

A new approach to the polygons rasterization in GRASS

Ludovico Biagi, Marco Negretti

*Politecnico di Milano – Laboratorio di Geomatica di Como
Via Valleggio 11, 22100 Como
ludo@geomatica.ing.unico.it, marco@geomatica.ing.unico.it*

Within the GIS framework the rasterization of polygons vectorial themes is needed when a raster approach has to be applied in the theme analysis: for instance in the overlay between some normative zones (i. e. a polygon vectorial theme) and one or more environmental phenomena (typically represented by a raster model).

It is well known that, formally, an univocal algorithm doesn't exist; the most popular approaches are the center, the area prevalence and the importance methods. Up to now in GRASS software the first approach was the only one available.

Because of the numerous environmental applications that require one of the other two methods, we have studied and implemented an "approximate" version of both. This is realized by using the center method on a working raster, characterized by an higher resolution than the final one; thus the value of every final cell is computed by evaluating the working cells contained in it. The user can choose the method and the computation approximation, e. g. the ratio between the final and the working raster resolutions; for the importance method an "importance scale" must be specified too. In the paper the approach and the results obtained both on example and on real data analyses are discussed.

THE VECTORIAL MODEL

A *vectorial data structure* models phenomena as geometric primitives composed by points, lines, surfaces and volumes. The vectorial model name is due to the fact that the spatial entities are defined by their coordinates. In the two dimensional domain a point position is described by a set of two coordinates; an arc is described by the two terminal points (nodes) and eventually by the sequence of the geometric discontinuity points (vertices); an area (alternatively named as a polygon) is described by the arcs that compose its perimeter. In three dimensions a volume is described by its boundary surfaces, and so on.

The vectorial approach is based on the emphasis given to the discrete nature of the world: indeed it implies the definition of discontinuity boundaries, composed of simple geometric elements, between different classes or entities; for this reason it is also called *object based description*. Obviously the vectorial approach, or the object based description, is the optimal choice to describe entities that are intrinsically discontinuous, such as the state boundaries or the road networks.

The object based description uses the set of vectorial primitives to describe the more complex spatial entities. The primitives choice and their combination method depend, in addition to the entities complexity, both on the observation scale and on the generalization level. At a small scale, for instance, the cities can be represented by points; at an higher resolution they must be represented by polygons; increasing further on the resolution, the boundaries of the normative zones should also be represented within the city. Note that, typically, in order to completely describe an entity, one or more attributes are associated to its vectorial description: for instance an oil pipe can be modeled by an arc to which its code is associated.

THE RASTER MODEL

The *raster description* is based on the division of the geographical domain in cells (or voxels for three or more dimensional spaces): it is alternatively defined as *location based approach*. Limiting the discussion to the two dimensional space, rectangular and regularly spaced cells are usually adopted; nevertheless other irregular figures can also be chosen. In the first case the cells set is numerically stored as a matrix; in addition, to georeference the matrix, the positions of its outermost corners and the cell dimensions are stored: so, each cell is characterized by its value or attribute and its row and column indexes, that univocally indicate the cell geographic position.

The raster model represents the optimal choice to numerically describe a quasi-continuous surface or phenomenon (i. e. a field) such as the terrain elevation as well as the acoustic noise pollution. The cell dimensions determine the resolution at which the field variations can be represented; typically, either the field value occurring in the center of one cell or the mean field value within the cell is stored.

The raster model can be adopted to describe discontinuous entities too, such as the normative zoning of a city or the presence/absence of some interest activity: in such a case an approach has to be defined in order to convert the original vectorial data to the raster model.

THE VECTOR TO RASTER CONVERSION

In the following the word *rasterization* will be adopted to mean the process of conversion from a vectorial model (Figure 1/a) to a raster model (Figure 1/b). At first, it's worth to note that the conversion implies an unavoidable approximation, that depends on the cells dimensions: the approximation effects can be minimized by choosing a resolution equal to the errors involved by the scale of the original vectorial map; alternatively, if the rasterization is finalized to overlay the vectorial map with a raster, it is reasonable and consistent to adopt the overlay raster resolution. However this is a user problem and not a methodological one.

