WEATHER FORECASTING WITH SCILAB

Authors: *Massimiliano Margonari*

Keywords. Scilab; Experimental data; Metamodeling

Abstract: In this paper Scilab and its DACE toolbox are used to shown how it is possible to forecast the temperature field starting from a set of measurements and from information regarding the terrain of the region.

Contacts m.margonari@openeering.com

1. Introduction

The weather is probably one of the most discussed topic 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 to the happy family who wants to spend a week-end 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 forecasts is the most awaited moment by the public of a television news.
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 measures 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 in points an times where no direct measures are available.
Nowadays, very sophisticated models are used to forecast the weather conditions, based on registered measures, such as the temperature, the atmospheric pressure, the air humidity as so forth.
It is quite obvious that the largest the dataset of the measures 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 also the temperature maps, such the one reported in Figure 1, where the predicted temperature at the terrain level for the Trentino is reported for a chosen instant. These maps are based on a set of measures available from the stations: an algorithm is able to predict the temperature field in all the point 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 predictive model and gmsh (see [2]) as a tool to display the results.
Probably one of the most popular algorithm in the geo-sciences domain used to interpolate data is Kriging (see [5]). This algorithm has the notable advantage to exactly interpolate known data; it is also able to potentially capture non-linear responses and, finally, to provide an estimation of the prediction error. This last valuable 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 faster implement our procedure because we can use the toolbox as a sort of black-box, avoiding in this way to spend time implementing a non-trivial algorithm.

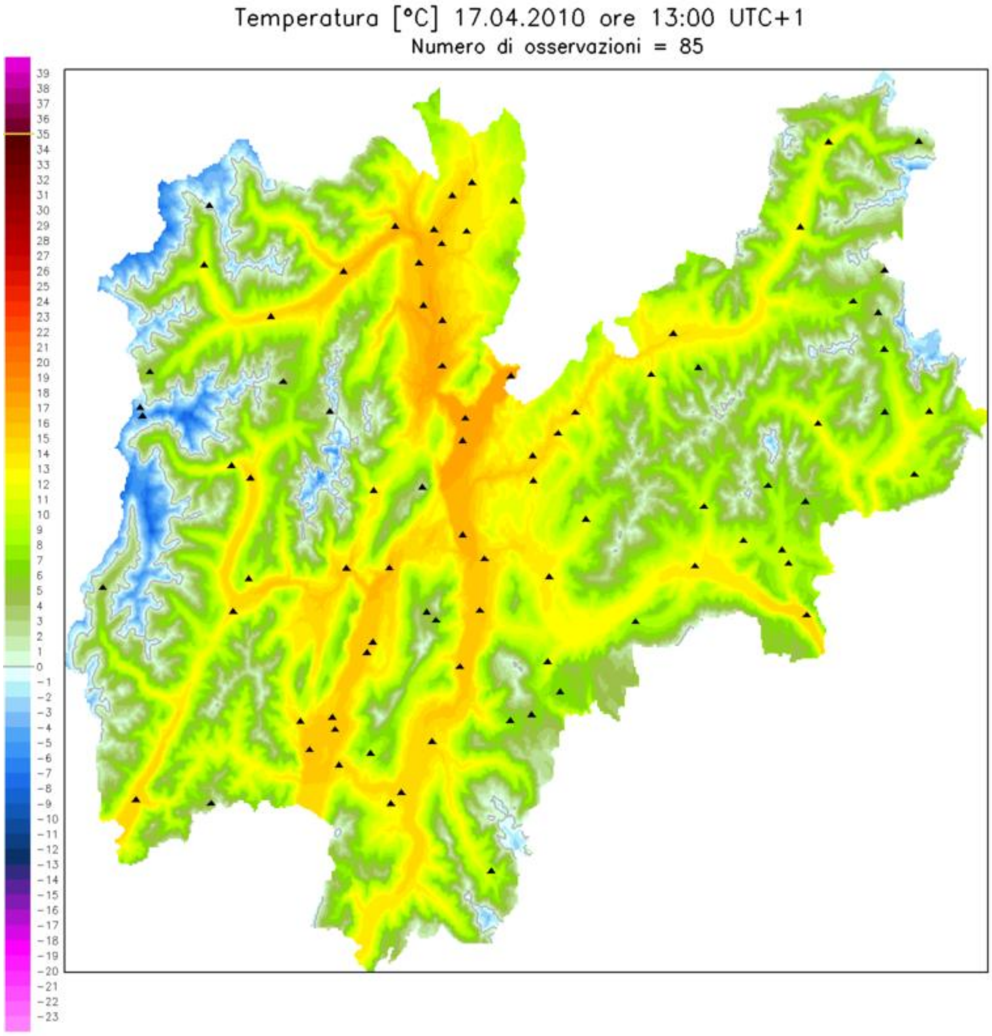

Figure 1: The predicted temperature field in Trentino by Meteo Trentino (see [3]). The black triangles indicate the stations which provide temperature measures at the 17th of April 2010, 13:00.

