Revista Brasileira de Cartografia (2017), Nº 69/8, Edição Especial "Geovisualização, mídias sociais e participação cidadã: apoio à representação, análise e gestão da paisagem": 1402-1416.
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

OPTIMIZATION, DYNAMIZATION AND VISUALIZATION IN THE USE OF LIDAR DATA IN PAMPULHA, BELO HORIZONTE: FROM DESKTOP TO SPATIAL DATABASE

Otimização, Dinamização e Visualização no Emprego de Dados LiDAR na Pampulha, Belo Horizonte: de Aplicações Desktop ao Banco de Dados Espaciais

Christian Rezende Freitas[1,2], Ana Clara Mourão Moura[2]

[1]Layer Geotecnologias Ltda.
E Av. Afonso Pena, 3924 - Belo Horizonte, MG, cep 30130-009
christianrezende@alomeioambiente.com.br

[2]Universidade Federal de Minas Gerais – UFMG
Escola de Arquitetura - Laboratório de Geoprocessamento
Rua Paraíba 697, sala 410a, Savassi. Belo Horizonte, MG, cep 30130-140
anaclara@ufmg.br

Received on November, 2016/ Accepted on May 7, 2017
Recebido em 9 de Novembro, 2017/ Aceito em 7 de Maio, 2017

RESUMO

Os usuários de geoinformação têm tido a oportunidade de trabalhar com expressiva quantidade e relativa qualidade de dados, mas isto exige o manuseio de coleções robustas e complexas. Ao mesmo tempo, a exigência por rápidas respostas e análises remete à cartografia dinâmica, que precisa da agilidade de processos para a imediata atualização de resultados de análises. Juntamente com a ampliação da qualidade do dado e a necessidade de rápida produção de informação aumenta-se a utilização das informações espaciais relacionadas às questões legais. Essa nova demanda exige uma evolução a partir da utilização das Infraestruturas de Dados Espaciais (IDE) para uma base de acesso a serviços, que transformam dado em informação, favorecendo a produção de conhecimento. É para este cenário que o artigo tem como objetivo propor uma metodologia de trabalho segundo a lógica de banco de dados e não apenas em ambiente de Sistema de Informações Geográficas. Apresenta contribuições em termos de reflexões críticas e propõe uma metodologia de manuseio de dados de ortofotos de alta resolução e de informações altimétricas de captura LiDAR. Apresenta o estudo de caso da modelagem para análise de volumetria de vegetação e volumetria de edificações na regional Pampulha, Belo Horizonte, Minas Gerais. O foco é demonstrar para usuários de aplicativos *desktop* por softwares de SIGs, que hoje são maioria entre usuários de tecnologias de geoinformação, que é possível mudar para a lógica de construção de análises espaciais no formato de banco de dados, em aplicativos de livre acesso e que representam ganhos significativos no manuseio de grandes quantidades de dados e de necessidade de aplicação de rotinas de trabalho. Nesse sentido, é um artigo de interface entre a era da tecnologia de informação desktop representada pelos grandes softwares comerciais e a era de emprego de lógicas de *ETL* e banco de dados. A proposta é de emprego da lógica de banco de dados espaciais e ferramentas de *ETL* (*extração, transformação e carga*) para ganho de desempenho em termos de tempo, emprego de equipe técnica e segurança da informação. Comprova numericamente o ganho de performance e contextualiza o processo em relação ao estado-da-arte das tecnologias de geoinformação.

Palavras chaves: Novos Parâmetros Urbanísticos, Suporte à Tomada de Opiniões, Banco de Dados Espacial, Ferramentas de *ETL*, Dados LiDAR.

ABSTRACT

The users of geoinformation have been given the opportunity to work with impressive amount of data with some good quality, but it demands knowledge to handle complex and robust collection of information. At the same time the demand for quick answers and analysis requires dynamic cartography capable of execute processes with agility for immediate update of outputs. Together with the improvement of the data quality and the need of fast information production the legal issues is also developed, progressing from the first steps of the Spatial Data Infrastructure to a services access, via internet, which change data into information, favoring the knowledge production. It is for this scenario that the paper presents as objective to propose a methodology based on database and not only on Geographic Information Systems. It represents contribution as critical reflections and introduction of a methodology for handling high resolution information extracted from orthophotos and altimetry from LiDAR data. It presents the case study of the modeling to assess vegetation and buildings´ volumes in Pampulha region, in the city of Belo Horizonte, Minas Gerais State. The focus is to demonstrate to users of desktop applications based on GIS software, that nowadays are the majority among users of geoinformation technologies, that it is possible to change to the construction of spatial analyzes using the logic of database, in applications of free access, and with significant gains in the handling of large amounts of data and the need to apply routines. In this sense, it is an interface between the era of desktop information technology represented by large commercial software and the era of *ETL* and database logics. The use of spatial geodatabase logics and *ETL* tools (extract, transform and load) are proposed and tested to improve the performance in terms of time, technical team demand and information security. It numerically proves the performance improvement and contextualizes the process regarding the state-of-art of geoinformation technologies.

Keywords: New Urban Parameters, Support to Opinion Making, Spatial Geodatabase, *ETL* Tools, LiDAR Data.

1. INTRODUCTION

The digital environment representation is living now one more significant change which demands new manners to handle geographical information. The generic term that best expresses this moment is "big data" but we can add to the expression the terms "crowdsourcing", meaning not only the availability and need of handling a large amount of data as well as the expressive access and participation of many and different groups of users. Although the term is related to the almost "anarchical" data collection, it is used the concept to define the access to a great amount, speed and variety of data which require the user has the capacity to use the specific technological resources and analytic methods for the data extraction and for composing information from data.

