

A NEW MULTI-RESOLUTION ALGORITHM TO STORE AND TRANSMIT COMPRESSED DTM

Ludovico Biagi *, Maria Antonia Brovelli, Giorgio Zamboni
Politecnico di Milano, DIAR, Geomatics Laboratory at Como Campus, via Valleggio 11, 22100 Como, Italy
ludovico.biagi@polimi.it, maria.brovelli@polimi.it, giorgio.zamboni@polimi.it

Commission IV, WG IV/1

KEY WORDS: Internet/Web, LIDAR, Modelling, Archiving, Compression, DEM/DTM, Algorithms, GIS

ABSTRACT

WebGIS and virtual globes allow DTMs distribution and three dimensional representations to the Web users' community. In these applications, the database storage size represents a critical point.

DTMs are obtained by some sampling or interpolation on the raw observations and typically are stored and distributed by data based models, like for example regular grids. A new approach to store and transmit DTMs is presented. The idea is to use multi-resolution bilinear spline functions to interpolate the observations and to model the terrain. More in detail, the algorithm performs the following actions.

1) The spatial distribution of the observations is investigated. Where few data are available, few levels of splines are activated while more levels are activated where the raw observations are denser: each new level corresponds to an halving of the spline support with respect to the previous level.

2) After the selection of the spline functions to be activated, the relevant coefficients are estimated by interpolating the observations. The interpolation is computed by batch least squares.

3) Finally, the estimated coefficients of the splines are stored.

The model guarantees a local resolution consistent with the data density and can be defined analytical, because the coefficients of a given function are stored instead of a set of heights.

The approach is discussed and compared with the traditional techniques to interpolate, store and transmit DTMs, considering accuracy and storage requirements. It is also compared with another multi-resolution technique. The research has been funded by the INTERREG HELI-DEM (Helvetia Italy Digital Elevation Model) project.

1. INTRODUCTION

Nowadays, Digital Terrain Models (DTMs, Li et al, 2005) represent fundamental databases for the Geographical Information Systems (GIS, O' Sullivan and Unwin, 2003). Up to few years ago, DTMs were only used in specific applications of territorial analyses, typically by the scientific community. The coming and the diffusion of the new technologies based on Web GIS and virtual globes have changed the perspective: altimetric analyses and three dimensional representations of the terrain are still object of new researches but also praxis.

With respect to the traditional programs for the two dimensional representation, virtual globes have introduced the third dimension and, consequently, a simpler usage and a greater visual consistence between the digital representation and the real world. At present, the new acquisition techniques provide information with a never seen accuracy. Therefore, virtual globes are no more merely qualitative viewers for low resolution global data, but can become scientific instruments to process and analyze high accuracy geographical information.

Virtual globes and Web GIS cannot be properly compared, but they share a fundamental principle: the geographic information (satellite and aerial images, height data, vectorial objects) is accessed via Web. Particularly, the servers provide the data according to specific transmission standards that have been defined mainly by the Open Geospatial Consortium (OGC, 2006, 2010a, 2010b).

The new DTMs provide height information with never seen accuracy and spatial density (El-Sheimy et al., 2005). However, in the Web distribution of geographical information, the databases storage size represents a critical point. Given a specific interest area, typically the server needs to perform some preprocessing, the data have to be sent to the client, that applies some additional processing: the efficiency of all these

actions is crucial to guarantee a near real time availability of the information.

Generally speaking, the terrain surface is composed by an infinite number of points: a DTM is obtained by interpolating the available height observations and extracting a finite dataset that allows the reconstruction of the whole surface at a given accuracy. DTMs can be stored and transmitted according to two different approaches: the data based models and the analytical models (Biagi et al., 2011).

In the data based models, the DTM is stored and transmitted by a sample of interpolated heights that are used to reconstruct (interpolate) the terrain heights in other points. On the contrary, an analytical model implies the storage of a dataset of coefficients that, in a one to one relation with a given function, allows the computation of the height everywhere.

The purpose of this paper is to discuss a new analytical approach, that is based on an least squares interpolation by multi-resolution bilinear splines and has been already discussed in its preliminary implementation in (Brovelli and Zamboni, 2009). The raw observations are interpolated by a linear combination of splines with compact support, whose resolutions and positions vary in space and are automatically chosen according to the distribution of the raw observations. In the following, they will be called multi resolution splines. For each spline, the resolution level, the position and the coefficient are stored by the server and are transmitted to the clients. The coefficients and their auxiliaries metadata allow the complete reconstruction of the terrain at any point and different detail levels can be provided, according to the required accuracy. The purpose of the proposed approach is the saving of storage requirements with respect to the traditional models without any loss of accuracy.