2. 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 2) containing the maximum, the minimum and the mean temperature with a step of an hour. In our work we only consider the “good” values of the 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”.

```
"Time","T0096","","T0096","","T0096",""
"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,T0096 - Moena (Diga Pezze') Lat:46.38363633 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,        2.6,      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,       -0.6,      1,       -0.6,      1
06:00:00 12/09/1990,         0.1,     1,       -0.6,      1,        0.7,      1
```

Figure 2: 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.

3. The terrain data

Another important information we need is certainly 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.

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 refers to the Gauss Boaga Roma 40 system.

```
NCOLS 164
NROWS 143
CELLSIZE 40
XLLCENTER 1621780
YLLCENTER 5128540
NODATA_VALUE -9999
 2486.029 2490.659 2497.481 2502.897 2511.207 2512.758 2525.367 2553.460 2589.987 2610.638 2588.106 2572.209 2565.491
 2478.876 2482.370 2485.615 2488.923 2490.825 2492.892 2512.861 2542.153 2588.106 2602.204 2575.600 2559.558 2552.199
 2474.163 2476.293 2477.181 2479.083 2479.765 2486.504 2507.941 2536.199 2577.791 2594.948 2564.333 2548.664 2539.527
 2470.008 2473.026 2473.667 2474.535 2475.776 2484.499 2507.424 2536.902 2594.204 2592.302 2558.029 2543.558 2538.494
 2468.851 2471.538 2472.014 2472.076 2474.804 2483.362 2507.094 2553.584 2607.558 2592.137 2558.773 2541.243 2533.243
 2465.378 2469.430 2471.393 2470.422 2475.548 2495.187 2522.949 2579.424 2599.434 2586.122 2553.191 2533.450 2525.740
 2460.148 2463.724 2466.122 2466.949 2474.184 2497.440 2538.287 2578.948 2575.724 2552.447 2526.132 2520.551 2513.626
 2456.965 2458.722 2457.006 2462.608 2479.807 2508.644 2570.122 2567.228 2546.949 2524.499 2506.225 2502.628 2506.494
 2456.572 2457.006 2457.006 2462.670 2486.628 2527.083 2557.450 2545.191 2523.528 2500.313 2485.491 2488.158 2492.520
 2454.050 2461.823 2465.213 2474.535 2499.073 2523.755 2522.721 2511.517 2489.460 2475.528 2467.921 2477.657 2475.404
 2443.611 2469.409 2478.814 2478.814 2495.807 2499.094 2498.411 2477.781 2458.846 2452.520 2446.753 2457.337 2460.624
 2431.683 2454.732 2474.598 2476.355 2476.747 2477.119 2463.476 2446.980 2434.143 2428.624 2431.580 2443.197 2451.735
 2419.094 2438.484 2465.461 2471.042 2463.021 2454.009 2447.146 2429.533 2420.251 2418.577 2426.763 2439.166 2442.122
 2409.419 2429.781 2451.735 2454.877 2447.435 2440.014 2434.432 2420.562 2412.127 2417.957 2426.556 2435.900 2435.342
 2396.024 2413.388 2428.251 2430.195 2419.714 2417.626 2417.399 2408.427 2407.497 2413.244 2419.631 2429.843 2429.471
 2375.518 2385.440 2388.996 2388.583 2380.562 2392.014 2393.957 2392.800 2391.125 2400.179 2410.598 2425.419 2426.474
 2351.642 2356.086 2349.492 2350.980 2348.603 2349.926 2370.371 2374.980 2376.159 2388.975 2407.022 2420.396 2419.011
 2318.340 2317.286 2314.040 2311.725 2315.673 2331.756 2348.541 2357.430 2363.590 2372.934 2389.327 2406.587 2409.585
 2277.306 2280.076 2281.358 2278.733 2294.919 2315.136 2323.694 2336.634 2345.192 2355.342 2364.458 2379.053 2389.885
 2237.803 2249.854 2248.015 2239.498 2277.658 2286.898 2302.919 2316.872 2323.963 2338.577 2343.022 2350.567 2358.195
 2208.800 2224.118 2216.407 2224.656 2245.803 2265.772 2279.332 2292.604 2308.728 2328.758 2329.482 2326.340 2332.996
```

Figure 3: 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.