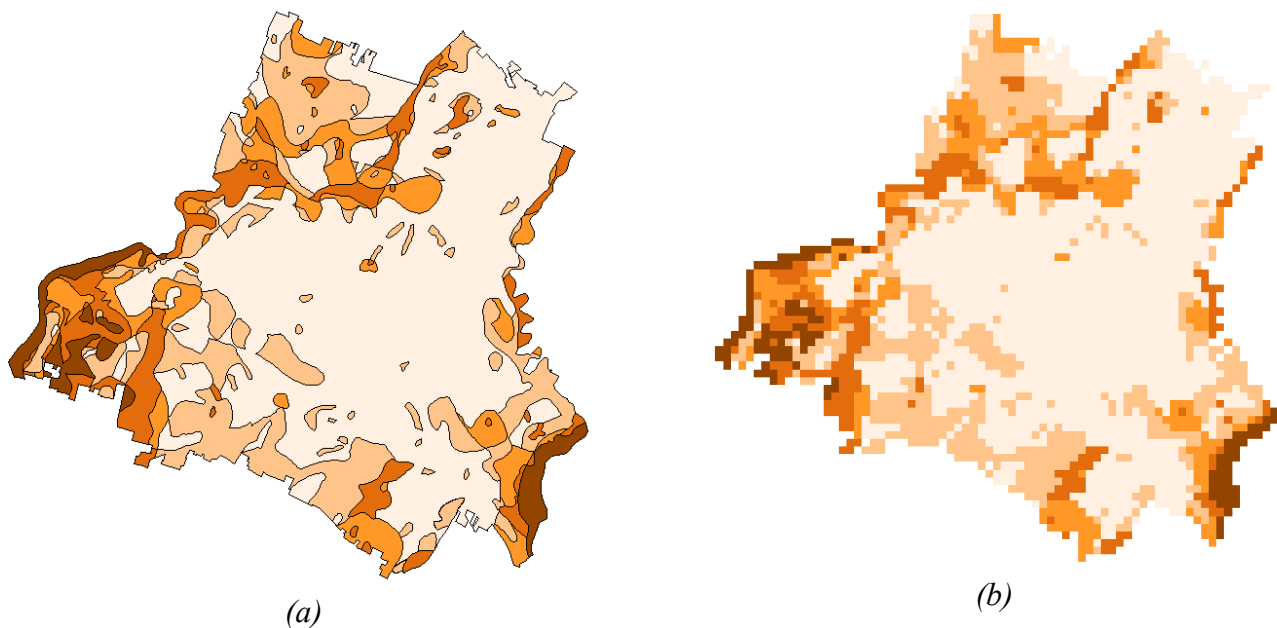


Figure 1 – The same map, according to the vectorial (a) and the raster (b) models

The rasterization process is univocally defined when a method to classify the cells is formalized. The problems involved in the conversion are quite different according to the typology of the

vectorial entities to be rasterized; particularly, limiting the discussion to the two dimensional case, rules must be defined for the following processes:

1. the points to raster conversion;
2. the arcs to raster conversion;
3. the polygons to raster conversion.

Our study and the new GRASS command concern the third process.

Referring to the example map shown in the (Figure 2), many possibilities exist; obviously an unambiguous value can be attributed to a cell when it contains only one polygon: the C value will be attributed to the North-West cell. An ambiguity arises when, within the cell, more polygons exist: in such a case precise conversion rules must be defined. The most popular methods are:

1. the area prevalence method;
2. the center method;
3. the importance method.

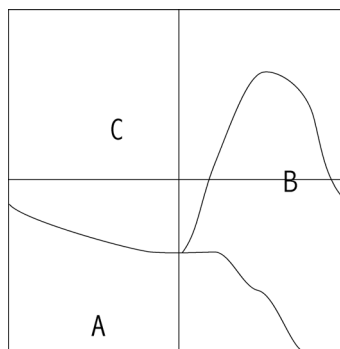


Figure 2 – The vectorial map and the raster cells

The prevalence method (Figure 3/a) attributes to the cell the value of the polygon spatially prevalent in the whole cell. The approach is quite intuitive and it is suitable for generic purposes.

