Creation of climatic maps for Trentino from spread data

Chiara SBOARINA (*)

(*) Centro di Ecologia Alpina, Viote del Monte Bondone (TN), 0461 939555, sboarina@cealp.it

Abstract

The aim of this study is to create climatic maps for one particular region, Trentino in the north-east of Italy. The availability of a great number of measurement stations all over the region and the use of a geographic information system are the characteristics of this work. Interpolating values of climate variables from points to large areas is important in a variety of disciplines, therefore it is fundamental to minimize the extent of interpolation errors by using a suitable interpolation method. We compared the performance of three interpolation methods: Inverse Distance Weighting (IDW), Regularized Spline with Tension (RST) and Ordinary Kriging. The first two procedures are integrated in GIS GRASS while for Kriging it has been taken advantage of an external program named Gstat. Before interpolating it was found that there is no relationship between precipitation and elevation or coordinates of measure stations, while as to temperature it was found the known relationship with elevation. Annual average rain data were interpolated using the three different methods and the results were compared on the base of statistics and cross validation. Monthly average temperature data for all year months were interpolated after being taken back to sea level. The choice of the best interpolation method for temperature was made for December and August data because they are the coldest and the warmest month of the year. The final result is a map of annual precipitation and twelve maps, one for every month, of temperature for all Trentino.

Introduction

Estimates of meteorological values such as temperature and precipitation rate are required for a number of landscape scale models, including those of regeneration, growth, and mortality of forest ecosystems. Interpolating values of climate variables from measurement stations to large areas is therefore fundamental and requires to minimize the extent of interpolation errors by using a suitable interpolation method. Given a set of meteorological data, it's possible to use a variety of stochastic and deterministic interpolation methods to estimate meteorological variables at unsampled locations. The choice of spatial interpolator is especially important in mountainous regions, like Trentino, where data collection are sparse and variables may change over short spatial scales. In the specific case of this work a great number of measurement stations all over the region provides an enormous quantity of precipitation and temperature data. The best way to work with this georeferred data is to use a Geographic Information System because it allows to manage a lot of data quickly and without loosing their geographic information. This research is therefore concerned with the spatial interpolation of meteorological data to obtain climatic maps as layers in a GIS.

The climatic data

The climatic data used in this research are from the Hydrographic Bureau of Autonomous Province of Trento and from the Istituto Agrario of S.Michele all'Adige. The data cover a period of ten years from 1990 to 1999 but some stations present holes in data series or few years of data. The daily temperature data were provided by the Province Bureau in text file of random length, so it was necessary to patch all files for each station into one file only. During this process of patching, the data were analyzed and many errors were adjusted. There were a lot of repeated days and data for the 29th of February of no-leap years. The daily precipitation data were provided in table format, one for each year and station, with monthly and annual mean value. The data of the Istituto Agrario of S.Michele were downloaded from the internet into text tables one for each year. The aim of this work is to produce climatic maps for Trentino and it was chosen to create an annual precipitation map and twelve monthly temperature maps, one for each month. The average annual precipitation was quickly calculated in the period of ten years for all the stations. The temperature data preparation required such a lot of time because of the data holes. These prevented the automatic calculation of monthly means. In cases where some days were missing but less of 10 for that given month, the mean was calculated from the available daily data. This data averaging was necessary to ensure an adequate number of stations for the 10 years mean. Even with this process, same stations within the test region were dropped from the analysis due to data fewness. Then in the area remained 75 measurement stations for precipitation and 64 for temperature. In fig. 1 are shown the locations of used measurement stations.

