Weather Forecasting with Scilab

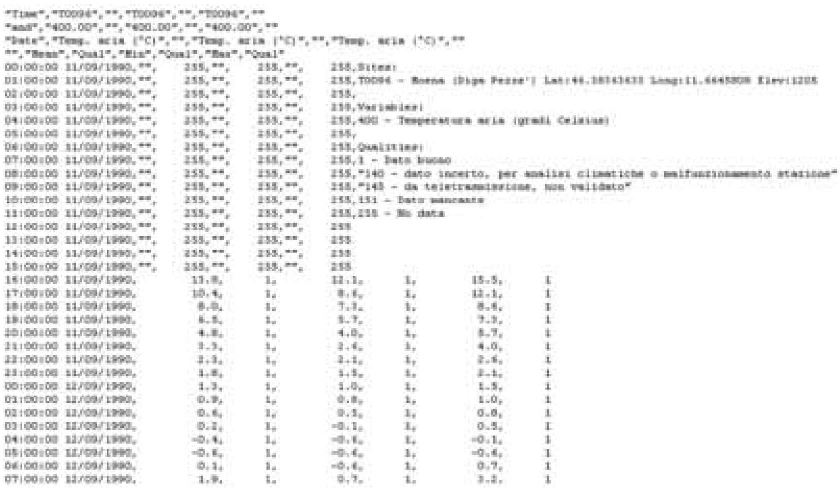
The weather is probably one of the most discussed topics all around the world. People are always interested in weather forecasts, and our life is strongly influenced by the weather conditions. Let us just think of the farmer and his harvest or of the happy family who wants to spend a weekend on the beach, and we understand that there could be thousands of good reasons to be interested in knowing the weather conditions in advance.
This probably explains why, normally, the weather forecast is the most awaited moment by the public in a television newscast.
Sun, wind, rain, temperature… the weather seems to be unpredictable, especially when we consider extreme events. Man has always tried to develop techniques to master this topic, but practically only after the Second World War the scientific approach together with the advent of the media have allowed a large diffusion of reliable weather forecasts.

To succeed in forecasting, it is mandatory to have a collection of measurements of the most important physical indicators which can be used to define the weather in some relevant points of a region at different times. Then, we certainly need a reliable mathematical model which is able to predict the values of the weather indicators at points and times where no direct measurements are available.
Nowadays, very sophisticated models are used to forecast the weather conditions, based on registered measurements such as the temperature, the atmospheric pressure, the air humidity as so forth.
It is quite obvious that the larger the dataset of measurements the better the prediction: this is the reason why the institutions involved in monitoring and forecasting the weather usually have a large number of stations spread on the terrain, opportunely positioned to capture relevant information.

This is the case of Meteo Trentino (see [3]), which manages a network of measurement stations in Trentino region and provides daily weather forecasts.
Among the large amount of interesting information we can find in their website, there are the temperature maps, where the predicted temperature at the terrain level for the Trentino province is reported for a chosen instant. These maps are based on a set of measurements available from the stations: an algorithm is able to predict the temperature field in all the points within the region and, therefore, to plot a temperature map.
We do not know the algorithm that Meteo Trentino uses to build these maps, but we would like to set up our own procedure able to obtain similar results. To this aim, we decided to use Scilab (see [1]) as a platform to develop such a predictive model and gmsh (see [2]) as a tool to display the results.
Probably one of the most popular algorithms in the geo-sciences domain used to interpolate data is Kriging (see [5]). This algorithm has the notable advantage of exactly interpolating known data; it is also able to potentially capture non-linear responses and, finally, to provide an estimation of the prediction error. This valuable last feature could be used, for example, to choose in an optimal way the position of new measurement stations on the terrain.
Scilab has an external toolbox available through ATOMS, named DACE (which stands for Design and Analysis of Computer Experiments), which implements the Kriging algorithm. This obviously allows us to implement more rapidly our procedure because we can use the toolbox as a sort of black-box, avoiding in this way spending time implementing a non-trivial algorithm.