The evolution of the representation stages was based on the preparation of the georeferenced geometric primitives (CAD/CAM) for the combination of cartographic with alphanumeric data (Desktop Mapping). The next step was the improvement of availability of spatial assessment tools (GIS), for the specific use of such tools in specialists' issues and modeling (Expert Information Systems). The present stage is characterized by a great change provided by the advent of the Web 2.0 and the WebGis possibilities (Figure 1).

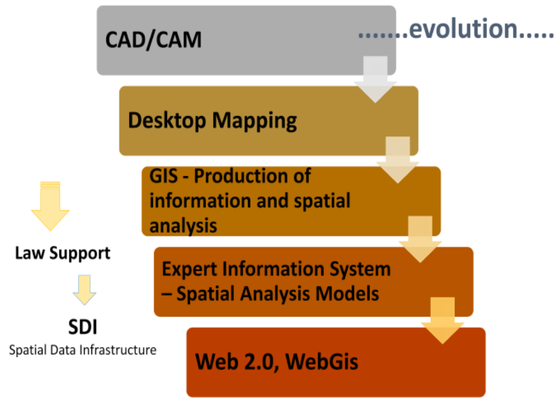


Fig. 1 – The evolution of technologies in geoinformation. Adapted from Moura (2013).

As an example, we can present the drawing of a road (CAD), followed by the evolution of road representation linked with tables with attributes (Desktop), followed by related tables linked with roads but with the possibility to analyze capillarity and accessibility (GIS), followed by the capacity of predicting the best areas to construct car-sharing facilities (Expert and Models), followed by the possibilities of using crowdsourcing information to receive data from traffic and to create a network of users of car-sharing.

In parallel to this, we can highlight the development of the legal support, with rules that favor the wide spreading and access the georeferenced data.

The improvement of data access and the consequent possibility of using the georeferenced information arose many challenges to geoinformation professionals. The image presented by Jorge Luís Borges (1899-1986) in the tale "The Aleph", from 1949, named "the two kings and the two labyrinths" has never been so present and adequate. The tale tells the story of the dispute of two kings, being the desert one king´s domain, related to the idea of lack the information, and the other king´s domain is the labyrinth which is related to the idea of the information excess. The tale presents the dilemma of the information excess being as problematic as the lack of information. And it can be observed in the present when the advent of the web 2.0 favors the establishment of the webs and connections that completely change the information media, ways and producers.

However, it is necessary to make clear the difference between data, information and knowledge. Some authors have introduced the concepts in a really good manner, among them we quote Xavier-da-Silva (2001), who defends that information generation can result in a knowledge gain, but it's not true that all the data information and all information can result in knowledge. We also mention Steinitz (2012) who presents the spatial assessment by the Geodesign framework (Figure 2).

According to Geodesign framework (Steinitz, 2012), reality decoding is made by models that facilitate the understanding of how the area can be described based on data collection (representation models); on how the area operates, what is produced changing data into spatially distributed information (process models); and on the understanding about how well the area operates (evaluation models), this third step producing knowledge about the area from data and information detailed by previous stages.

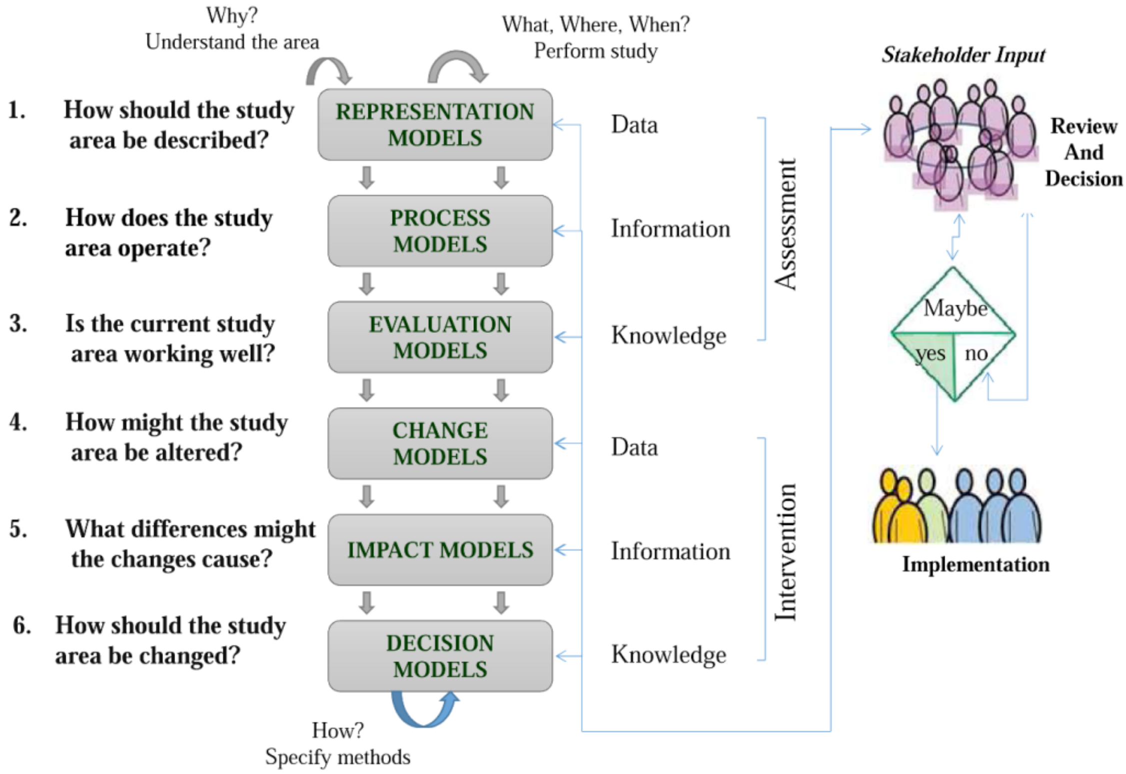


Fig. 2 - The framework to Geodesign proposed by Steinitz. Source: Steinitz (2012, p. 25)

Geodesign methodology goes further that, it is not limited to interpretation of the area conditions, but mainly about proposing alternative futures. In these new stages, the determination of where such changes are to be made (change models) dictate the creation of data. Information is produced as consequences of such changes (impact models) and, finally, knowledge is obtained regarding alternative futures for an area (decision models). In this sense, geoinformation technologies widely facilitate the evolution of data changed into information and transformed into knowledge, but this last one understood as knowledge acquirement.