The center method (Figure 3/b) attributes to the cell the value of the polygon that occupies its center. The method is usually considered as an approximation of the previous one, because the evaluation is performed on a single cell point rather than on the whole cell surface.

The importance method (Figure 3/c) requires the definition of a hierarchy among the values: in the example map the hierarchy $A > B > C$ is adopted. The most important value among those existing in a cell is attributed to the whole cell; moreover, the attribution can be done disregarding the percentages of area occupation as well as superimposing a minimum threshold value. Usually, the approach is adopted when it is desired to emphasize the presence of a particular value (i.e. attention or risk class) within the cells.

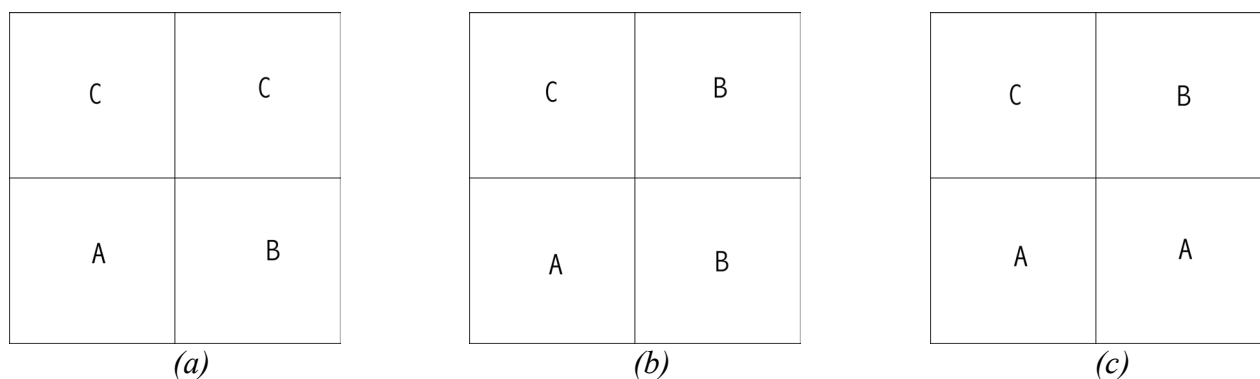


Figure 3 – The rasterization results

THE NEW GRASS COMMAND: THE COMPUTATION PROCESS

The standard GRASS command (`v.to.rast` command in the 5.0B8 release) performs the center method; our aim was to implement the other two approaches.

To do that we started from the center method, in such a way to exploit the existing GRASS command: the algorithm works in the following manner.

1. Once given a vectorial map (Figure 4) and defined the resolution of the final raster map (for example 2×2 , Figure 5/d), the implemented algorithm generates a raster (working raster, Figure 5/a) whose resolution is higher than the final one.
2. each cell of the working raster is classified according to center method (Figure 5/b).
3. for each cell of the final raster, a count of the values assigned to all the working cells belonging to it is performed: the cell is then classified applying the prevalence or importance method (Figure 5/c and Figure 5/d) to the computed statistics.

Obviously, our implementation furnishes an approximated approach both to the “exact” prevalence and to the “exact” importance methods. The resolution ratio between the working and the final raster has to be specified by the user: an higher resolution (ratio) allows a lesser, or negligible, approximation but it is numerically slower; a ratio equal to one corresponds to the standard command. Moreover, in the importance method, the user must specify the class hierarchy and the minimum threshold value.