The following part of the introduction shortly summarizes the data models and interpolation methods that are typically

adopted to store and compute DTMs. Section 2 discusses our new approach and its storage requirements. In Sect. 3, the performances of the different models are analyzed, by comparing their application to a case study. The conclusions and the future developments follow.

1.1 Data based Models

Different data models can be adopted to store and transmit a DTM. Contour lines, grids (or elevation matrix) and Triangular Irregular Networks (TIN) are the standard.

Contour lines are obtained by connecting with a line all the points with the same height. The lines are drawn at given, equally spaced in height, intervals. Contour lines are useful to visualize heights on maps in 2D applications, but seldom are used for analysis purposes, and are stored and transmitted following the general rules of vector objects.

Gridded DTMs (in the following DTM_{GRID}) are georeferenced as regular grids of nodes, whose heights are stored. The storage of a grid requires a set of metadata that allow its georeferencing (see Sect. 3) and are listed in the so called header. The heights are stored in an ordered sequence. DTM_{GRID} is a very simple conceptual model and can be easily accessed, visualized and spatially analyzed by map algebra. However, the choice of the grid resolution is a crucial point, because the storage size is inversely proportional to the square of the gridding interval. If rough terrain (for example mountains) alternates to flat terrain (plains), the high resolution needed to accurately describe the first causes a useless redundancy in the second. To continuously describe the heights between the nodes, either a bilinear or a bicubic interpolation is typically applied.

In a TIN, the DTM (in the following DTM_{TIN}) is described by a set of planar triangular faces that are obtained by connecting sparse points, whose horizontal coordinates and heights are given. Usually, the Delaunay criterion is applied to triangulate the points. By a TIN model, more points can be stored where the terrain is rough and less points are used in flat areas. Each point of a TIN is represented by its three (X, Y, height) coordinates. Moreover, to reconstruct the topology of the triangles, the labels of the three vertices of each triangle are needed. This simple data model requires long computation times for the processing and the analysis of the 3D surface: therefore, in the practice, more complex topological models are applied, like for example the node based, the triangle based or the edge based data structures. These models reduce computation times but require an overhead of information that is stored and transmitted to the clients. When a TIN model is used, the height within each triangle is linearly interpolated from its three vertices.

1.2 Interpolation techniques

To produce a DTM, several interpolation techniques exist: a first classification can be into *exact* and *approximate* interpolations. An exact interpolator passes for all the observations and allows the complete reconstruction of all the discontinuities existing in the dataset. However, the observation errors are not filtered and propagate into the model. A classical example of exact interpolator is given by the *Inverse Distance Weighting* (IDW). Approximate interpolators apply statistical methods to estimate a smoother function from the observations: in this way, the errors can be filtered and both the observations accuracy and the function correctness can be assessed. However, actual details and discontinuities can be lost in the smoothing. *Local Polynomial* (POL) is an approximate interpolator when the coefficients are fewer than the observations and are estimated by least squares.

In the *deterministic* interpolation, either exact or approximate, the analytical model of the surface is a priori chosen and the observations are used to estimate it: IDW and POL are examples of deterministic interpolators. In the *stochastic* interpolation (Christakos, 1992), the observations are considered as a sample of a random field (the surface) that is completely described by spatial stochastic properties like, for example, the covariance function. The stochastic properties are estimated analyzing the observations and then applied to interpolate the surface. Collocation and kriging are the classical examples of stochastic interpolators.

Note that the most popular interpolation techniques, as reported in scientific and technical literature, cannot be easily and efficiently used to implement an analytical model because the interpolating functions cannot be described by a small number of parameters or coefficients. In IDW and POL, the interpolation coefficients and domain are a function of the positions of both the interpolation point and the observations: to reply the model, all the observations must be stored and distributed. Radial Basis technique uses a linear combination of radial functions that interpolate exactly the observations and are characterized by the minimum curvature. These methods (Regularized Spline, Spline with Tension, Thin Plate Spline) differ in the function choice, all of them could be analytically described by a finite set of coefficients but the needed coefficients are at least as many as the raw observations.

Let consider a stochastic interpolator, for example the collocation. The height in a point is provided by the $\hat{h}(P) = \mathbf{c}^T \boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is the vector of the observations multiplied by its inverse covariance matrix, \mathbf{c} is the cross-covariance vector between the point and the observations. \mathbf{c} can be built by knowing the covariance function of the surface and the positions of the observations, while $\boldsymbol{\xi}$ needs to be stored: also in this case, an analytical model would require as many data as the original observations.