The expressive production, access and handling of data, added to the rising interest of new users that are expanding numerically and in different profiles in the use of geographical data, require better conditions of the mechanisms to change data into information and changing information into knowledge.

Regarding the change of information into knowledge, there are researches applications that have been established for visualization of geographical information, aiming the promotion of shared information codes to help different kind of users to better understand spatial reality and to feel themselves invited to take part in common decisions, speaking and interfering in spatial proposals and management. The idea is, according to McCormick et al. (1987), to “see the unseen”, to achieve the new understanding and ideas. Regarding visualization, there can be quoted the works of MacEachren at al. (2004), Manovich (2004), Kingston (2007), Andrienko at al. (2011), Pensa et al. (2013) and Pensa and Masala (2016), among others.

Visualization has an expressive impact on transforming data into information and in the support to opinions making and in decision making, when agents involved with the shared decision on the alternative future of a land are analyzing and understanding possibilities.

Before the stage of information and knowledge, there´s the stage of: production and handling the data to support the diagnostic, prognostic and propositional analysis. The production of data change, involving citizens as producers and sensors of reality, as defined by Davis Jr. et.al (2016) who approaches the issues of crowd sourcing and crowd sensing, resulting from collective actions, also defined by Onsrud at al., 2004; Goodchild, 2007; Maguire, 2007; Bao et al., 2015, among others. The authors report the social media potentials on the behavioral data production. It is the big data world, but it requires proper manners to receive the data to properly change it into information and information into knowledge.

In addition to the data production, there is also significant improvement of the spreading and wide access to the data, a process that is developed from the existence of legislation which allows, manages and promotes that information is properly and efficiently shared. It is about the attention on SDI – Spatial Data Infrastructure, in Brazil named IDE - *Infraestrutura de Dados Espaciais*. According to Campagna (2016), IDE is a data network organized to facilitate and share spatial information. In Europe, the Directive 2007/02/EC established the *INfrastructure for SPatial InfoRmation in Europe* (*INSPIRE*) which regulates the access to public areas spatial data and to the services regarding 34 topics selected as most significant to the environmental preservation and, consequently, on the urban and territorial planning.

In Brazil, the process was started after the Decree 6666/2008 which establishes the National Infrastructure of Spatial Data. From this initiative, the city and state governments have also promoted initiatives to provide access to the geographical information organized within the respective fields of action.

The IDE action brings along the possibility of wide access to large collection of data, but also requires the processes of changing structured data into information. It is the implementation of ‘services’ combined with data, and not only data access. The services are structured as software, infrastructure and platforms. The use of database based on services is increased by the possibility of working in a cloud for activities and for visualization. The production of data requires a new manner to work with services, as defended by Agrawal et al. (2009; 2010). But even though the network media is not used, the access to the large collection of data demands to review and rethink the way information is been produced, and specially how to include the logic of database, based on services.

Explaining the process in simple words, most of the users of geoinformation technologies used to work with tools, such as algorithms and applications provided by the software., what means that the steps to work with data are more conditioned to the tools that exists in a software, so that the user thinks about how to work with the data according to the applications provided by the software. For example, if the goal is to transform from points to calculate the volume of the vegetation mass, using GIS tools the users must follow many steps, as to convert from points to multipoints, to select the points that correspond to footprints of trees, to construct a volume and calculate its values, and so one.

All these steps produce partial results that are stored in the computer. In the software, to execute more complex procedures as a process chain, the available resource is to create models which manage steps and connect available tools. Once the model is created, the process is triggered leading to the expected result. It is usual that many secondary data are generated and stored as partial results from each step of the process, and it is also usual that many of the steps go through actions that require visual recovery, which is quite complex once the data collection is too large.

The focus is to demonstrate to users of desktop applications based on GIS software, that nowadays are the majority among users of geoinformation technologies, that it is possible to change to the construction of spatial analyzes using the logic of database, in applications of free access, and with significant gains in the handling of large amounts of data and the need to apply routines. In this sense, it is an interface between the era of desktop information technology represented by large commercial software and the era of ETL and database logics.

This paper contribution is presented for that matter. To optimize the use of data generated by the LiDAR capture (Light Detection and Ranging) related to topographic and surface information, and the creation of DTM (Digital Terrain Modeling) and DST (Digital Surface Modeling), also known as DSMn (Digital Surface Model Normalized). The target is more specifically data related to the difference of topography and surface, the altimetry information of urban vegetation cover masses.

The Municipal Government of Belo Horizonte has performed a great effort on producing data and information regarding the city land use and occupation, mostly three-dimension data. The first data acquisition was executed in 2007, and a new acquisition was held in 2015. After an agreement signed by UFMG and Prodabel/PBH it was possible to access data composed by orthophotos of 20 cm resolution containing infrared and visible ranges, and the LiDAR captured in irregular distribution of around 1 pt/m2. In addition to that recently acquired data, there were also provided the Prodabel/PBH vector data related to blocks, lots and buildings projections.