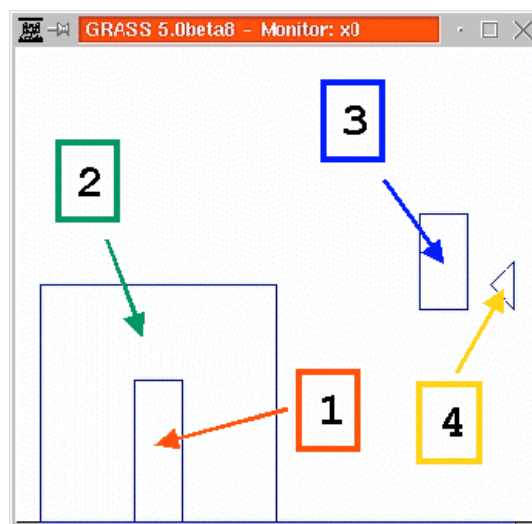


Figure 4 – The test vectorial map

THE TESTS ON SIMULATED DATA

We tested the new command on several synthetic data set. In the following some examples are discussed; the test vectorial map is the same shown in the (Figure 4); for the reader convenience, it is repeated in (Figure 6/a) and (Figure 7/a).

The (Figure 6) refers to a final raster of 2×2 cells, while the (Figure 7) raster is composed of 4×4 cells; in both the examples the applied working/final resolution ratio was equal to 10.

In the first example the center method (Figure 6/b) attributes to the S-W cell the value 1, that occupies the cell center; the prevalence method (Figure 6/c) attributes to the cell the value 2, that occupies the major part of the cell. Note that, at the present, our algorithm manages the null values as a value: so the other cells are not classified because null values are prevalent.

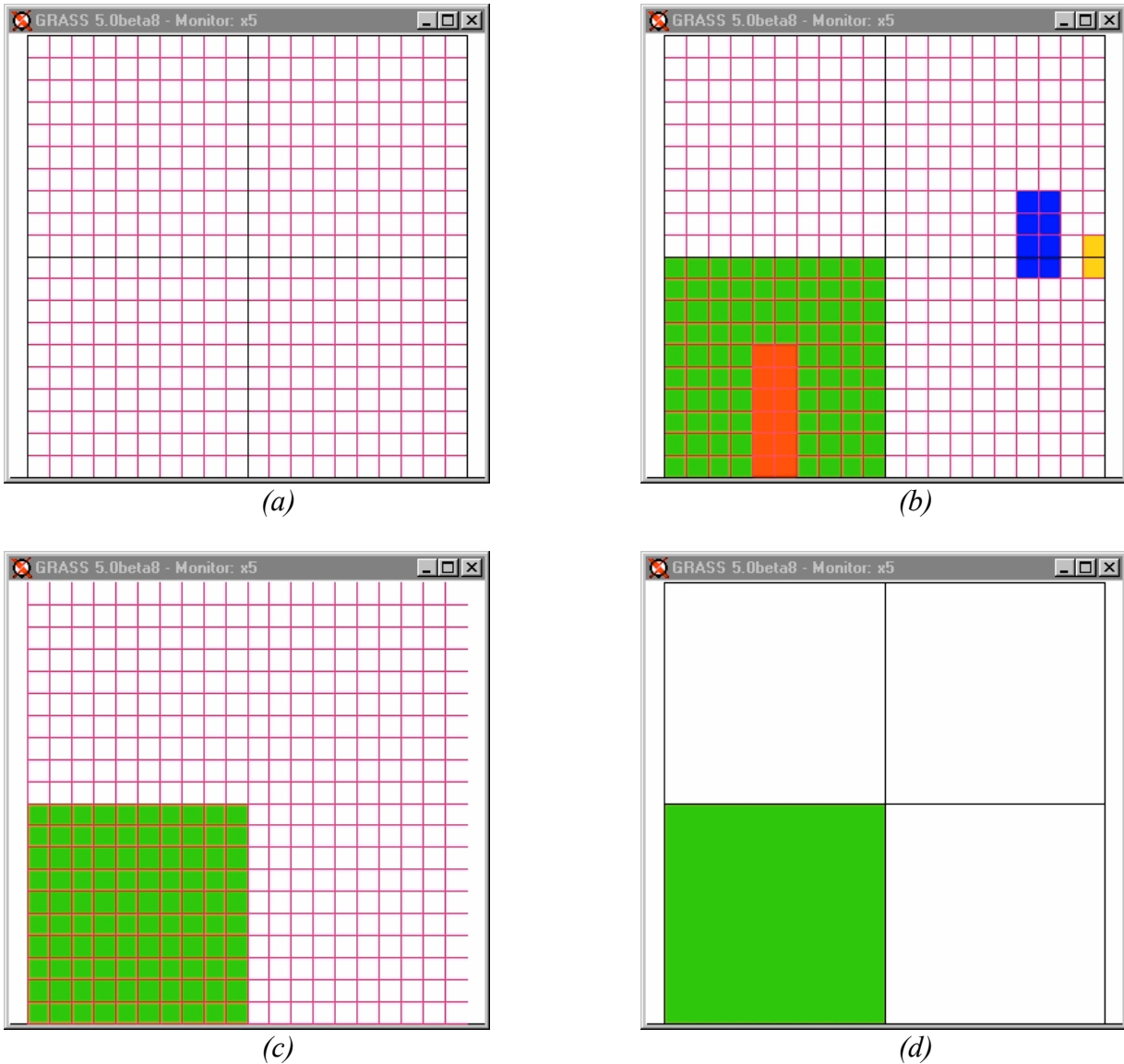
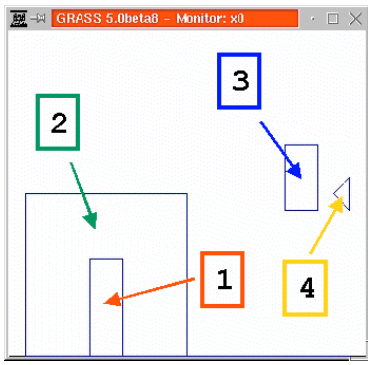
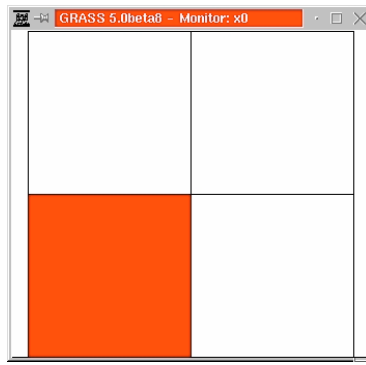