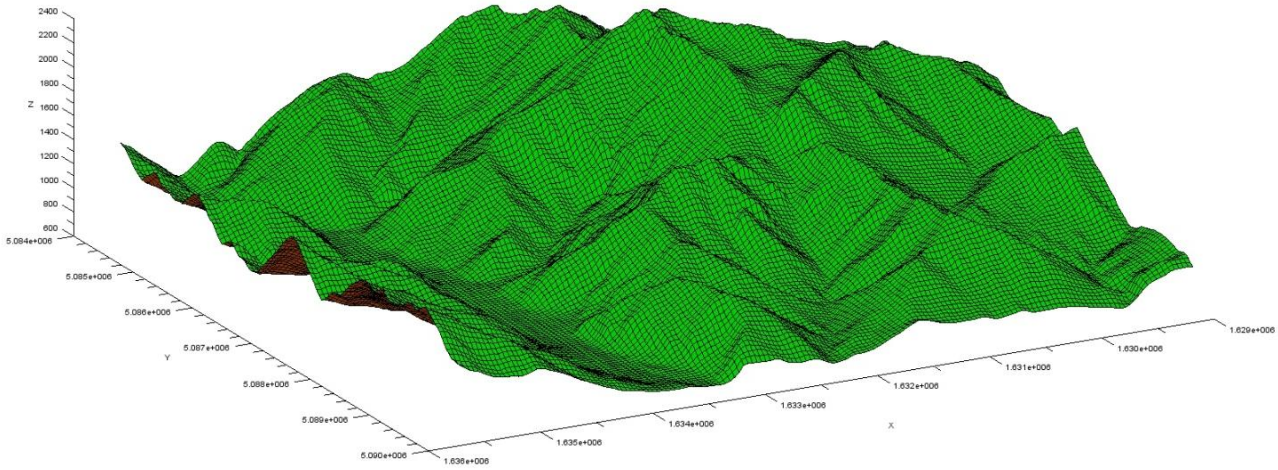

Figure 4: The information contained into one DTM file is graphically rendered. As a results we obtain a plot of the terrain.

4. 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 should be able to build a Kriging model, thanks to the DACE toolbox. Once the model is available, we can ask for the temperature in 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 expert derivation 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 of variation: this is quite interesting when forecasting an environmental temperature.

5. 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 6 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 could 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 5, 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 these two approaches some differences appear, especially down in the valleys: the

Kriging model seems to give more detailed results.
If we consider January the 20th, the temperature can no longer be computed as a function of only the terrain height. It immediately appears, looking at Figure 9, 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 8, is able to capture also local 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 10) is larger than the one obtained for 17th of May (Figure 7): this let us imagine that the temperature is in this case much more difficult to capture correctly.