The classical bilinear splines estimated by least squares provide a twofold interpretation, because they can be thought as both data based and analytical models. Given the required spatial resolution, the observations are interpolated to estimate the coefficients of the splines, that are used to predict the heights on a regular grid, that represents the data based model. If the splines and the grid have the same spatial resolution, the coefficients of the splines and the heights of the relevant grid nodes are equal. Moreover, the coefficients of the bilinear splines used to interpolate from the four neighboring nodes of a regular grid are exactly the relevant four heights: indeed each local bilinear spline assumes the maximum in its node, while annihilates on all the other nodes. In this case, the analytical model has exactly the same complexity of the data based model.

The adoption of a new multi-resolution splines interpolation has been studied, that represents a true analytical model and provides actual storage and distribution saving with respect to a data based model.

2. THE MULTI-RESOLUTION SPLINES APPROACH

The approach has been previously discussed in a preliminary way by (Brovelli and Zamboni, 2009): it is an approximate deterministic method whose estimation principle is based on Least Squares (LS, Kock, 1987).

The founding idea is to combine splines with different width in order to guarantee the resolution adequate to the data density in every region of the field, exploiting all available information implicitly stored in the sample. Different levels of splines,

corresponding to different halving steps, are considered. A new level halves the width of the support of the previous level.

We suppose that a field $h(\mathbf{t}) = h(x, y)$ has been sampled at m locations $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m$, $\mathbf{t}_i = (x_i, y_i)$. The interpolation domain is $[\mathbf{t}_{\min}, \mathbf{t}_{\max}]$. The field observations y_{0i} are modelled by means of a suitable combination of bilinear spline functions and noise v_i : $y_{0i} = h(\mathbf{t}_i) + v_i$.

Let define the following

$$\varphi(q) = \begin{cases} 1-q & 0 \leq q < 1 \\ 0 & q > 1 \\ \varphi(-q) & q < 0 \end{cases}$$

$$\varphi_{\Delta}(q) = \varphi(q/\Delta)$$

The height field is given by the

$$h(\mathbf{t}) = \sum_{h=0}^{M-1} \left[\sum_{i_x=0}^{N_h-1} \sum_{i_y=0}^{N_h-1} \lambda_{h,i_x,i_y} \varphi_{\Delta x_h}(q_x) \varphi_{\Delta y_h}(q_y) \right] \quad (1)$$

where $q_x = x - \Delta x_h i_x - x_{\min}$, $q_y = y - \Delta y_h i_y - y_{\min}$, $\Delta x_h = \frac{x_{\max} - x_{\min}}{2^{h+1}}$, $\Delta y_h = \frac{y_{\max} - y_{\min}}{2^{h+1}}$. M is the number of levels, λ_{h,i_x,i_y} is the coefficient of the spline at the grid node (i_x, i_y) , N_h is the number of nodes at the h level, $N_h = 2^{h+1} + 1$.

In the estimation, all the field observations are tiled in a vector $\mathbf{y}_0 = \mathbf{y} + \mathbf{v} = \mathbf{A}\boldsymbol{\lambda} + \mathbf{v}$, $E\{\mathbf{v}\} = \mathbf{0}$, $\mathbf{C}_{yy} = \mathbf{C}_{vv} = \sigma_0^2 \mathbf{I}$ where $\boldsymbol{\lambda}, \mathbf{A}$ are respectively the vector containing all the λ_{h,i_x,i_y} coefficients to be estimated and the design matrix obtained by applying (1) to the observations. The estimation of $\hat{\boldsymbol{\lambda}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}_0$ is based on the well known LS principle. Two innovative aspects characterize our interpolation approach.

Given a level, each local spline is individually activated if no spline of some lower level has the same application point. Moreover the spline is activated if at least $f, f \geq 1$, observations exist in at least k ($k=1,2,3,4$) quarters of the spline support: f, k are input by the user. They must be chosen according to two criteria. Clearly $f=1, k=1$ correspond to no redundancy in the estimation, while bigger values smooth the interpolating field. Moreover, particular spatial configurations of the observations can produce a LS system that, although redundant, is either rank deficient or ill conditioned: in these cases, f and k should be increased independently of redundancy considerations. The individual activation of the splines guarantees a real multi-resolution interpolation.

The levels are activated iteratively from level 0 to level M . A new level is activated if and only if its splines significantly improve the accuracy of the interpolation.

