interpolation methods and even may be desired in some cases to reduce errors. It is important to realize that the present KNMI products based on splines are (unrealistically) smooth for several reasons, but widely accepted by the general public. When KNMI starts to deliver less smooth products that may be "a change in culture" and this should be communicated with care.

5.2 Correct use of the interpolation method

The correct use of the interpolation method means that assumptions like stationarity, intrinsic hypothesis, isotropy and being unbiased are being fulfilled. It is also important to choose the best explanatory variables (ancillary data) and to include them correctly. The discussion based on this document with KS-KA and the pilot study in phase 2 should give this information which should be documented for future use for each variable.

5.3 Test several methods

As shown in section 3 there is no single general method for all problems. Therefore different methods should be tested based on the most promising techniques (section 5.1), with different sets of ancillary data and parameterization. Testing also involves the selection of suitable processing and data-distribution software.

GMT does only support trend surfaces (splines) and nearest neighbour interpolation and is therefore not the preferred solution for this project. However, because of the powerful tools for visualization of the data it may be considered to be used in the final stage of map production.

ESRI ArcMap is very powerful for visualization and processing with user intervention. It is therefore the preferred choice for visualization (and possibly calculation) of the climate atlas and scenario products. It is not flexible enough for automatic processing: it asks a lot of system resources and is available for the Windows platform only.

The R/GSTAT combination includes all necessary functions for statistical analysis and interpolation and has proven to be robust and useful in an operational environment (INTAMAP, 2008). Especially the combination linear regression modelling/residual Kriging can be done efficiently. It is the preferred solution for automatic processing and it may be useful to provide the input data for manual processing.

MISH and PRISM could be tested in a limited pilot study. Interaction between the manual and automatic platform should also be tested. OGC web services for data distribution should be tested in a later phase, but the requirements (as far as available) should be kept in mind while choosing methods and software platforms.

5.4 Validation

At present any measure of success is lacking for the interpolated KNMI products. We should strive to provide this essential information both during the pilot study to optimize the methods and during production to indicate the quality of the products. A measure of success may be the results of cross validation and/or the Kriging variance. The limited number of automatic meteorological stations does not allow data splitting, it may be possible with the 325 voluntary precipitation measurements, but the spatial variation in precipitation is rather high for a successful application of data splitting.

5.5 Future research

At the end of the Cost-719 action the three most promising trends within meteorological interpolation were described:

1) interpolation approaches which enable regional model parameters, 2) the use of weather-type information (circulation patterns) and 3) the use of numerical weather models to provide reliable first guess fields. The regional approach is not so useful in the Netherlands because of the limited extent of the country and the limited number of automatic meteorological stations. The use of numerical weather models as ancillary data sources is a promising technique for the Netherlands. At the moment, implementation is still too complex but the advancement of data delivery systems (results of the ADAGUC project) makes it an interesting candidate for future research, especially for implementation at the weather department (10-minute scales). The use of weather-type information is linked to the numerical weather models but is easier to implement.

6 Conclusion

The following conclusions and recommendations can be drawn from this literature research. These conclusions will give an outline for the second phase of this research (August-December 2008) when different techniques will be tested and methodology and procedures will be documented in detail.

1. Research should focus on the most used and promising techniques: universal Kriging and linear regression models in combination with Kriging (residual Kriging) or IDW. When less than 30 measurements are available IDW or splines may be a better choice. The added value compared to the present KNMI technique (splines) should be tested for all methods. The use of Artificial Neural Networks (ANN) is not preferred.
2. The methods MISH and PRISM should be investigated in more detail in a (limited) pilot study.
3. Measures of success should be given for all data products by the results of cross validation and/or the Kriging variance.
4. Different methods should be tested based on the most promising techniques with different sets of ancillary data and parameterization.
5. Climate reference maps (normals) should be based on the interpolation of absolute values. Monthly or seasonal maps should be based on the interpolation of the anomalies. Daily and/or synoptic maps (10-minutes) should be based on the interpolation of absolute values.
6. The work of Schuurmans could be continued for optimizing the interpolation of precipitation by precipitation radar data.
7. The coordinate system should match the current systems used at KNMI. RijksDriehoekstelsel (RD) is a good candidate as well as the European system ERTS89.
8. Other projections can be supported through OGC services.
9. A cell size of 1*1 km or smaller is recommended during the interpolation process, individual products should be provided on a cell size of 1*1 or larger.
10. It should be realized that the present KNMI products based on splines are smooth; When KNMI starts to deliver less smooth products this should be communicated with care.
11. R/GSTAT is the preferred software for automatic processing. ESRI ArcMap is the preferred software for manual processing. GMT may be used for visualization at the web front-end.