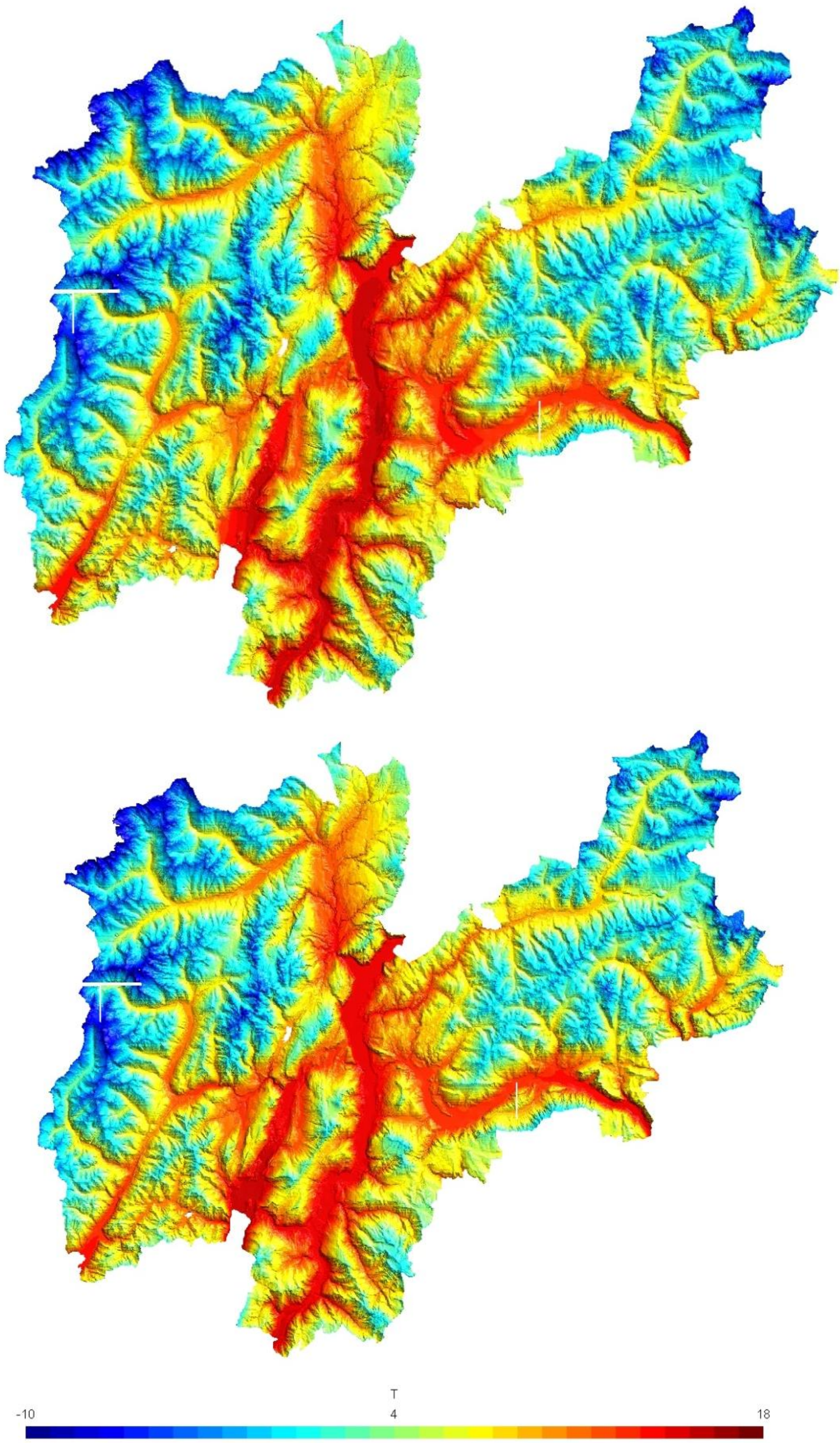

Figure 5: 6th May 2010 at 17:00. Top: the predicted temperature at the terrain level using Kriging is plotted. The temperature follows very closely the height on the sea level. Bottom: the temperature map predicted using a linear model relating the temperature to the height. At a first glance these plots could appear exactly equal: this is not exact, actually slight differences are present especially in the valleys.