Figure 5 – The working and the final raster

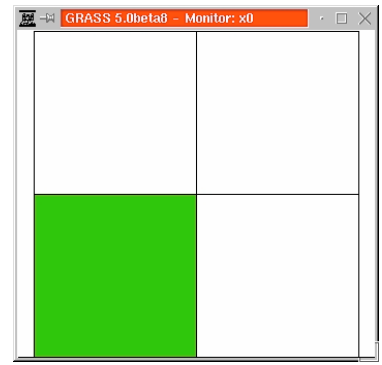
In the importance method tests the class hierarchy $1 > 3 > 2 = 4$ was imposed; in the first test a minimum threshold value equal to 5% was specified (Figure 6/d): correctly the value 1, that is the most important class in the cell and that occupies more than 5% of its area, is attributed to the S-W cell; analogously the N-E cell is classified in the class 3. Choosing an higher threshold value, 10% (Figure 6/e), the N-W cell is not classified because the value 3 occupies less than 10% of its area; when a 15% threshold value is adopted (Figure 6/f), the rasterization results are the same furnished by the prevalence method: indeed no “important” classes occupy more than 15% of the final cells. Consider now the 4 x 4 final raster (Figure 7). By using the center method (Figure 7/b), the value 3 is attributed to one cell (the S-E of the N-W block): indeed the value occupies the cell center, but it isn’t the prevalent one. Correctly the prevalence method (Figure 7/c) attributes the null value to the same cell. By using the importance method, with the previous hierarchy, an increase of the threshold value (Figure 7/d to Figure 7/f) leads to a decrease of the number of cells classified according to the hierarchy: a prevalence classification is furnished by the 30% threshold value.