The purpose is to start an investigation to give support to the proposal vegetation cover parameters in urban areas that can be applied to municipal master plans. In Brazil, the thematic of 'ambience' and environment quality are, generally, treated as secondary and not the main matter by the land use and occupation regulations. Ambience concerns the quality of the place, resulting from objective values (climate, temperature, light), but also subjective values related to well-being, all of them quite related to vegetation cover.

The existing parameters regarding such requirements in Brazil are just the ones that define soil permeability rates, and the buildings standards code establishes, in some cases, required trees planting for the sidewalks. The land use and occupation regulations establish tables of urban parameters to specify the volume of the buildings defined by morphometric references for setbacks, FAR – floor area ratio, volumetric coefficient and, in some cases, the maximum height.

To characterize the distribution of the urban vegetation cover, using the LiDAR data and high resolution orthophotos, was decided to load the data into a spatial database capable of relate the information regarding the vegetation height to its horizontal projection using simple *SQL* scripts without creating additional data, especially considering the amount of information necessary for the analysis. The horizontal projection of the vegetation cover surfaces was obtained by digital processing of orthophotos and the calculation of NDVI - *normalized difference vegetation index* (Rocha et al., 2016) and, once surfaces were obtained, it was necessary to use database tools to calculate their respective height, the combination of cartography and alphanumeric data, to produce necessary information for the vegetation cover volume calculation, and finally to indicate the ration per block.

These procedures of optimization in information production are absolutely necessary for urban planning and municipal management that use a large collection of data to create scenarios of the existing reality in a short period of time. The use of automatic services facilitates the test of many simulations, to make more than one combination and to avoid having limited purposes due to difficulties presented when data collection is too large.

The case study was developed for

Pampulha region, in Belo Horizonte. The coverage area has 51,034,800 square meters (51 km2), composed by 37 districts, 1946 blocks, 41360 lots, 95158 buildings (data provided by Prodabel/PBH), 224,376 people live there (IBGE, Censo 2010). The regional area is characterized by significant vegetation cover, what motivated the research on the respective conditions to be used as assessment and comparison factor regarding other areas of the city (Figure 3 and Figure 4).

The data set required for the calculation of the vegetation volume, composed by LiDAR data and orthophotos of high resolution (spatial and spectral), which resulted in Digital Surface Modeling and NDVI maps, represent a wide universe of geographical information to be processed, aiming the identification of the vegetation volume in each block. LiDAR data presents over 270 million points about altimetry information, while the vegetation cover spots provided by NDVI map are over 1.5 million.

The first phase of the information organization for processing was to divide the total amount of LiDAR data and NDVI results in a grid. The output is a set of areas of shape files, able to be edited with no need of using dedicated machines, and counting on processors and memory enough to assess the total volume of information to be used for the analysis (Figure 5).

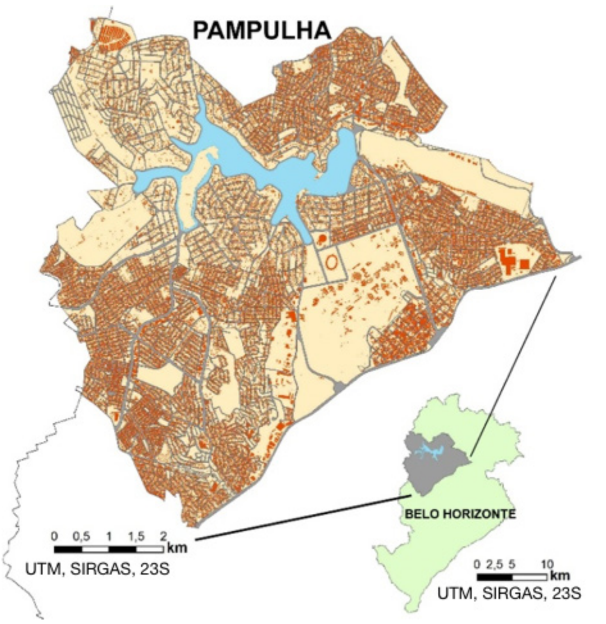


Fig. 3 – Pampulha region in Belo Horizonte, MG.

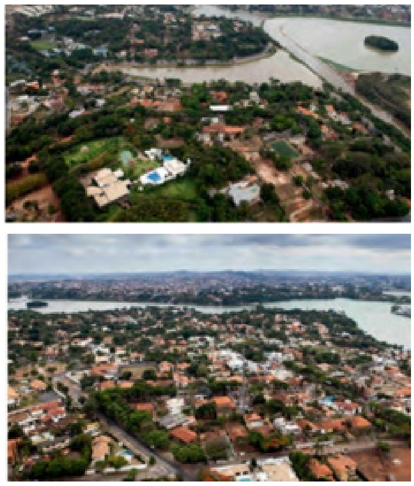


Fig. 4 – Images of Pampulha region. Source: Pampulha Modern Complex Application File for Inclusion to the List of World Heritage, 2015.

The data was divided into 180 sets (the number of cells created by the grid analysis) of LiDAR and NDVI Shapefiles, making 360 files. Therefore, it was adopted a specific tool to organize and process the information to provide the result in a short time and involving less effort regarding the human intervention. The process was then started by the preparation of logics based on database and services.