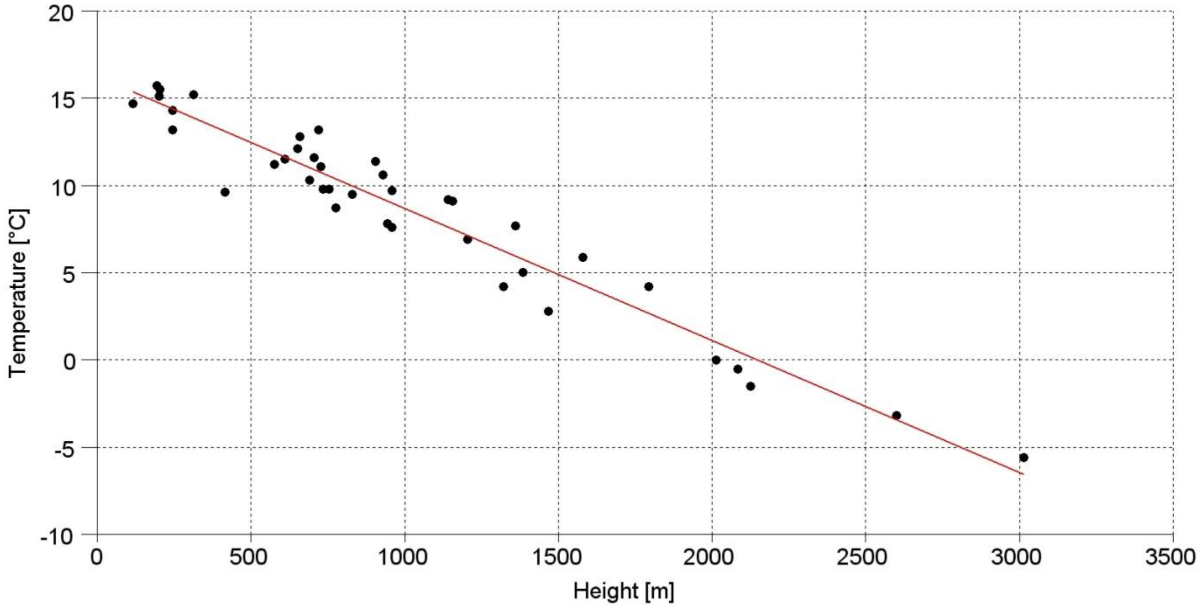

Figure 6: 6th May 2010 at 17:00. The measured temperatures are plotted versus the height on the sea level. The linear regression line, plotted in red, seems to be a good approximation: the temperature decreases 0.756 [°C] every 100 [m] of height.

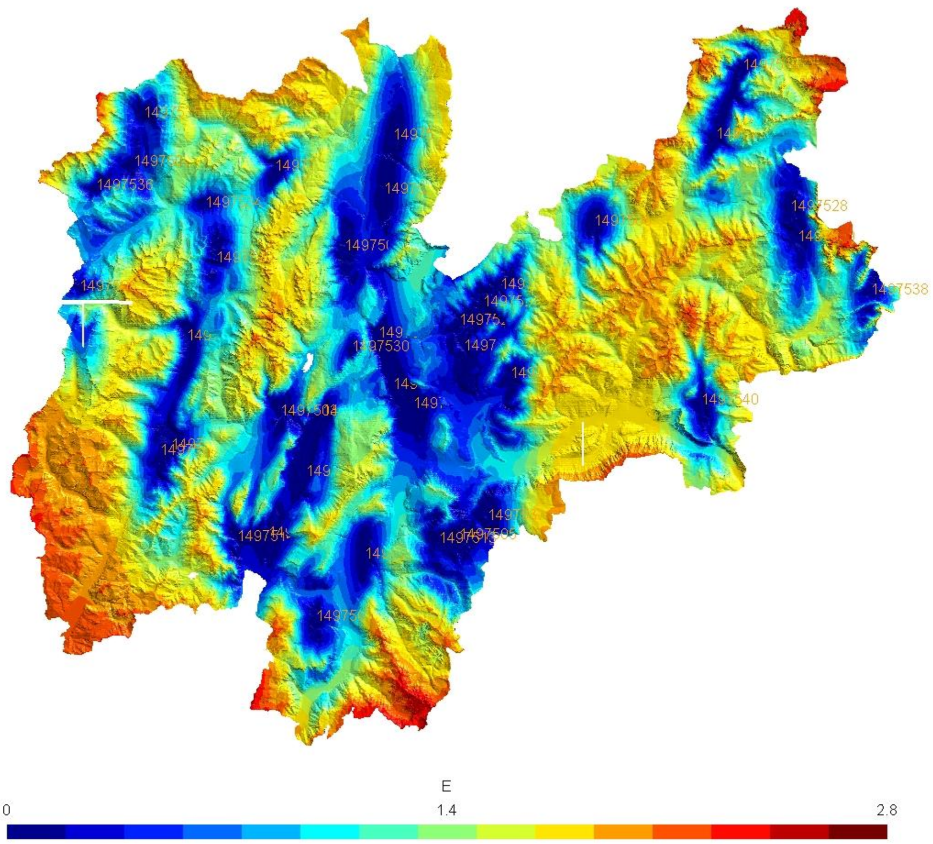

Figure 7: 6th May 2010 at 17:00. The estimated error in predicting the temperature field with Kriging is plotted. The measurement stations are reported on the map with a code number: it can be seen that the smallest errors are registered close to the 39 stations while, as expected, the highest estimated errors are typical of zones where no measure is available.