Let suppose that M ($h=0,1,\dots,M-1$) levels have been already activated, for a total number of N_M splines estimated coefficients. The criterion to activate or not the $M+1$ -th level is based on a significance analysis. Let suppose that N_{M+1} is the number of splines activated with the new level and use

$\hat{\boldsymbol{\lambda}}_{M+1}$ to indicate the vector containing all the splines coefficients of the new level. We want to evaluate the following hypothesis

$$H_0 : \left\{ \boldsymbol{\lambda}_{M+1} = E\{\hat{\boldsymbol{\lambda}}_{M+1}\} = \mathbf{0} \right\} \quad (2)$$

If H_0 is true, the coefficients of the new level are not significant, the relevant estimates can be discarded and the iterative process can be stopped. Otherwise, the coefficients should be kept and a new higher level should be tested. Let define the following quantities

$$\mathbf{v}_M = \mathbf{y}_0 - \hat{\mathbf{y}}_M, \hat{\sigma}_M^2 = \mathbf{v}_M^2 / (m - N_M)$$

$$\mathbf{v}_{M+1} = \mathbf{y}_0 - \hat{\mathbf{y}}_{M+1}, \hat{\sigma}_{M+1}^2 = \mathbf{v}_{M+1}^2 / (m - N_M - N_{M+1})$$

where $\mathbf{y}_0 = [h_1, \dots, h_m]^T$ is the vector containing all the field observations, $\hat{\mathbf{y}}_M, \hat{\mathbf{y}}_{M+1}$ are the a posteriori estimates provided by LS. From a geometrical point of view the situation is depicted in Fig. 1.

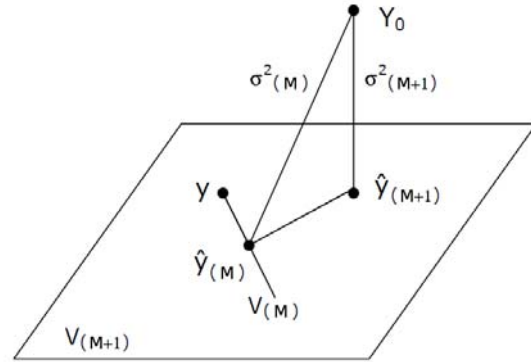


Figure 1. Geometric interpretation of the significance analysis of a new level.

If (2) holds, $\mathbf{y} = E\{\mathbf{y}_0\} \in V_M$, and the usual significance analysis on the a posteriori variances can be applied

$$\frac{(m - N_M) \hat{\sigma}_M^2 - (m - N_M - N_{M+1}) \hat{\sigma}_{M+1}^2}{N_{M+1} \hat{\sigma}_{M+1}^2} \approx \frac{\chi_{N_{M+1}}^2 / N_{M+1}}{\chi_{m - N_M - N_{M+1}}^2 / (m - N_M - N_{M+1})} \mathbf{v}_{M+1} = F_{N_{M+1}, m - N_M - N_{M+1}} \quad (3)$$

where $\chi_i^2, F_{i,j}$ indicate respectively a chi square variable with i degrees of freedom and a Fisher variable with (i, j) degrees of freedom. A threshold value F_α with significance α can be set and the zero hypothesis (2) can be tested by (3).

2.1 Storage requirements

To evaluate the DTM storage requirement of the different models, at first a numerical comparison between grids, TINs and our multi-resolution approach is here presented.

Particularly, an occupation of 64 bits (8 bytes) is hypothesized for the horizontal coordinates and the height of a point. In the

following $S(\alpha)$ is the storage space required by the model type α . To compute it, the following formulas are used.

$$S(\text{DTM}\alpha) = S(\text{metadata}_{\text{DTM}\alpha}) + S(\text{data}_{\text{DTM}\alpha})$$

To georeference a grid, the minimum needed metadata are four geographic coordinates and two integers: typically, the X (LL_X) e Y (LL_Y) coordinates of the lower left node of the grid, the spacing in X and Y between the nodes (DX and DY) and the number of the nodes in X and Y directions (N_X and N_Y) are provided. Other ways can be adopted but the number of minimally needed metadata does not change. Moreover, a field should be devoted to the conventional identifier of no-data (ND). So, the following holds

$$\begin{aligned} S(\text{metadata}_{\text{GRID}}) &= \\ S(LL_X) + S(LL_Y) + S(DX) + S(DY) + S(N_X) + S(N_Y) + S(ND) \\ &= 7 \times 64 \text{ bits} = 56 \text{ bytes} \\ S(\text{data}_{\text{GRID}}) &= N_X \times N_Y \times S(\text{height}) = N \times 8 \text{ bytes} \end{aligned}$$

As TINs are concerned, the minimal model, without additional topological information, is discussed.

$$\begin{aligned} S(\text{metadata}_{\text{TIN}}) &= S(N_V) + S(N_F) \\ S(\text{data}_{\text{TIN}}) &= S(\text{data}_{\text{NODES}}) + S(\text{data}_{\text{FACES}}) = \\ N_V \times 3 \times 64 \text{ bits} + N_F \times 3 \times \text{Ceil}[\log_2(N_V)] \text{ bits} \end{aligned}$$

N_V and N_F are the number of vertices and faces. The vertices are stored as 3D points (X , Y and height). The triangular faces are stored simply by the list of the three relevant vertices. Consider that k bits can address 2^k points: if the number of vertices is N_V , the size in bits needed for each label is $\text{Ceil}(\log_2(N_V))$, where Ceil is the rounding to the greater integer.