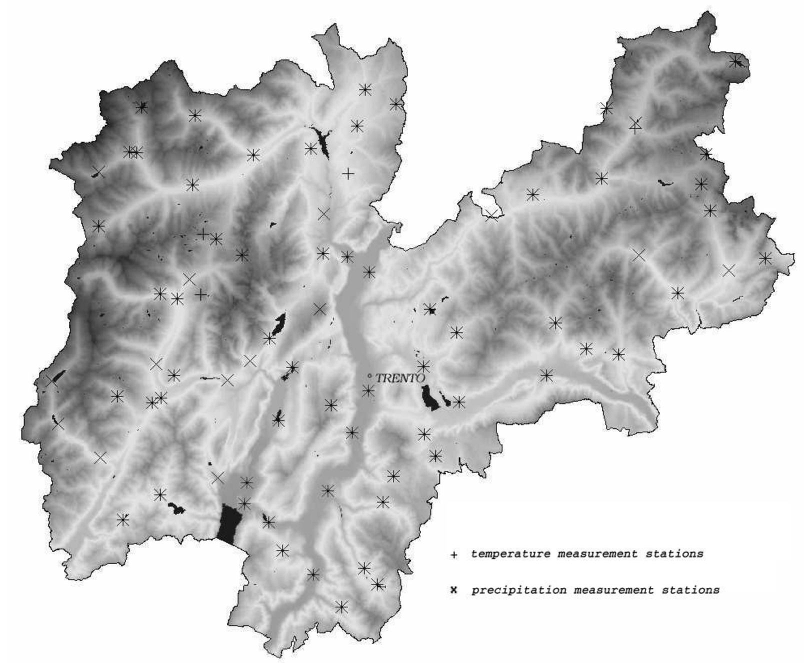

Figure1- Measurement stations

Since the area of Trentino is of 6208 Km^2 each station represents an average area of about 90 km^2. Comparing this value with other analogue works (Hargrove 1995, Hutchinson et al. 1996, Linch 1998) was chosen a grid resolution of 100 meters, so every cell represents a hectare. For topographically dependent representations of temperature and rainfall surface would be more appropriate a mesoscala with spatial resolutions from 200 meters to 5 kilometers. But since a fine mesoscala with spatial resolutions from 50 to 200 meters, is well suited to modelling aspect related to microclimatic variations, particularly in solar radiation, evaporation and associated vegetation patterns (Hutchinson, 1998), the choice of a fine mesoscala can be appropriate for the future use of the climatic maps. Before beginning with interpolation, it was analyzed the relationship between temperature and elevation and between precipitation and elevation or stations coordinates. It was found the known linear relationship between temperature and elevation. The angolar coefficient of the line varies from -0.0033 to -0.006 °C/m, it is smaller for winter months and bigger for summer months. Precipitation doesn't present any relationship with measurement stations coordinates and a light relationship with the elevation: higher values at medium elevation.

GIS and interpolation

The best way to work with georeferred data, like the climatic data from measurement stations, is to use a Geographic Information System because it allows to manage a lot of data quickly and without loosing their geographic information. In particular the GIS used for this work is GRASS (Geographic Resource Analysis Support System). It has the great advantage to be free and to manage very well raster format. As said before to interpolate climatic data is very important to consider elevation. This is very simple using a GIS like GRASS having a Digital Terrain Model available. The DTM was provided by the Autonomous Province of Trento and has a resolution of 10 meters. The interpolation of temperature is a problem with two independent position variables (north and east coordinates) and a single linear dependence on elevation. The assumption of this work is to separate the process of interpolation in two parts. Temperatures were taken back to see level before interpolating using the known lapse rate and the DTM. The result of interpolation was then taken to originally elevation doing the inverse path. An advantage of this process is that the coefficient of the linear sub-model (the lapse rate) is automatically determined from the data, thus it hasn't to be specified beforehand. About precipitation the interpolation was done without considering elevation since no strong relationship was found.

Three methods of interpolation were used in this research: inverse distance weighting, regularized spline with tension and ordinary kriging. The first two are integrated in GIS GRASS while the third required the use of an external program named Gstat. This is a free program for the modelling, prediction and simulation of geostatistical data in one, two and three dimensions.

Inverse Distance Weighting

Inverse distance weighting interpolation is a deterministic estimation method where values at unsampled points are determined by a linear combination of values at known sampled points. The degree of influence, or the weight, is expressed by the inverse of the distance between points raised to a power. A power of 1.0 means a constant rate of change in value between points, and the method is called linear interpolation. A power of 2.0 or higher suggests that the rate of change in values is higher near a known point and levels off away from it. The GIS GRASS includes this technique of interpolation in a module named s.surf.idw that fills a raster matrix with interpolated values generated from a set of irregularly spaced data points using a power of 2.0 for distance. The default number of points considered in interpolation is 12.

Regularized Spline with Tension

Regularized smoothing spline with tension is a radial basis function method for interpolation from scattered data. The interpolation is flexible through the choice of a tension parameter which controls the properties of the interpolation function and a smoothing parameter which enables to filter out the noise. The function has regular derivatives of arbitrary order and can be used for interpolation in arbitrary dimension. The spline mathematical functions are akin to the flexible ruler that was historically used to produce a smooth curve when joining a set of points. Splines have the advantage of creating curves and contour lines which are visually appealing. The module of GRASS GIS that interpolates with regularized spline with tension allows to set many parameters. The tension parameter tunes the character of the resulting surface from thin plate to membrane. For noisy data, it is possible to define either a constant or a variable smoothing parameter. In this work the tension and the smooth parameters were set to the default values while the minimum number of points was set to 3 and the distance below which two points are too close was set at 500 meters.