The weather data

We decided to download from [3] all the available temperatures reported by the measurement stations. As a result we have 102 formatted text files (an example is given in Figure 1) containing the maximum, the minimum and the mean temperature with a timestep of one hour.
In our work we only consider the "good" values of the



```
"Time","T0094","","T0094","","T0094",""
"and","400.00","","400.00","","400.00",""
"Date","Temp. aria (°C)","","Temp. aria (°C)","","Temp. aria (°C)",""
"","Mean","Qual","Min","Qual","Max","Qual"
00:00:00 11/09/1990,"",     255,"",     255,"",     255,Sites:
01:00:00 11/09/1990,"",     255,"",     255,"",     255,T0094 - Moena (Diga Pezze') Lat:46.38163633 Long:11.6645808 Elev:1205
02:00:00 11/09/1990,"",     255,"",     255,"",     255,
03:00:00 11/09/1990,"",     255,"",     255,"",     255,Variables:
04:00:00 11/09/1990,"",     255,"",     255,"",     255,400 - Temperatura aria (gradi Celsius)
05:00:00 11/09/1990,"",     255,"",     255,"",     255,
06:00:00 11/09/1990,"",     255,"",     255,"",     255,Qualities:
07:00:00 11/09/1990,"",     255,"",     255,"",     255,1 - Dato buono
08:00:00 11/09/1990,"",     255,"",     255,"",     255,"140 - dato incerto, per analisi climatiche o malfunzionamento stazione"
09:00:00 11/09/1990,"",     255,"",     255,"",     255,"145 - da teletrasmissione, non validato"
10:00:00 11/09/1990,"",     255,"",     255,"",     255,151 - Dato mancante
11:00:00 11/09/1990,"",     255,"",     255,"",     255,255 - No data
12:00:00 11/09/1990,"",     255,"",     255,"",     255
13:00:00 11/09/1990,"",     255,"",     255,"",     255
14:00:00 11/09/1990,"",     255,"",     255,"",     255
15:00:00 11/09/1990,"",     255,"",     255,"",     255
16:00:00 11/09/1990,       13.8,       1,      12.1,      1,      15.5,      1
17:00:00 11/09/1990,       10.4,       1,       8.6,      1,      12.1,      1
18:00:00 11/09/1990,        8.0,       1,       7.3,      1,       8.6,      1
19:00:00 11/09/1990,        6.5,       1,       5.7,      1,       7.3,      1
20:00:00 11/09/1990,        4.8,       1,       4.0,      1,       5.7,      1
21:00:00 11/09/1990,        3.3,       1,       2.4,      1,       4.0,      1
22:00:00 11/09/1990,        2.3,       1,       2.1,      1,       2.6,      1
23:00:00 11/09/1990,        1.8,       1,       1.5,      1,       2.1,      1
00:00:00 12/09/1990,        1.3,       1,       1.0,      1,       1.5,      1
01:00:00 12/09/1990,        0.9,       1,       0.8,      1,       1.0,      1
02:00:00 12/09/1990,        0.6,       1,       0.5,      1,       0.8,      1
03:00:00 12/09/1990,        0.2,       1,      -0.1,      1,       0.5,      1
04:00:00 12/09/1990,       -0.4,       1,      -0.6,      1,      -0.1,      1
05:00:00 12/09/1990,       -0.6,       1,      -0.4,      1,      -0.6,      1
06:00:00 12/09/1990,        0.1,       1,      -0.6,      1,       0.7,      1
07:00:00 12/09/1990,        1.9,       1,       0.7,      1,       3.2,      1
```

Fig. 1 - The hourly temperature measures for the Moena station: the mean, the minimum and the maximum values are reported together with the quality of the measure.

mean temperature: there is actually an additional column which contains the quality of the reported measure which could be "good", "uncertain", "not validated" and "missing".