DTM_{MR} requires the storing of metadata relevant to the coefficients, that are needed to define the position and the resolution of each activated spline. Once defined the global interpolation domain (lower left and upper right corners), the record corresponding to a particular level is stored in the following way:

$$N_h, i_{1x}, i_{1y}, \lambda_{h,i_{1x},i_{1y}}, i_{2x}, \dots, i_{N_h x}, i_{N_h y}, \lambda_{h,i_{N_h x},i_{N_h y}}$$

where N_h is the number of activated splines in the level; i_{j_x} and i_{j_y} are the row and column indexes of the node occupied by the J -th spline; $\lambda_{h,i_{j_x},i_{j_y}}$ is the coefficient of the J -th spline.

At level h , the maximum number of active splines is $N_{\text{max}} = (2^{h+1} + 1)^2$. It is easy to show that

$\text{Ceil}(\log_2(2^{h+1} + 1)) = (h + 2)$. The storage requirements for level h , ($h = 0, \dots, M$) are:

- $2 \times (h + 2)$ bits to store the number of active splines,
- $(h + 2) \times 2 \times N_h$ bits needed to store the row and column indexes,
- $64 \times N_h$ bits needed to store the coefficients.

3. A CASE STUDY

In order to evaluate the proposed approach and compare it with the data based models, we have analyzed one case study. The data stem from a LiDAR survey of a promontory overlooking the lake of Como in Northern Italy. The horizontal spacing of the pre-processed grid is 2 m x 2 m and its vertical accuracy (Rood, 2004) is of about 20 cm.

The first step is to extract three different samples in order to simulate three dataset of sparse observations with different accuracies from which extract the relevant DTMs.

For this reason, four TINs have been extracted from the grid, with different sampling tolerances. By fixing the tolerance equal to 5m, 2m and 1m, we have created respectively the training datasets TR5, TR2 and TR1 containing scattered data (i. e. the nodes of the TIN's). By fixing the tolerance equal to 20 cm (and removing TR5, TR2 and TR1), we have finally created the test dataset TE to use for cross-validate the results. The original dataset is shown in Figure 2. In Table 1 the statistics of the datasets are reported.

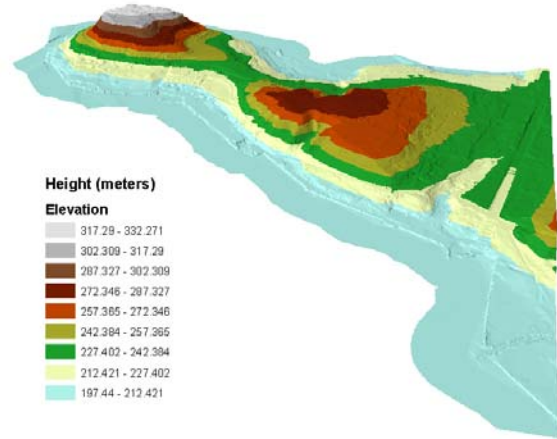


Figure 2. The original DTM

Using the three training sets as raw observations, TIN and grid models corresponding to the height accuracies of 1, 2 and 5 m have been built. By construction, the training sets directly provide the DTM_{TIN} at the different accuracy levels: in Table 2 the storage requirements of the DTM_{TIN} are reported.

To produce DTM_{GRID} , five deterministic interpolation techniques have been tested: the Inverse Distance Weighting (IDW), the 1° Order Local Polynomial (POL), the Completely Regularized Spline (CRS), the Spline with Tension (SWT) and the Thin Plate Spline (TPS). The interpolations have been computed in ArcGIS, applying the parameters automatically optimized by the software itself.

| | DTM | TR5 | TR2 | TR1 | TE |
|-------|--------|--------|--------|--------|--------|
| Count | 422610 | 3274 | 9256 | 21656 | 81869 |
| Min | 197.44 | 197.44 | 197.44 | 197.44 | 197.47 |
| Max | 332.27 | 332.27 | 332.27 | 332.27 | 332.23 |
| Mean | 225.27 | 214.33 | 225.75 | 230.81 | 235.59 |
| RMS | 27.80 | 28.58 | 30.85 | 30.36 | 27.83 |