Kriging

Kriging is a stochastic technique similar to inverse distance weighted averaging in that it uses a linear combination of weights at known points to estimate the value at an unknown point. Kriging uses a semivariogram, a measure of spatial correlation between two points, so the weights change according to the spatial arrangement of the samples. Unlike other estimation procedures investigated, kriging provides a measure of the error or uncertainty of the estimated surface. Kriging requires a two-step process: the fitting of a semivariogram model function (of distance) followed by the solution of a set of matrix equations. When the number of data points is large this technique is computationally very intensive and the estimation of the variogram is not simple. Gstat, the program used to do kriging, allows to calculate the semivariogram from punctual data setting cutoff and lag; in this case cutoff was set to 100000 meters while lag to 7500 meters. The resulting semivariogram must be fitted to a curve that for temperature resulted to be exponential and for precipitation to be the combination of an exponential and a gaussian. In fig. 2 is shown the fitted semivariogram for precipitation. Known the semivariogram equation the program can proceed to interpolation via ordinary kriging. Same parameters are needed such as radius of interpolation, set to 100000, and the minimum and maximum number of points to be used, respectively set to 2 and 10.

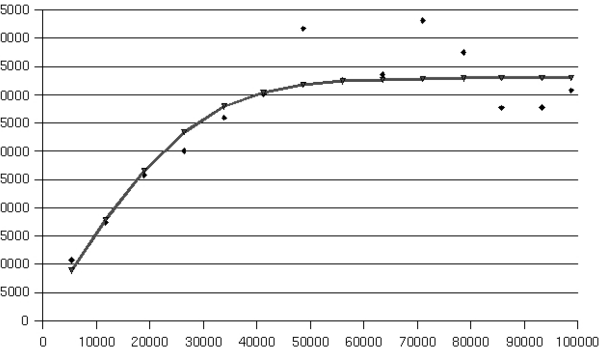

Figure 2 - Fitted semivariogram for precipitation

Cross Validation and interpretation of results

The different point estimation methods were compared on the basis of bias, mean absolute error (MAE), and mean squared error (MSE). As the true temperature and precipitation surfaces were not known, the comparison statistics were obtained using the cross validation technique. This allows to compare estimated and true values only using the information available in the known data set. In cross validation the estimation method is tested at the locations of existing samples. The value at a particular location was temporarily discarded from the sample data set, the value at the same location was then estimated using the remaining samples. Once the estimate was calculated it was compared to the true sample value that was initially removed. This procedure was repeated for all available samples. The resulting true and estimated values were compared using summary statistics like mean, standard deviation and so on.

Precipitation

In the case of precipitation the cross validation was done for all measurement station without one that was too near to region border. In figure 3 there are the statistics of bias in millimeters. It can be seen that errors mean is considerably lower in case of regularized spline with tension than the other cases. One can objects that this mean is not reliable because error values are positive and negative. But looking the mean absolute error and the mean squared error it can be seen that the lowest values remain those of RST.

	IDW	RST	Kriging
m	-7.7	-3.35	-6.13
σ	132.15	129.89	131.73
IQR	156.71	146.68	162.29
MAE	99.76	97.51	100.67
MSE	17286.92	16654.14	17157.14

Figure 3 - Statistics for errors of precipitation

The result of interpolation is the precipitation map in millimeters shown in figure 4, where values of rainfall increase from white to black. In this map also the isohytes of precipitation are represented in black.

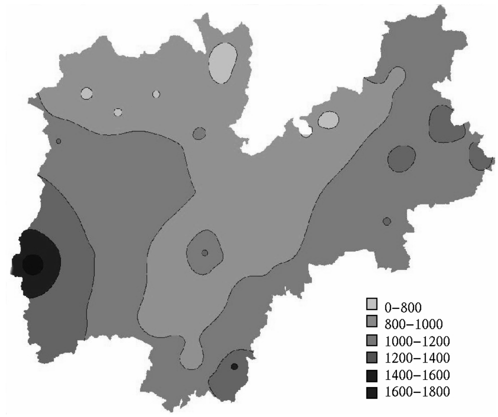

Figure 4 - Precipitation map with isohyets in black

Temperature

Because the technique of cross validation requires a lot of time in the case of temperature it was made only for two months: the coldest and the warmest. For Trentino the coldest month is December while the warmest is August. In figure 5 are shown statistics in °C for August while in figure 6 statistics for December.

	IDW	RST	Kriging
m	0	-0.01	0.03
σ	0.82	0.82	0.84
IQR	1.17	1.17	1.22
MAE	0.66	0.67	0.68
MSE	0.65	0.67	0.69

Figure 5 - Statistics for errors of August mean temperature

	IDW	RST	Kriging
m	0.51	0.46	0.48
σ	1.23	1.22	1.23
IQR	1.42	1.2	1.24
MAE	1.01	0.98	1
MSE	1.75	1.68	1.73

Figure 6 - Statistics for errors of December mean temperature

The results are a little different for the two months in fact for August the best method is IDW while for December is RST. Given the minimum difference between the MAE and MSE of IDW and RST for August and the lower values of RST for December the regularized spline with tension was chosen as interpolation method for all months. The final result in the case of temperature are twelve maps of mean monthly temperature. In figure 7 is reported the map of April where temperature increases from white to black.