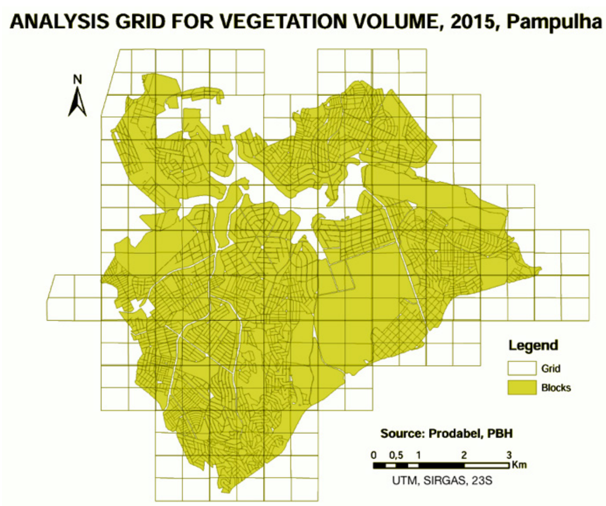


Fig. 5 – Grid to develop the analysis, demonstrating the necessity to cut the area in many parts.

The contribution of the paper herein is the presentation of a methodological process based on the use of geographical database for the modeling of services, for the optimization to change data into information. The case study is intended to map and characterize the urban vegetation cover, with the use of LiDAR capture data and high resolution orthophotos, to encourage the discussion about new parameters in urban planning related to ambience produced by vegetation cover.

2. METHODOLOGY

The total volume of data to process was 360 files divided into altimetry points from laser survey (LiDAR) and vegetation spots created from a NVDI map, a product of high resolution images. The division of the set into grid cells aimed to enable the information base processing in the mapping desktop software.

The division into smaller files means slight gain from the perspective of time and effort involved and imposed to the technical user. Even representing a great effort to attempt processing smaller files, data fragmentation requires more time and attention by the technician during the processing of information in a continuous manner, assuring timely and reliable data input.

From such reality, the use of tools that enable process automation represents a gain as it reduces the need of technical intervention during the analysis, being the main concern to make the right information flow during the desired analysis. In this case study, a flowchart (Figure 6) is established for the use of two analysis tools in the process.

After dividing the information into grid cells, such cells were loaded in a geographical database using the *ETL* tool. Having the data loaded to the database, once again the *ETL* tool was used to organize and sequence the spatial analysis performed by the database, aiming to produce the required information. At the end, once again the *ETL* tool is used to extract obtained information in Desktop Mapping software standard format.

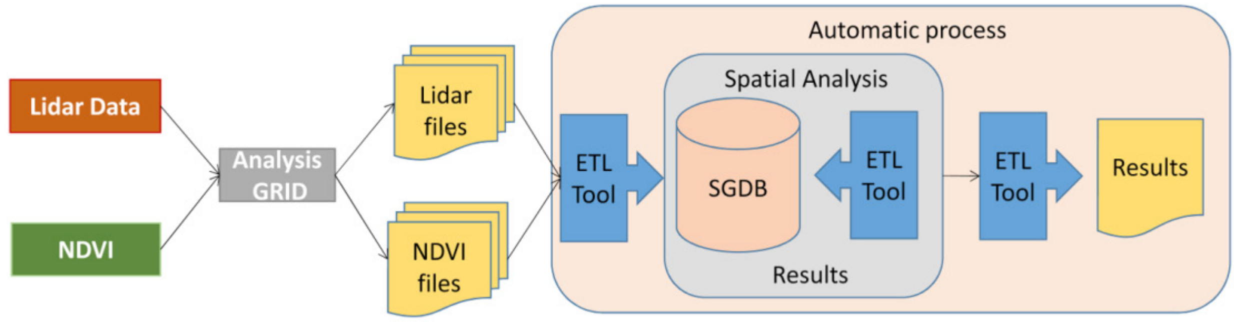


Fig. 6 – Developed processes flowchart.

3. DEVELOPMENT AND RESULTS

Considering the article is about methodological contribution, its development means the presentation of the steps undertaken and the obtained partial and final results.

The first application used was the software *Geokettle*. This free domain tool is an *ETL – Extract, Transform and Load*, able to process large volumes of data in a continuous way, providing a large set of tools to change and transform information, being information whether geographical or not, in a model builder style interface. The main task of the software in the analysis was to perform the data loading in the Geographical Database, and the ordering of the volume calculation in interface with the database for the set of 360 geographical files used.

Together with the *ETL Geokettle* software, the database *PostgresSQL/PostGIS* was used as a data storage and geographic analysis tool. The advantage of transferring geographical data from a desktop environment to a database environment is about facilitating the creation of processing which makes unnecessary the user´s intervention along the several phases of the vegetation volume calculation. The databases can store and process geographical information, providing several spatial functions for analysis and recovery of information based on the spatial characteristics. Currently the *PostgresSQL*, with the extension *PostGIS*, is the main open source geographical database being used.

3.1 Data Loading

After dividing the data into 360 files, according to the analysis grid, the files were loaded in the geographical database. In the database, the files have been kept separated, that is, each shapefile was stored in the database in its respective table, with its geographic spatial reference.

An *ETL* tool was used to execute this task. For the LIDAR data, as well as for the vegetation spots obtained by NDVI, the loading consisted of, firstly, the creation of a "txt" type file, containing the list of shapefiles related to LiDAR and NDVI issues (Figure 7).


listveg.txt - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
shape
D:\veget_topo\4556_4_vegtop.shp
D:\veget_topo\4654_1_vegtop.shp
D:\veget_topo\4654_4_vegtop.shp
D:\veget_topo\4655_1_vegtop.shp
D:\veget_topo\4655_4_vegtop.shp
D:\veget_topo\4656_1_vegtop.shp
D:\veget_topo\4656_2_vegtop.shp
D:\veget_topo\4656_3_vegtop.shp
D:\veget_topo\4657_1_vegtop.shp
D:\veget_topo\4657_2_vegtop.shp
D:\veget_topo\4657_3_vegtop.shp
D:\veget_topo\4658_1_vegtop.shp
D:\veget_topo\4658_4_vegtop.shp
D:\veget_topo\4659_1_vegtop.shp
D:\veget_topo\4659_2_vegtop.shp
D:\veget_topo\4659_3_vegtop.shp
D:\veget_topo\4659_4_vegtop.shp
D:\veget_topo\4660_1_vegtop.shp
D:\veget_topo\4660_2_vegtop.shp
D:\veget_topo\4660_3_vegtop.shp
D:\veget_topo\4660_4_vegtop.shp
D:\veget_topo\4661_1_vegtop.shp
D:\veget_topo\4752_1_vegtop.shp

Fig. 7 - Example list of vegetation shapefile files. Should be made a similar list for shapefiles points LiDAR.