Table 1. Statistics of the sampled datasets. Values in m.

| TR | N_V | S_V (bytes) | N_F | S_F (bytes) | S (KB) |
|----|-------|-------------|-------|-------------|--------|
| 1m | 21656 | 519744 | 41343 | 294374 | 795 |
| 2m | 9256 | 222144 | 16580 | 110430 | 325 |
| 5m | 3274 | 78576 | 4644 | 28320 | 104 |

Table 2. Characteristics of the three sampled TIN. N_V : number of vertices; S_V : storage space for vertices; N_F : number of faces; S_F : storage space for faces; S : total storage size.

If both the accuracy and the storage size of a grid have to be considered, the optimal compromise is given by the coarser grid that guarantees the desired accuracy. Therefore, different

spatial resolutions have been tested in the interpolation, starting from 10 m and decreasing it by steps of one meter. The residuals of each resulting DTM_{GRID} have been computed with respect to the starting TR and to the set of check points (TE). For each interpolation function and for each starting TR, the coarser DTM_{GRID} whose accuracy is consistent with the starting data set has been selected. In Table 3, the statistics and the characteristics of the final DTM_{GRID} are reported.

In the interpolation of TR1 and TR2, POL function does not provide consistent accuracies, even pushing the resolution of DTM_{GRID} to one meter. The other algorithms provide satisfactory results, the best being CRS that provides one meter accuracy already at the resolution of three meters. As expected, the TIN model requires less storage space than the grid model. The saving ranges from 72% to 97% for TR1, from 68% to 80% for TR2, from 59% to 67% for TR5.

| T | TR | R (m) | NX | NY | S(KB) | ResObs | | ResCheck | |
|-----|----|-------|------|------|-------|--------|----------|----------|----------|
| | | | | | | M (m) | RMSE (m) | M (m) | RMSE (m) |
| IDW | 1m | 1 | 1830 | 1784 | 25506 | -0.0 | 0.1 | -0.1 | 1.1 |
| | 2m | 5 | 366 | 357 | 1021 | -0.3 | 2.0 | -0.2 | 2.0 |
| | 5m | 10 | 183 | 179 | 256 | -0.9 | 5.0 | -0.7 | 4.8 |
| POL | 1m | 1 | 1830 | 1784 | 25506 | -0.2 | 1.9 | -0.2 | 1.5 |
| | 2m | 1 | 1830 | 1784 | 25506 | -0.4 | 3.5 | -0.4 | 2.6 |
| | 5m | 8 | 229 | 223 | 399 | -0.8 | 5.5 | -0.9 | 5.0 |
| CRS | 1m | 3 | 610 | 595 | 2836 | -0.2 | 1.3 | -0.1 | 1.1 |
| | 2m | 5 | 366 | 357 | 1021 | -0.4 | 2.4 | -0.1 | 1.8 |
| | 5m | 10 | 183 | 178 | 255 | -0.9 | 5.2 | -0.8 | 4.9 |
| SWT | 1m | 2 | 915 | 892 | 6376 | -0.2 | 1.1 | -0.1 | 1.1 |
| | 2m | 5 | 366 | 357 | 1021 | -0.4 | 2.3 | -0.3 | 2.0 |
| | 5m | 9 | 203 | 198 | 314 | -0.7 | 4.7 | -0.7 | 4.8 |
| TPS | 1m | 1 | 1830 | 1784 | 25506 | -0,10 | 0,50 | -0,03 | 1,19 |
| | 2m | 4 | 458 | 446 | 1596 | -0,44 | 2,06 | -0,10 | 2,05 |
| | 5m | 10 | 183 | 178 | 255 | -1,39 | 5,71 | -0,77 | 4,99 |

Table 3. Characteristics of the interpolated grids. T: technique used to interpolate; TR: interpolated dataset; R: final resolution; NX and NY: number of nodes in X and Y; S: storage size; ResObs, ResCheck: statistics of the residuals on the used observations and the check points; M: mean; RMSE: root mean square error.

| TR | NSplines | Number of coefficients for each level L | | | | | | | | ResObs | | ResCheck | | S (KB) |
|----|----------|---|----|----|-----|-----|-----|------|------|--------|----------|----------|----------|--------|
| | | L0 | L1 | L2 | L3 | L4 | L5 | L6 | L7 | M (m) | RMSE (m) | M (m) | RMSE (m) | |
| 1m | 5616 | 9 | 9 | 34 | 106 | 336 | 955 | 2080 | 2087 | 0.0 | 1.1 | 0.0 | 1.4 | 55 |
| 2m | 1767 | 9 | 9 | 34 | 106 | 292 | 610 | 707 | / | 0.0 | 1.8 | 0.0 | 2.2 | 17 |
| 5m | 349 | 9 | 9 | 34 | 85 | 112 | 100 | / | / | 0.0 | 3.4 | -0.4 | 4.8 | 3 |