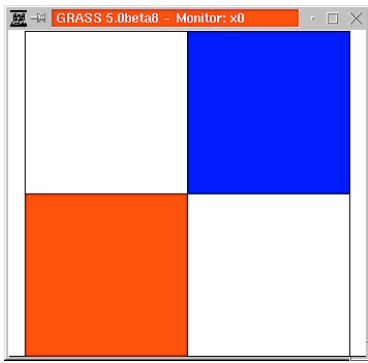
(a)



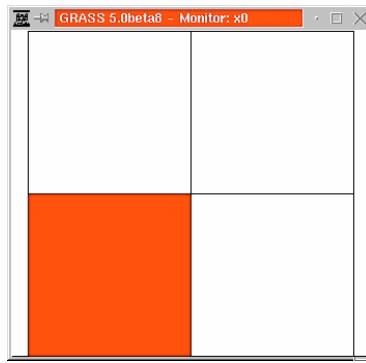
(b) center



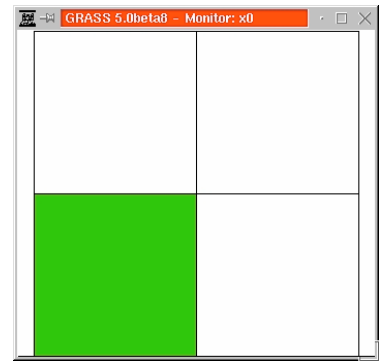
(c) prevalence



(d) importance
threshold = 5%

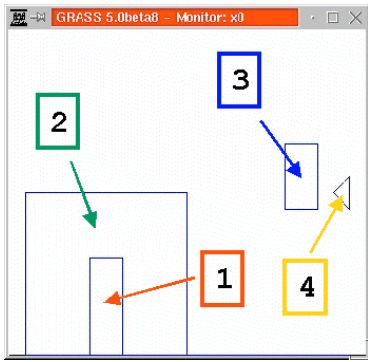


(e) importance
threshold = 10%

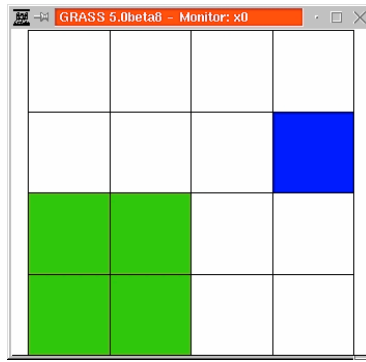


(f) importance
threshold = 15%

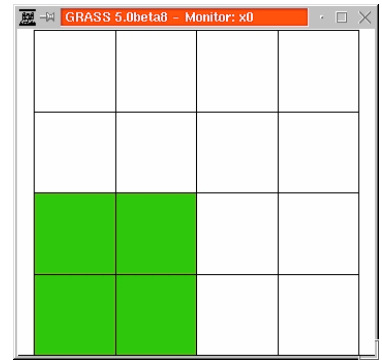
Figure 6 - 2x2 rasterization



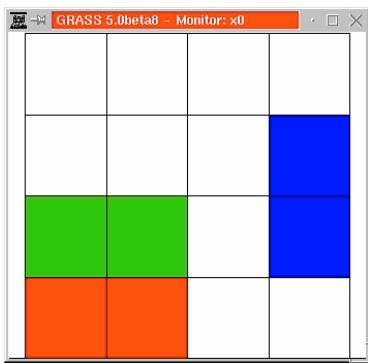
(a)



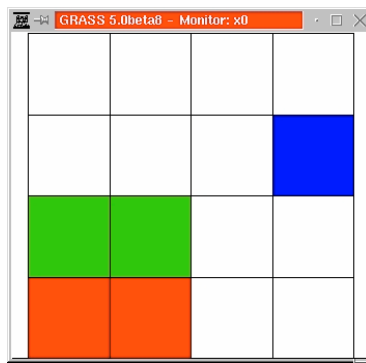
(b) center



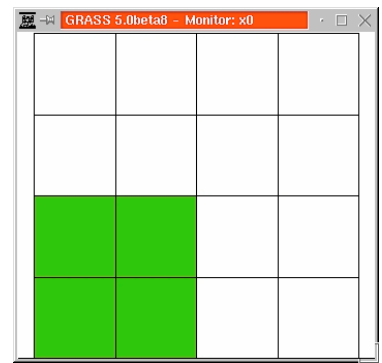
(c) prevalence



(d) importance
threshold = 5%



(e) importance
threshold = 15%



(f) importance
threshold = 30%

Figure 7 - 4x4 rasterization

THE TESTS ON REAL DATA

The new rasterization command has been tested on real data to verify its numerical performances: two examples are shown, related to the catchment basins map of the Modena (Italy) municipality (Figure 8).