The terrain data

Another important piece of information we need is the "orography" of the region under exam. In other words we need to have a set of triplets giving the latitude, the longitude and the elevation of the terrain. This last information is mandatory to build a temperature map at the terrain level.

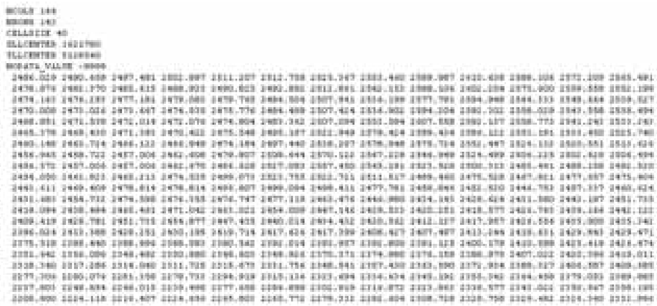


Fig. 2 - An example of the DTM file formatted to the ESRI standard. The matrix contains the elevation of a grid of points whose position is given with reference to the Gauss Boaga Roma 40 system.

To this aim we downloaded the DTM (Digital Terrain Model) files available in [4] which, summed all together, contain a very fine grid of points (with a 40 meters step both in latitude and longitude) of the Trentino province. These files are formatted according to the ESRI standard and they refer to the Gauss Boaga Roma 40 system.

Set up the procedure and the DACE toolbox

We decided to translate all the terrain information to the UTM WGS84 system in order to have a unique reference for our data. This operation can be done just once and the results stored in a new dataset to speed up the following computations.
Then we have to extract from the temperature files the available data for a given instant, chosen by the user, and store them. With these data we are able to build a Kriging model, thanks to the DACE toolbox. Once the model is available, we can ask for the temperature at all the points belonging to the terrain grid defined in the DTM files and plot the obtained results.
One interesting feature of the Kriging algorithm is that it is able to provide an expected deviation from the prediction. This means that we can have an idea of the degree to which our prediction is reliable and eventually estimate a possible range