The figure 8 shows an example of tables structures in the database using OMT-G data model documentation.

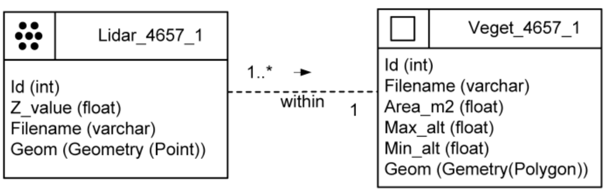


Fig. 8 – Example of tables in OMT-G data model.

From the files list, it was set a flow in the *ETL* software which performs a definitive data load into the geographical database. For this, the first phase consists of loading the list with the shapefiles, and it is defined the name of the table to which they belong to into the database. This phase consists of 3 different steps, as illustrated in the flowchart (Figure 9).

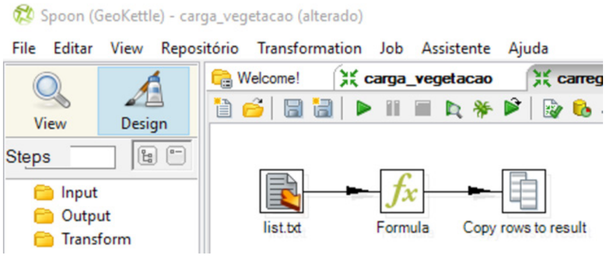
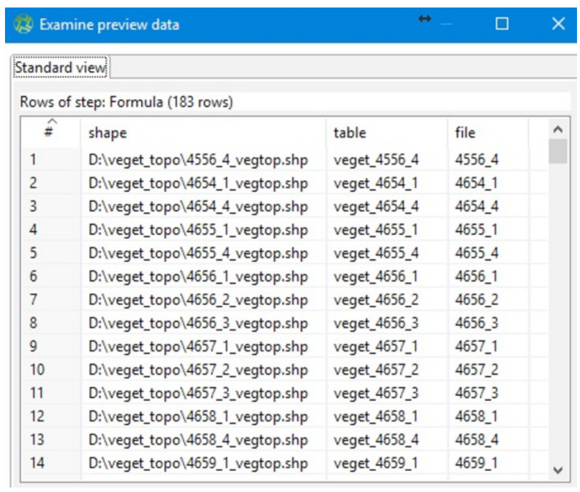


Fig. 9 – Phase 1 of the data loading flow into the database.

The first step (*list.txt*) loads the list into the software. The second step (*Formula*) generates the field with the name of the geographical table that stores the data in the respective shapefile (Figure 10) type file. The third step is responsible for the storage of the results of the first step in the memory, to be used in the phase 2 of the loading process.

In this second phase, there are five steps. The first step (*Get rows from result*) is responsible to recover information created in the first phase. The second step (*Shapefile File Input*) runs the reading of the shapefiles, using for that the file path described under the "shape" column (Figure 10).

The third step (*Select values*) is used to execute the changes of the existing columns in the shapefiles, such as deleting undesired columns and/or renaming the columns. The fourth step (*Set SRS*) is responsible for the definition of the geographical projection for the spatial data, in this case the UTM WGS84 Zone 23S (EPSG:32723). At last, the fifth step (*Table output*) is responsible for transferring spatial information into the *Postgres/PostGIS* database, into the respective tables using the column "table" (Figure 10).


Examine preview data
Standard view
Rows of step: Formula (183 rows)

#	shape	table	file
1	D:\veget_topo\4556_4_vegtop.shp	veget_4556_4	4556_4
2	D:\veget_topo\4654_1_vegtop.shp	veget_4654_1	4654_1
3	D:\veget_topo\4654_4_vegtop.shp	veget_4654_4	4654_4
4	D:\veget_topo\4655_1_vegtop.shp	veget_4655_1	4655_1
5	D:\veget_topo\4655_4_vegtop.shp	veget_4655_4	4655_4
6	D:\veget_topo\4656_1_vegtop.shp	veget_4656_1	4656_1
7	D:\veget_topo\4656_2_vegtop.shp	veget_4656_2	4656_2
8	D:\veget_topo\4656_3_vegtop.shp	veget_4656_3	4656_3
9	D:\veget_topo\4657_1_vegtop.shp	veget_4657_1	4657_1
10	D:\veget_topo\4657_2_vegtop.shp	veget_4657_2	4657_2
11	D:\veget_topo\4657_3_vegtop.shp	veget_4657_3	4657_3
12	D:\veget_topo\4658_1_vegtop.shp	veget_4658_1	4658_1
13	D:\veget_topo\4658_4_vegtop.shp	veget_4658_4	4658_4
14	D:\veget_topo\4659_1_vegtop.shp	veget_4659_1	4659_1

Fig. 10 – Results of the "Formula" step – table example containing the path of the shapefile file and its respective table name in the database.

Once loaded the files list, the phase 2 of the loading process is responsible for the definitive loading of information from the shapefiles into the geographical database (Figure 11).