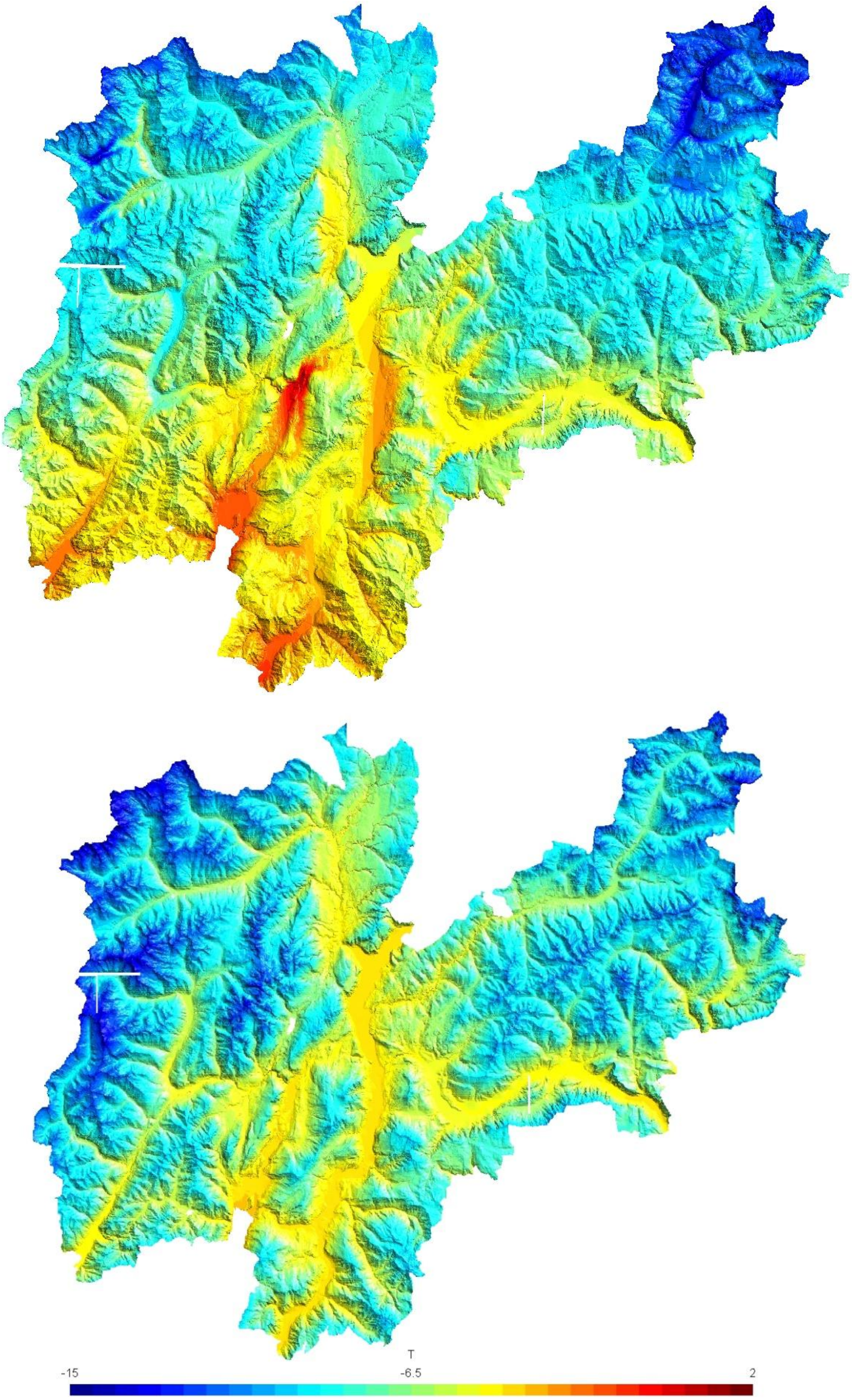

Figure 8: 20th January 2010 at 08:00. Top: the predicted temperature with the Kriging model at the terrain level is plotted. Globally, the temperature still follows the height on the sea level but locally this trend is not respected. Bottom; the temperature map predicted using a linear model relating the temperature to the height.