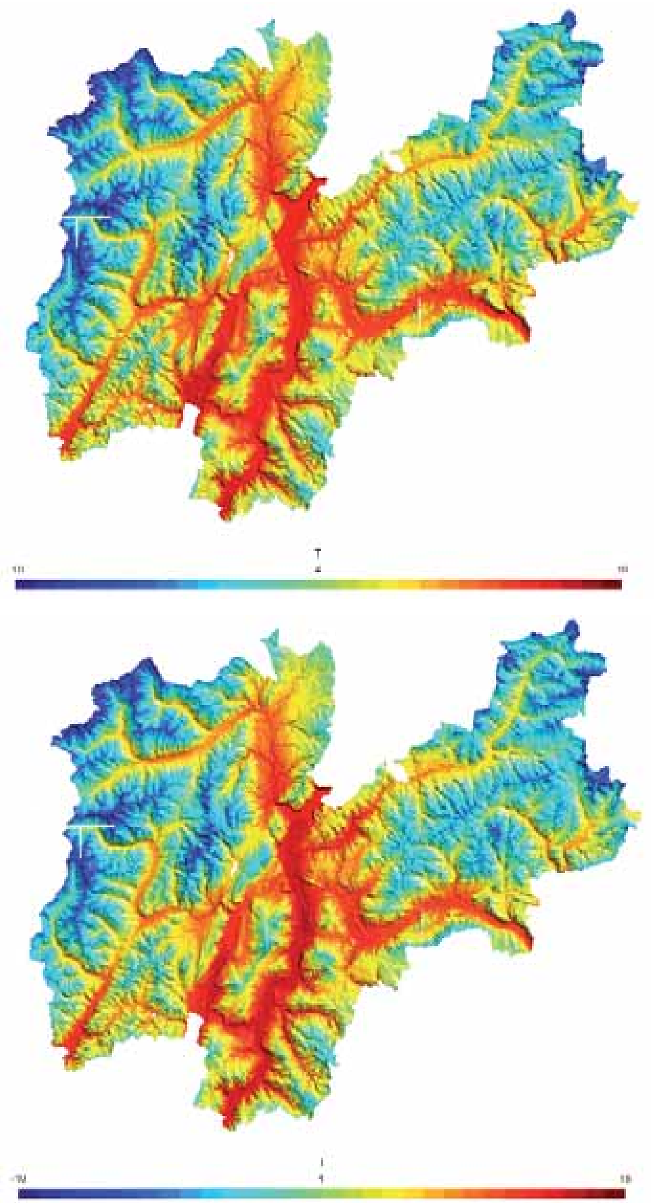


Fig. 4 - 6th May 2010 at 17:00. Top: the predicted temperature at the terrain level using Kriging is plotted. The temperature follows very closely the height on the sea level. Bottom: the temperature map predicted using a linear model relating the temperature to the height. At a first glance these plots could appear exactly equal: this is not exact, actually slight differences are present especially in the valleys.

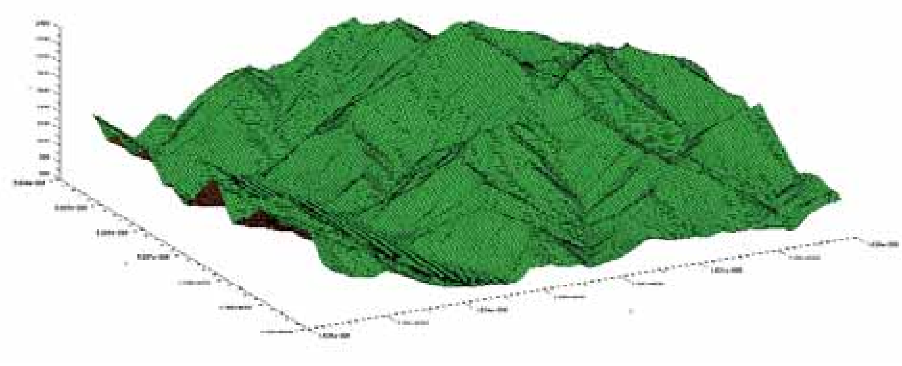


Fig. 3 - The information contained into one DTM file is graphically rendered. As a results we obtain a plot of the terrain.

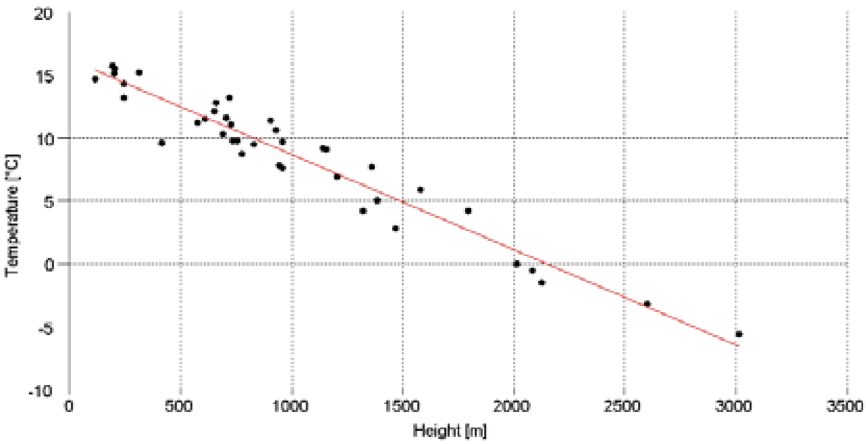


Fig. 5 - 6th May 2010 at 17:00. The measured temperatures are plotted versus the height on the sea level. The linear regression line, plotted in red, seems to be a good approximation: the temperature decreases 0.756 [°C] every 100 [m] of height.

of variation: this is quite interesting when forecasting an environmental temperature.