The tests were performed on a PC whose characteristics are here summarized:

- O.S.: Linux Red Hat 6.2;
- Processor: Pentium III 450 MHz;
- RAM: 128Mbyte;
- HD: SCSI Quantum Atlas IV 9 WIS.

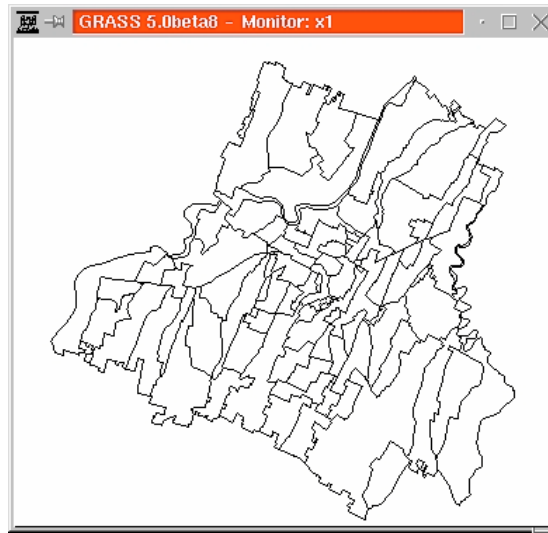


Figure 8 – The catchment basins map

The two examples differ in the final raster resolution; both are generated by the prevalence method with a working/final resolution ratio equal to 10. Note that this ratio corresponds to a division of each final cell into 100 working cells.

In the (Figure 9/a) the region settings relevant to the first test are shown; the final raster (100 m x 100 m resolution) is depicted in the (Figure 9/b). The working raster is composed by almost 600000 cells: the whole computation required 3 minutes. The second test relates to a final resolution of 100 m x 100 m (Figure 10/a). The working cells number exceeds 3.5 millions: the whole computation required 15 minutes (Figure 10/b).

The results provided by our approximated implementation have been compared with those obtained by applying the “exact” prevalence method, implemented by the *polygrid* command of ARC/INFO[®], release 8.0. At first note that both the ARC/INFO tests required less than three minutes of computation time; they were performed on a different, Windows NT based, PC: so, a rigorous performance comparison is not possible but undoubtedly the ARC/INFO command is numerically more efficient than the our implementation.

The accuracy comparison (Table 1) is restricted to the cells where a real value is prevalent; in the same table the errors involved by the center approximation are also shown. It is evident that the proposed algorithm works well and that it constitutes a real improvement with respect to the original center method; this holds especially when the wanted final raster has a coarse resolution. Obviously the results could be improved further on, by augmenting the working resolution ratio; on the other hand that implies an increase in the computation time

```

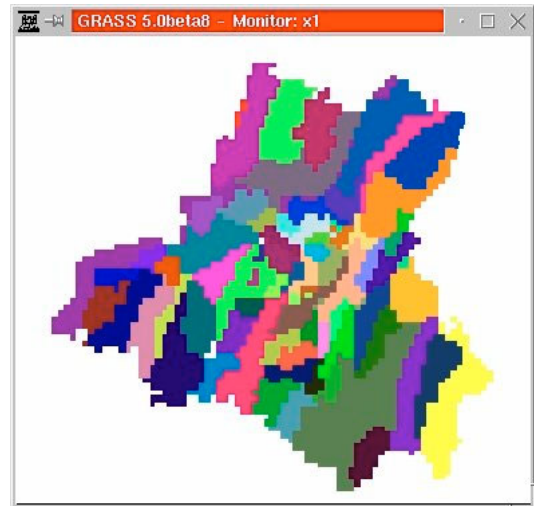
projection: 1 (UTM)
zone:      13
north:    955160
south:    935910
east:     660200
west:     640950

e-w res:  250
n-s res:  250

total rows:      77
total cols:      77
total cells:    5929