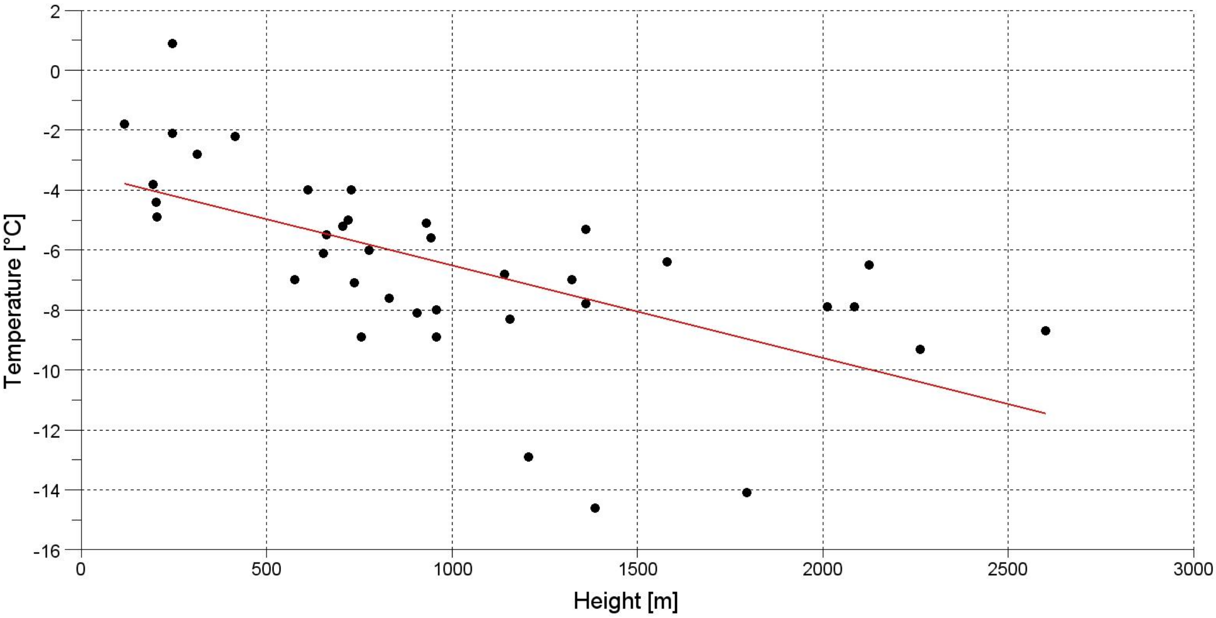

Figure 9: 20th January 2010 at 08:00. The measured temperatures are plotted versus the height on the sea level. The linear regression line, plotted in red, says that the temperature decreases 0.309 [°C] every 100 [m] of height but it seems not to be a good approximation in this case; there are actually very large deviations from the linear trend.