Some results

We chose two different days of 2010 (the 6th of May, 17:00 and the 20th of January, 08:00) and ran our procedure to build the temperature maps.

In Figure 5 the measured temperatures at the 6th of May are plotted versus the height on the sea level of the stations. It can be seen that a linear model can be considered as a good model to fit the data. We can conclude that the temperature decreases linearly with the height of around 0.756 [°C] every 100 [m]. For this reason one could be tempted to use such model to predict the temperature at the terrain level: the result of this prediction, which is reported in Figure 4, is as accurate as the linear model is appropriate to capture the relation between the temperature and the height. If we compare the results obtained with Kriging and this last approach some differences appear, especially down in the valleys: the Kriging model seems to give more detailed results.

If we consider January 20th, the temperature can no longer be computed as a function of only the terrain height. It immediately appears, looking at Figure 8, that there are large deviations from a pure linear correlation between the temperature and the height. The Kriging model, whose result is drawn in Figure 7, is able to capture also local

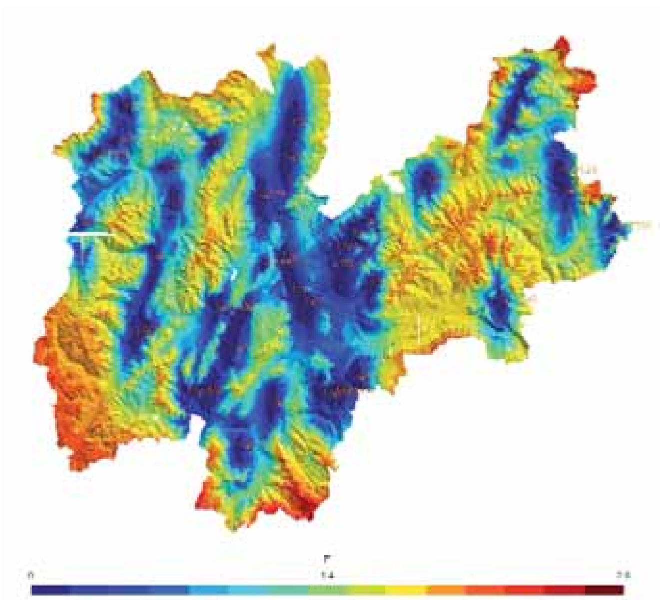


Fig. 6 - 6th May 2010 at 17:00. The estimated error in predicting the temperature field with Kriging is plotted. The measurement stations are reported on the map with a code number: it can be seen that the smallest errors are registered close to the 39 stations while, as expected, the highest estimated errors are typical of zones where no measure is available.