A similar process was made to load data related to LiDAR files and the file of Pampulha region blocks. At the end, it was set the Database with the tables related to be used to calculate the volume of the vegetation (Figure 12).

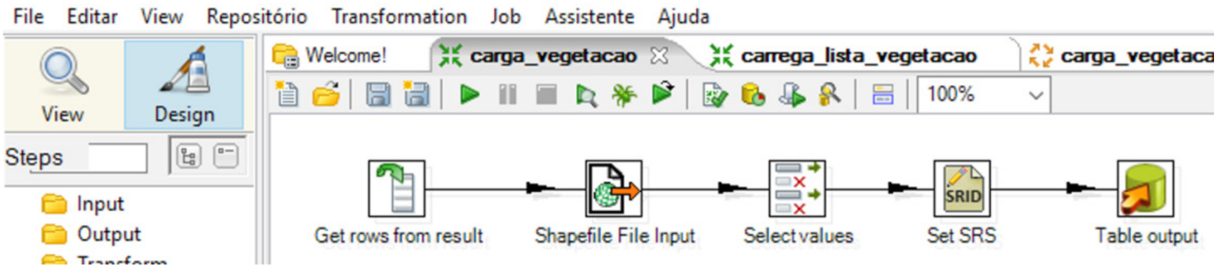


Fig. 11 – Phase 2 of the data flow loading into the database.

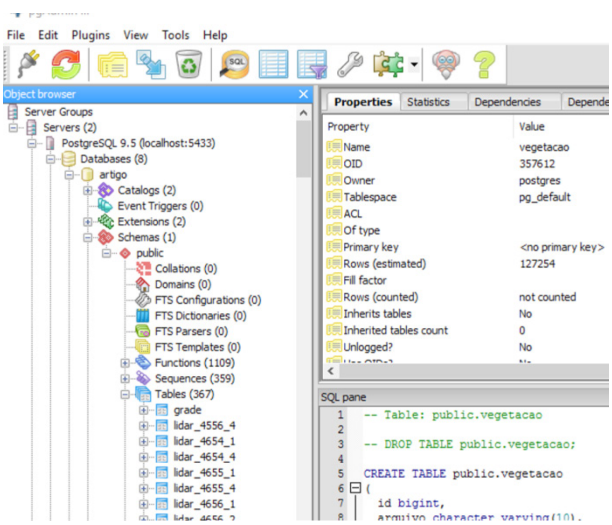


Fig. 12 - Geographic Database.

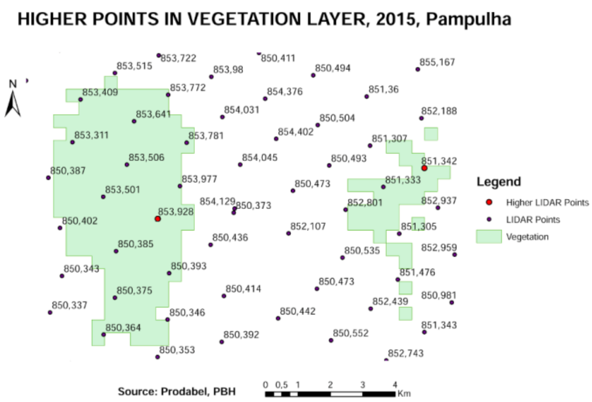


Fig. 13 – Identification of maximum height points in vegetation spots. In green vegetation plots, inside them points with elevation data, in red the highest elevation that is chosen by the system.

3.2 Calculation of the Maximum Height

For the calculation of vegetation volumes, it was necessary to identify, for each vegetation spot, the highest value among all the points from the LiDAR that were internal to the spot polygon (Figure 13).

Together with the base points values, already calculated for the spots by information collection on the DTM (*Digital Terrain Modeling*, generated by the topography altimetry points), it was calculated the distance between the highest point and the topography base point of each spot, resulting in the calculation of the spot height. The height, together with the value of the projected area of the polygon, enables the calculation of the vegetation volume.

For the identification of the maximum height it was executed an *SQL* script, to store the maximum value in the vegetation table. With the blank field duly created, it was executed an *ETL* process responsible for running the script for each of the 180 tables of vegetation spots existing in the database (Figure 14).

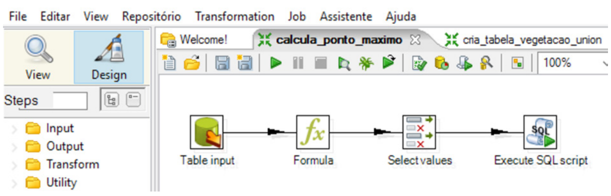
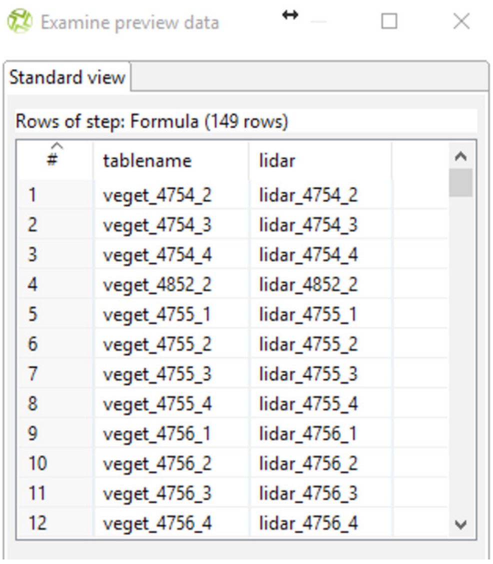


Fig. 14 – *ETL* flow for Maximum Point

For this procedure, the first step (*Table input*) extracts the existing vegetation tables from the list, using a simple *SQL* script (Figure 15):

Fig. 15 – Vegetation tables list script.