7 References

Buishand, T.A., J.J. Beersma, R. Sluiter & T. Kroon (2008). Definitiestudie rasterdata meteorologie. De Bilt, KNMI / RWS Waterdienst, 2008

Burrough, P.A. & R.A. McDonnell (1998). Principles of Geographical Information Systems. Oxford: Oxford University Press.

De Smith, M.J., M.F. Goodchild & P.A. Longley (2007). Geospatial Analysis - A Comprehensive Guide to Principles, techniques and Software Tools. Leicester: Matador.

Dobesch, H., P. Dumolard & I. Dryas, Eds. (2007) Spatial interpolation for climate data: the use of GIS in climatology and meteorology Geographical Information Systems Series. London: ISTE ltd.

Dryas, I. & Z. Ustrnul (2007). The Spatial Analysis of the Selected Meteorological Fields in the Example of Poland. In: H. Dobesch et al (Ed.) Spatial interpolation for climate data: the use of GIS in climatology and meteorology pp. 87-96. London: ISTE Ltd.

ESRI (2008). ESRI GIS and Mapping Software. [14/11/2008] http://www.esri.com/.

EU (2008). INfrastructure for SPatial InfoRmation in Europe. [14/11/2008] http://inspire.jrc.ec.europa.eu/.

Heijboer, D. & J. Nellestijn (2002). Klimaatatlas van Nederland - De Normaalperiode 1971-2000. Rijswijk: Uitgeverij Elmar B.V.

Hutchinson, M.F. (2008). ANUSPLIN Version 4.3. [14/11/2008] http://fennerschool.anu.edu.au/publications/software/anusplin.php.

INTAMAP (2008). Interoperability and Automated Mapping. [14/11/2008] http://www.intamap.org/.

KNMI (2005). COST-719 official website. [14/11/2008] http://www.knmi.nl/samenw/cost719/.

KNMI (2008a). Geografische overzichten van het weer in Nederland. [14/11/2008] http://www.knmi.nl/klimatologie/geografische_overzichten/index.cgi.

KNMI (2008b). Weer Actueel. [14/11/2008] http://www.knmi.nl/actueel/.

Mitchell, J. (2008). Home of Ensembles-eu.org. [14/11/2008] http://www.ensembles-eu.org/.

OGC (2008). he Open Geospatial Consortium, Inc.®. [14/11/2008] http://www.opengeospatial.org/.

Perry, M. & D. Hollis (2005). The generation of monthly gridded datasets for a range of climatic variables over the UK. International Journal of Climatology 25, pp. 1041-1054.

PRISM (2008). PRISM Group. [14/11/2008] http://www.prism.oregonstate.edu/.

R-Project (2008). The R Project for Statistical Computing. [14/11/2008] http://www.r-project.org/.

Schuurmans, J.M., M.F.P. Bierkens & E.J. Pebesma (2007). Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar. Journal of Hydrometeorology 8, pp. 1204 - 1224.

Schwarb, M. (2001). The Alpine Precipitation Climate. Evaluation of a High-Resolution Analysis Scheme Using Comprehensive Rain-Gauge Data. Zuercher Klimaschriften 80,

Szentimrey, T., Z. Bihari & S. Szalai (2007). Comparison of Geostatistical and Meteorological Interpolation Methods (What is What?). In: H. Dobesch et al (Ed.) Spatial interpolation for climate data: the use of GIS in climatology and meteorology pp. 45-56. London: ISTE Ltd.

Tveito, O.E. (2007). The Developments in Spatialization of Meteorological and Climatological Elements In: H. Dobesch et al (Ed.) Spatial interpolation for climate data: the use of GIS in climatology and meteorology pp. 73-86. London: ISTE Ltd.
Tveito, O.E., M. Wegehenkel, F. Van der Wel & H. Dobesch (2006). The Use of Geographic Information Systems in Climatology and Meteorology - Final Report COST Action 719.
Wessel, P. (2008). The Generic Mapping Tools.] http://gmt.soest.hawaii.edu/.