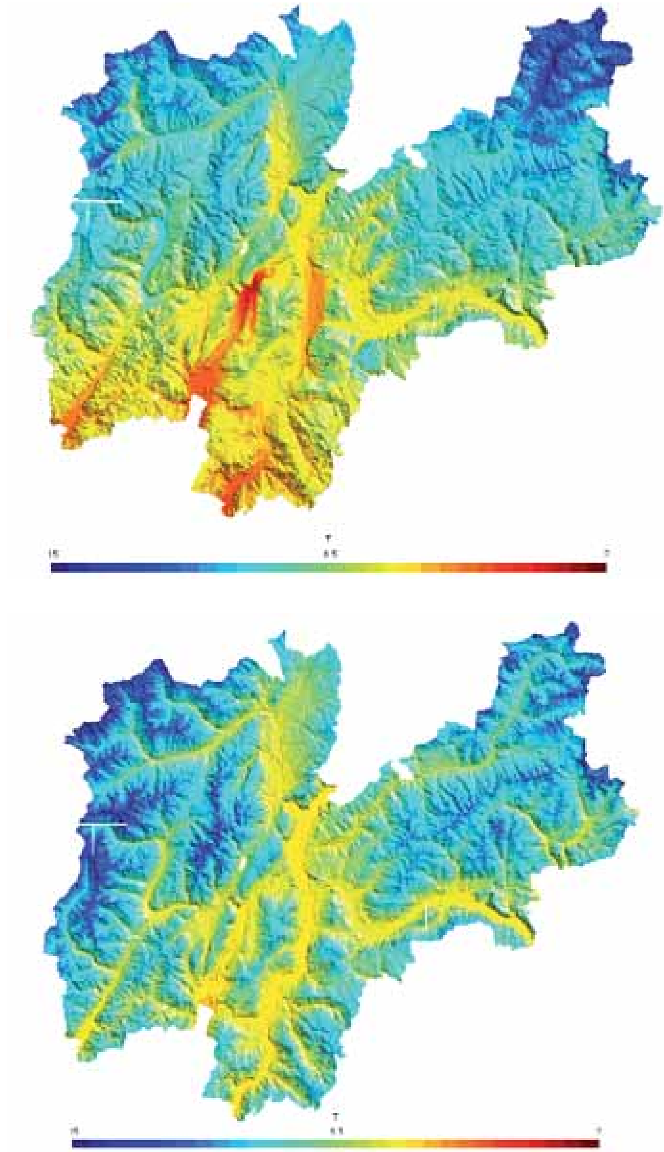


Fig. 7 - 20th January 2010 at 08:00. Top: the predicted temperature with the Kriging model at the terrain level is plotted. Globally, the temperature still follows the height on the sea level but locally this trend is not respected. Bottom: the temperature map predicted using a linear model relating the temperature to the height.

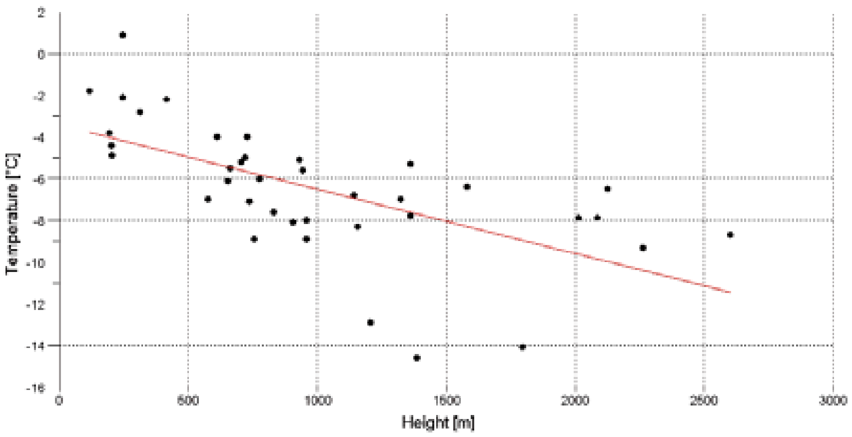


Fig. 8 - 20th January 2010 at 08:00. The measured temperatures are plotted versus the height on the sea level. The linear regression line, plotted in red, says that the temperature decreases 0.309 [°C] every 100 [m] of height but it seems not to be a good approximation in this case; there are actually very large deviations from the linear trend.