```

(a)



(b)

Figure 9 – The first real data test

```

projection: 1 (UTM)
zone:      13
north:    955160
south:    935960
east:     660200
west:     641000

e-w res:  100
n-s res:  100

total rows:      192
total cols:      192
total cells:    36,864

```

(a)

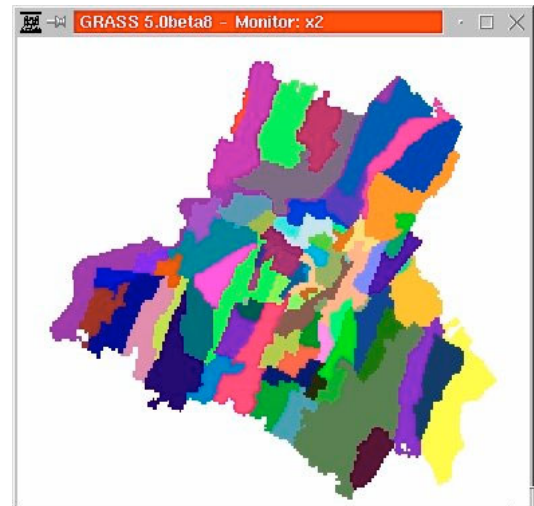


Figure 10 – The second real data test

Case	Cells Number (dimensions, meters)	Our method		Center Method	
		Cells N. Err.	Cells % Err.	Cells N. Err.	Cells % Err.
Test 1	2939 (250 x 250)	7	0.23	62	2.11
Test 2	18351 (100 x 100)	23	0.12	66	0.36

Table 1 – The errors of the proposed approximated method on a real data set. In the table the number and the percentage of the wrong classified cells by the new GRASS command (resolution ratio equal to 10) and by the center method are shown. The “truth” results are computed by the ARC/INFO[®] polygrid command.

THE COMMAND SYNTAX

v.to.rast2

input = name	vectorial map to be converted
output = name	final raster
[mod = name]	rasterization method: center (cnt) [default] prevalence (prv) importance (imp)
[optrast = value]	working/final resolution ratio [default=2]
[imp = name]	class hierarchy file name (imp. method)
[thr = value]	threshold value (imp. method)

CONCLUSIONS

The standard rasterization GRASS command (v.to.rast command in the 5.0β8 release) has been improved in order to allow the polygons to raster conversion according to the prevalence and the importance methods. The new command, named v.to.rast2, follows an approximated approach rather than the “exact” one; the approximation level can be chosen by the user, according to the desired classification accuracy. In the future we plan to numerically optimize the code reducing the computation time. At present the main goals are to add the option for the null values management and to work on the rasterization both of the points (sites) and the arcs themes.

REFERENCES

- Biagi L., Brovelli M. A., Muratori A., Rubbiani M. (1998) *Gestione di dati territoriali ed analisi di destinazione d'uso del territorio mediante GIS*, Atti della 2^a Conferenza Nazionale ASITA, Rilevamento, rappresentazione e gestione dei dati territoriali e ambientali
- L. Biagi, M. A. Brovelli, M. Negretti, (2000), *The ISOLA project, methodological and numerical problems: some solutions and a first process prototype*, Proceedings of UDMS2000, Urban and Regional Data Management – Common Problems – Common Solutions, Delft, The Netherlands
- C. B. Jones, (1997), *Geographical Information System and Computer Cartography*, Longman, Edinburgh.
- ESRI, (1992), ARC/INFO 8.0.2 User Manual, ESRI
- GRASS Developer Team, (2000), *GRASS 5.0 Reference Manual*, (<http://grass.itc.it/>)
- GRASS Developer Team, (2000), *GRASS 5.0 Programmer's Manual*, (<http://grass.itc.it/>)
- M. van Kreveld, J. Nievergelt, T. Roos, P. Widmayer, (1997), *Algorithmic Foundations of Geographic Information Systems*, Springer, Berlin