Table 4. Statistics of DTM_{MR} . TR: interpolated dataset. NSplines: total number of activated splines. ResObs, ResCheck: statistics of the residuals on the used observations and the check points. M: mean. RMSE: root mean square error. S: storage size.

| T | TR | R(m) | NX | NY | S(KB) | ResObs | | ResCheck | |
|-----|----|------|-----|-----|-------|--------|----------|----------|----------|
| | | | | | | M (m) | RMSE (m) | M (m) | RMSE (m) |
| IDW | 1m | 3 | 610 | 595 | 2836 | -0.2 | 1.1 | -0.1 | 1.2 |
| CRS | 1m | 4 | 458 | 446 | 1596 | -0.3 | 1.9 | -0.2 | 1.4 |
| SWT | 1m | 4 | 458 | 446 | 1596 | -0.3 | 1.9 | -0.2 | 1.4 |
| TPS | 1m | 3 | 610 | 595 | 2836 | -0.2 | 1.2 | -0.0 | 1.4 |

Table 5. Characteristics of the grids interpolated from TR1 with accuracy similar to DTM_{MR} . T: technique used to interpolate; R: spatial resolution; NX and NY: number of nodes in X and Y; S: storage size; ResObs, ResCheck: statistics of the residual on the used observations and the check points; M: mean; RMSE: root mean square error.

3.1 Interpolation and storage size of the multi-resolution model

By interpolating TR1, TR2 and TR5, the coefficients of the multi-resolution splines model (DTM_{MR}) have been estimated, according to the criteria described in Sect. 2. Then, the relevant accuracies and storage sizes have been analyzed. The statistics of the DTM_{MR} s obtained by interpolating T1, T2 and T5 are summarized in Table 4. For each TR, the total number of splines, the number of splines per level, the statistics of the residuals on the used observations and on the

checkpoints are reported. Moreover the storage size in Kbytes is given.

The interpolations of TR1, TR2 and TR5 activate eight, seven and six levels of splines respectively. However, the final accuracies of $DTM_{MR}(TR1)$ and $DTM_{MR}(TR2)$ are not completely satisfactory: indeed, their RMSEs exceed one and two meters respectively and are worse than those of the respective DTM_{GRID} s. The criteria that we are adopting to automatically activate the splines ($f=1$ and $k=4$) is conservative and, in these cases, provide sub optimal results: more analyses are needed and will be discussed in a final paper. Moreover, it should be taken into account that the decimation criterion

adopted to create TR1, TR2 and TR5 from the original grid is optimized to build a TIN model and not to apply a multi-resolution interpolation. At present, an optimized criterion for the automatic activation of the splines is under study. Moreover, new decimation algorithms will be applied.

To honestly compare the storage size required by different models, they should have the same accuracy. Except for the POL interpolation, all the final DTM_{GRID}'s interpolated from TR1 are more accurate than DTM_{MR}(TR1). In the storage comparison, coarser DTM_{GRID} have been chosen, with accuracies similar to that provided by our approach. Even the final DTM_{GRID}'s interpolated from TR2 have better accuracies than DTM_{MR}(TR2) but, in this case, the coarser grids are worse. Therefore, for DTM_{MR}(TR2) and DTM_{MR}(TR5) the storage requirements are compared with the grids of Table 3, for DTM_{MR}(TR1) with the grids reported in Table 5. The comparisons, in term of storage saving, are reported in Table 6. At the end, our approach has been compared also with the Multilevel B Spline Approximation, that represents a different multi-resolution interpolation approach by bicubic splines (Lee et al., 1997). To make the comparisons (Tab. 7), the lower level of MBA whose accuracies are similar to our approach has been chosen: also the storage requirements of MBA are significantly bigger than those of our approach.

| TR | [S(DTM _{MR}) / S(DTM _{GRID})] (%) | | | | | [S(DTM _{MR}) / S(DTM _{TIN})] (%) |
|----|---|------|------|------|------|--|
| | IDW | POL | CRS | SWT | TPS | |
| 1m | 1,9% | 0,2% | 3,4% | 3,4% | 1,9% | 6,9% |
| 2m | 1,7% | 0,1% | 1,7% | 1,7% | 1,1% | 5,2% |
| 5m | 1,3% | 0,8% | 1,3% | 1,0% | 1,3% | 3,1% |