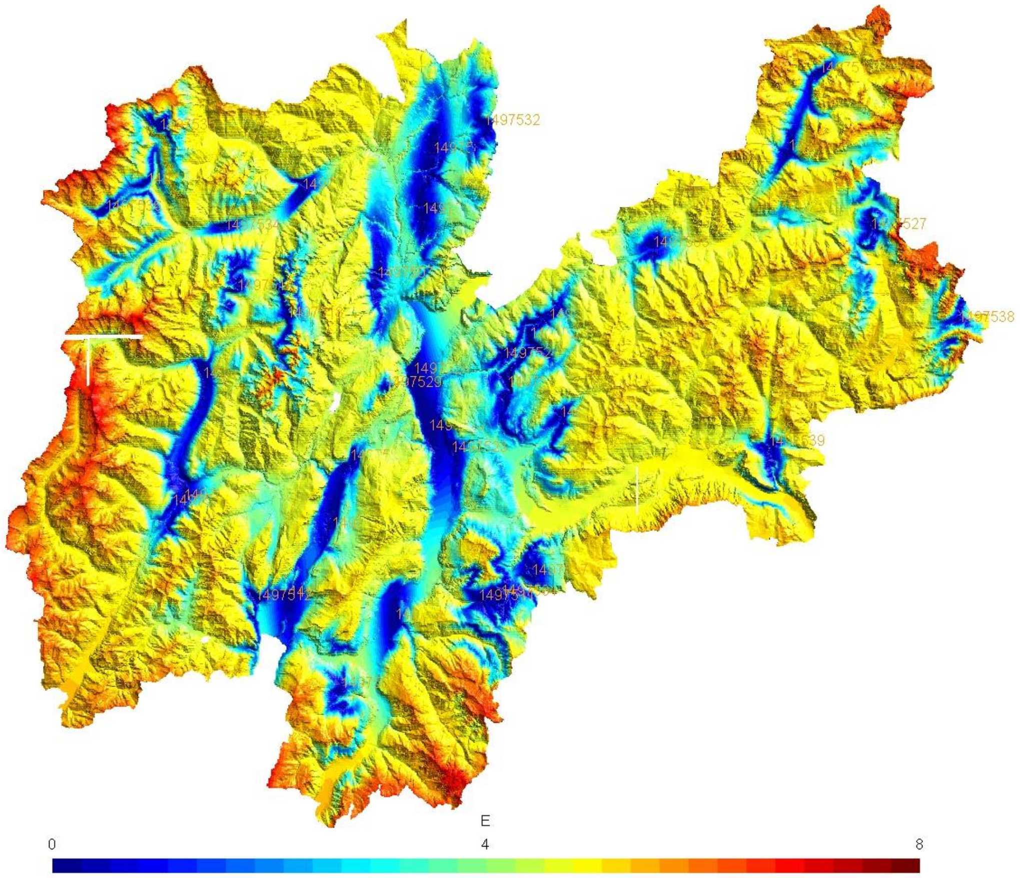

Figure 10: 20th January 2010 at 08:00. The estimated error in predicting the temperature field with the Kriging technique is plotted. The measurement stations are reported on the map with a code number: it can be seen that the smallest errors are registered close to the 38 stations.

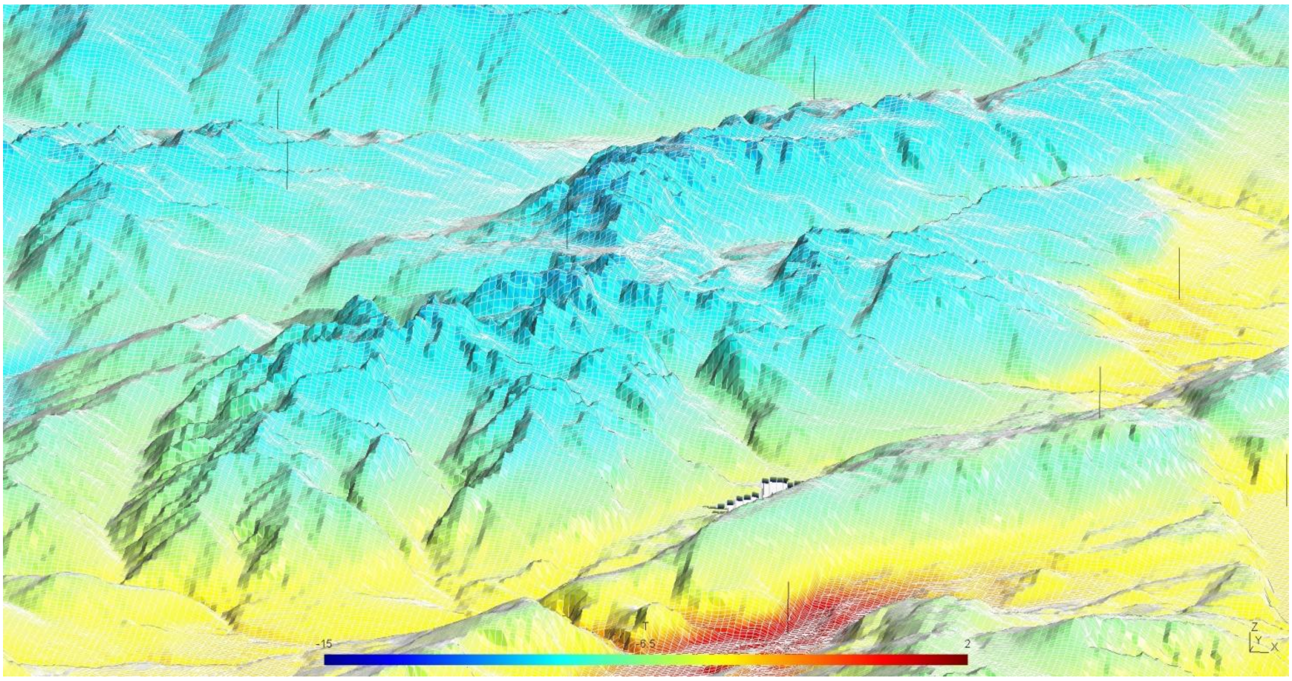

Figure 11: 20th January 2010 at 08:00. The estimated temperature using Kriging: a detail of the Gruppo Brenta. The black vertical bars reports the positions of the meteorological stations.

6. 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 measures and from some information regarding the terrain of the region.
We have shown that the Kriging algorithm can be used to have an estimated value and an expected variation around it: this is a very interesting feature which can be used to have 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.

7. 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.*