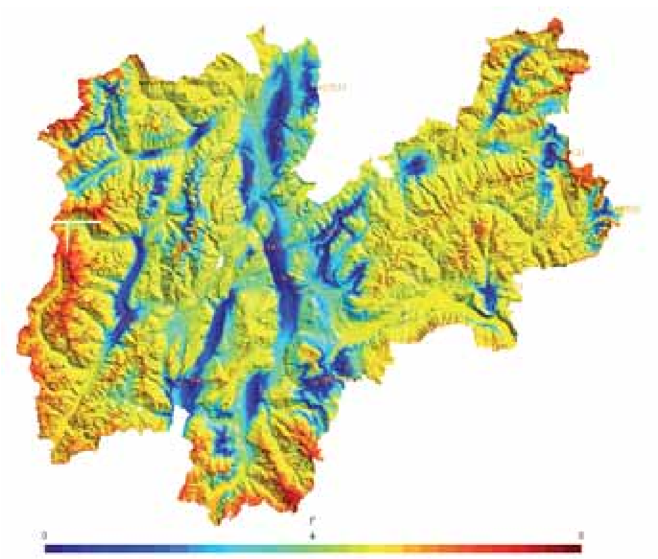


Fig. 9 - 20th January 2010 at 08:00. The estimated error in predicting the temperature field with the Kriging technique is plotted. The measurement stations are reported on the map with a code number: it can be seen that the smallest errors are registered close to the 38 stations.

positive or negative peaks in the temperature field, which cannot be predicted otherwise.

In this case, however, it can be seen that the estimated error (Figure 9) is larger than the one obtained for 17th of May (Figure 6): this lets us imagine that the temperature is in this case much more difficult to capture correctly.

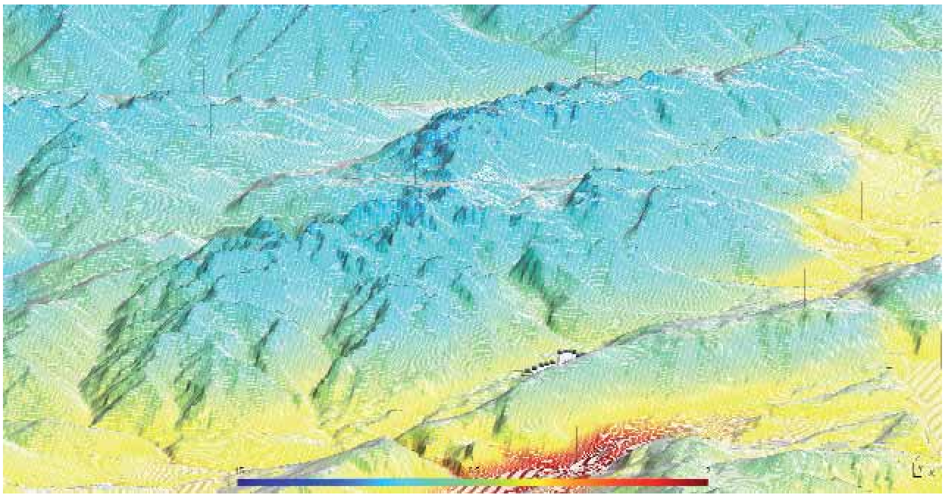


Fig. 10 - 20th January 2010 at 08:00. The estimated temperature using Kriging: a detail of the Gruppo Brenta. The black vertical bars reports the positions of the meteorological stations.

Conclusions

In this work it has been shown how to use Scilab and its DACE toolbox to forecast the temperature field starting from a set of measurements and from information regarding the terrain of the region.

We have shown that the Kriging algorithm can be used to get an estimated value and an expected variation around it: this is a very interesting feature which can be used to give a reliability indication of the prevision.

This approach could be used also with other atmospheric indicators, such as the air pressure, the humidity and so forth.

References

[1] http://www.scilab.org/ to have more information on Scilab.

[2] The Gmsh can be freely downloaded from: http://www.geuz.org/gmsh/

[3] The official website of Meteo Trentino is http://www.meteotrentino.it from where the temperature data used in this work has been downloaded.

[4] The DTM files have been downloaded from the official website of Servizio Urbanistica e Tutela del Paesaggio http://www.urbanistica.provincia.tn.it/sez_siat/siat_urbanistica/pagina83.html.

[5] Søren N. Lophaven, Hans Bruun Nielsen, Jacob Søndergaard, DACE A Matlab Kriging Toolbox, download from http://www2.imm.dtu.dk/~hbn/dace/dace.pdf.

Contacts

For more information on this document please contact the author:

Massimiliano Margonari - Openeering
m.margonari@openeering.com