Table 6. Storage size comparisons between DTM_{MR} and DTM_{GRID}, DTM_{TIN}

| TR | ResObs(m) | | ResCheck(m) | | TotCff | S (KB) | R (%) |
|----|-----------|------|-------------|------|---------------------|--------|-------|
| | M | RMSE | M | RMSE | | | |
| 1m | 0.0 | 0.90 | -0.0 | 1.50 | 1.5 10 ⁶ | 11019 | 0.5 |
| 2m | -0.0 | 1.22 | -0.0 | 1.76 | 3.5 10 ⁵ | 2779 | 0.6 |
| 5m | 0.0 | 3.66 | 0.1 | 2.99 | 2.3 10 ⁴ | 183 | 1.6 |

Table 7. Analysis of MBA. TR: interpolated dataset. ResObs, ResCheck: MBA statistics of the residuals on the used observations and the check points. M: mean. RMSE: root mean square error. TotCff: total number of coefficients of MBA. S: MBA storage size. R: ratio between the storage requirements of our algorithm and MBA.

4. CONCLUSIONS

In this paper a new approach has been presented to interpolate and store a DTM, aimed at saving storing size without losing in accuracy with respect to classical models. Multi-resolution bilinear splines are adopted to interpolate the observations and their coefficients are stored, instead of the interpolated heights. The coefficients can then be used to reconstruct the height at any point. The model is defined analytical instead of data based.

The classical models have been compared with our approach, considering accuracy and storage requirements. An original grid has been sampled to produce three TINs with tolerances of one, two and five meters respectively. Then, the TINs vertices have been interpolated by different deterministic techniques to produce grids at different spatial resolutions and the grids have been compared with the original data. Synthetically, different interpolation techniques provide similar results and the accuracy of the grids increases with their resolution: in particular, accuracies of one, two and five meters are obtained

respectively with one, five and ten meters of spatial resolution. At present, our approach reaches an accuracy slightly worse than the accuracies provided by the finest grids: this problem is probably due to a particular interpolation choice that still needs to be deeply analyzed and optimized. However, for similar accuracies, our approach allows a significant storage saving with respect to the classical models: indeed, its storage size is about 2% of the grids size and 5% of the TINs.

The results are quite satisfactory and justify further studies finalized to define a complete scheme for the managing of the data in the server, for their transmission and for their use by the clients.

ACKNOWLEDGEMENTS

The research has been funded by the INTERREG HELI-DEM (Helvetia Italy Digital Elevation Model) project.

REFERENCES

Biagi L., Brovelli, M. A. and Zamboni, G., 2011. A DTM multi-resolution compressed model for efficient data storage and network transfer, *Proceedings of WG IV/1,2,4,5,7 ICWG IV/VIII joint Workshop on Geospatial Data Infrastructure 20-21 October 2011, Guilin, China, 2011*.

Brovelli, M. A. and Zamboni, G., 2009. Approximation of terrain heights by means of multi-resolution bilinear splines. *VII Hotine-Marussi Symposium on Mathematical Geodesy, IAG Symposia Volumes*, Springer

Christakos G, 1992. *Random Field Models for Earth Sciences*. Academic Press, New York

El-Sheimy, N., Valeo, C., Habib, A., 2005. *Digital Terrain Modeling - Acquisition, manipulation, and applications*. Artech House

Lee, S., Wolberg G. and Shin S. Y., 1997. Scattered Data Interpolation with Multilevel B-Splines, *IEEE Transactions On Visualization And Computer Graphics*, Vol. 3, No. 3

Koch K. R., 1987. *Parameter estimation and hypothesis testing in Linear models*, Springer Verlag

Li, Z., Zhu, Q., Gold, C., 2005. *Digital Terrain Modeling - Principles and methodology*. CRC Press

O' Sullivan, D. and Unwin, D., 2003. *Geographic information analysis*. John Wiley & Sons., Inc., Hoboken, New Jersey.

OGC, 2006. OpenGIS Web Map Server Implementation Specification - OGC 06-042 - Version: 1.3.0.

OGC, 2010a. OpenGIS Web Feature Service 2.0 Interface Standard - 09-025r1 and ISO/DIS 19142 - Version: 2.0.0.

OGC, 2010b. OGC WCS 2.0 Interface Standard - Core - OGC 09-110r3 - Version: 2.0.0.

Rood, M., 2004. ASPRS Guidelines Vertical Accuracy Reporting for Lidar Data. *American Society for Photogrammetry and Remote Sensing Lidar Committee (PAD)*.