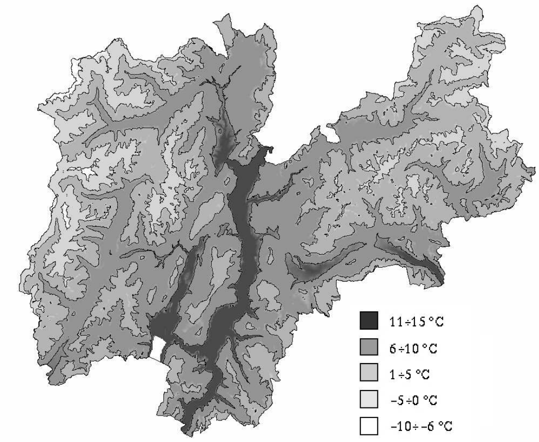

Figure 7 - Temperature map for April

Conclusions

The regularized spline with tension provided a simple and rapid interpolation method for temperature and precipitation. By incorporating a parametric sub-model for a linear dependence of temperature on elevation, regularized spline with tension offers an objective flexible tool and an elegant and simple solution (Kesteven and Hutchinson 1996). On the contrary the kriging method involves an enormous amount of work especially in semivariogram estimation and does not produce better results.

Error values for precipitation obtained from cross validation are 10% circa of the mean annual value. This results is good in comparison with the literature values shown in figure 8. Errors for temperature are respectively 0.67 °C and 0.98 °C for the mean value of the months of August and December. Error is larger for coldest month probably due to the process of thermic inversion that causes a change in the temperature elevation lapse rate. Temperature errors can also be compared with literature values. These errors are slightly smaller than the ones computed in this work but there is a considerable difference between the resolution of grid in literature and the resolution of grid used here. The dimension of cell adopted for the climatic maps of Trentino is 100 times smaller than the one of other works.

Authors	Tmin	Tmax	Precipitation	Resolution
Hutchinson et al. 1996	0.5-0.8 °C	0.5 °C	10-30%	5 km
Linch 1999	-	-	20%	1.6 km
Hutchinson 1998	0.3-0.5 °C	0.2-0.4 °C	5-15%	1 km?
Thornton et al. 1997	1.2 °C	0.7 °C	19.3 %	500 m - 32 km

Figure 8 - Error values of different literature works

In conclusions it can be said that regularized spline with tension is a valid tool to interpolate both precipitation and temperature in a reliable and quick way.

References

Collins, F.C. (1996) - *A comparison of spatial interpolation techniques in temperature estimation.* - Proc. Third International Conference/Workshop on Integrating GIS and Environmental Modeling CD-ROM. Santa Fe, New Mexico, USA. January 21-25.

GRASS Development Team - *GRASS 5.x Reference Manual.*

Hargrove, W.W. (1995) - *Interpolation of rainfall in Switzerland using a regularized spline with tension.* www.research.esd.ornl.gov/~hnw/sic97

Hutchinson, M; Nix, H.A; McMahon, J.P; Ord, K.D. (1996) - *The development of a topographic and climate database for Africa.* Proc. Third International Conference/Workshop on Integrating GIS and Environmental Modeling CD-ROM. Santa Fe, New Mexico, USA. January 21-25.

Hutchinson,M. (1998) - *Modelling Spatial and Temporal Variability of Climate and Terrain.*

Isaaks, E.H. Srivastava, R.M. (1989) - *An Introduction to Applied Geostatistics*. Oxford University Press.

Kesteven, J. Hutchinson,M. (1996) - *Spatial Modelling of climatic variables on a continental scale.* Proc. Third International Conference/Workshop on Integrating GIS and Environmental Modeling CD-ROM. Santa Fe, New Mexico, USA. January 21-25.

Lynch, S.D. (1999) - *Converting point estimates of daily rainfall onto a rectangular grid.* ESRI User Conference ’98. San Diego Convention Center, July 27-31 1998.

Linch, S.D.; Schulze, R.E. (199?) - *Techniques for estimating areal daily rainfall*. Proc. Seventh Hydrological Sciences Symposium, Rhodes University, Grahamstown, South Africa.

Pebesma, E.J. - *Gstat user's manual*. - Dept. of Physical Geography, Utrecht University.

Thornton, P.E.; Running, S.W.; White, M. A. (1997) - *Generating surfaces of daily meteorological variables over large regions of complex terrain*. Journal of Hydrology, *190*(3-4), 214-251.