The second step of the procedure (*Formula*) uses the information of the vegetation table name and creates the LiDAR respective table name (Figure 16).



Examine preview data

Standard view

Rows of step: Formula (149 rows)

#	tablename	lidar
1	veget_4754_2	lidar_4754_2
2	veget_4754_3	lidar_4754_3
3	veget_4754_4	lidar_4754_4
4	veget_4852_2	lidar_4852_2
5	veget_4755_1	lidar_4755_1
6	veget_4755_2	lidar_4755_2
7	veget_4755_3	lidar_4755_3
8	veget_4755_4	lidar_4755_4
9	veget_4756_1	lidar_4756_1
10	veget_4756_2	lidar_4756_2
11	veget_4756_3	lidar_4756_3
12	veget_4756_4	lidar_4756_4

Fig. 16 - Vegetation tables list with respective LIDAR table

The third step (*Select values*) consists of doubling the column with the name of the vegetation table, aiming to execute the script of identification of maximum height. The final step of the process (*Execute SQL script*) is composed by the script that enables the identification of the maximum point (*SQL description*). In this script, the question marks '?' are arguments and shall be replaced by the values existing in the table resulting from the previous step (*Select values*).

The *SQL* script for selection of maximum point is described by (Frame 1):

Frame 1: Script to select maximum point

```
1   update '?' as v
2     set max_alt = z.max_alt
3      from (select
4            ve.id,
5             max(l.z_value) as max_alt
6             from '?' as ve, '?' as l
7           where ve.area_m2 >=2.25 and
8             ST_Intersects(ve.geom,l.geom)=true
      and
9             ST_Within(l.geom,ve.geom)=true
10         group by ve.id) z
11     where v.id=z.id
```

Lines 1 and 2 indicate the script updates the column "max_alt" existing in each vegetation table. The symbol '?' is to be replaced by the name of the table provided by the previous step.

Line 3 provides a 'select' or research, which creates a table with the maximum point calculation. In the line 4 it is listed an 'id' column from the vegetation table, named in line 6. Attention is required for the fact that the vegetation table is named twice, in lines 1 and 6, and for that reason the column containing the name of the table is double using the step previously described (*Select values*).

Line 5 calls the second column of the 'select', using the join function 'max' to the 'z_value' from the LiDAR table. Line 6 highlights the tables to be used by the command 'from' the vegetation table with the alias "v", and the table LiDAR has the alias "l". The comma separating the two tables indicates the execution of a "cross join". For the identification of which of the LIDAR points are input into the vegetation spots it´s made necessary to cross join all the surfaces of both tables.

Lines 7, 8 and 9 aim to define the conditions so the database executes the joining of LiDAR points and vegetation spots. Line 7 highlights the condition to join only for the spots presenting area over or equal to 2.25m², eliminating, then, the small spots, for those are considered insignificant for the analysis in issue.

Line 8 calls the intersection spatial function "ST_Intersects( )" that compares the spots geometry "ve.geom" to the LiDAR points geometry "l.geom", requesting as only result the cross point where the intersection of the two geometries is, that is, "true". This type of function is executed in first place to improve the speed of the database spatial process using internal spatial index.

Line 9 makes sure will be cross joined only the records whose LiDAR points are input, "ST_Whitin( )=true" , in the respective vegetation spot.

At the end of the 'cross join' there is the repetition of spots, as usually there are more than one LiDAR point input into one same vegetation polygon. For that reason, the function 'Group by' is input to the result of the cross, using as reference the code 've.id' of the spot. In the grouping process, that is executed in line 5, it is used the 'max' joining function that makes sure that for that spot only the higher value point "l.z_value" is kept, as result of the 'select' run in line 3, having the alias 'z'.

To transfer the value between columns "z.max_alt" and "max_alt" described in line 2, it´s necessary to add the function 'where' shown in line 11, enabling the join of the table 'z' created by 'select' and the table called in line 1 of *SQL* command using the 'id' fields of both screens.

The *ETL* flow is run to execute the *SQL* script for the each one of the 180 tables of vegetation existing in the database, with no need of technician intervention.

3.3 Grouping vegetation volume per block

After calculating the maximum heights for the vegetation spots, those are transferred to one table in the database to facilitate the analysis of blocks grouping.

The figure 17 shows de structure of the tables used to make the final calculation of vegetation volume per block in Pampulha region.

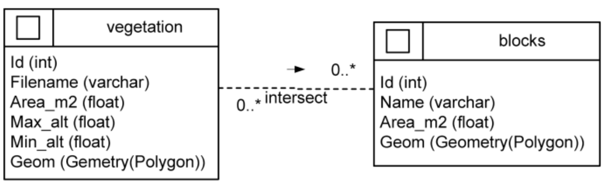


Fig. 17 – Table structures of vegetation and blocks in database.

This analysis was executed because the purpose of the study was to compare the built volume to the vegetation volume per block, as mentioned in the beginning of this paper. Having deleted the spots smaller than 2.25m², there are 127.254 remaining polygons related to significant vegetation. To produce results by block, data had to be grouped, and it was executed an *SQL* script that crossed the vegetation table and the blocks table, both loaded into the database. The script is (Frame 2):

Frame 2: Script to associate vegetation and blocks

```
1   Select
2   b.id as block_id,
3   b.name as region,
4   b.area_m2  as block_area,
5   sum(ST_Area(ST_intersection(v.geom,b.geom
6   ))) as aream2_veg,
7   sum(ST_Area(ST_intersection(v.geom,b.geom)
8   )* (v.max_alt - v.min_alt)) as volume_veg,
9   b.geom
10  from blocks as b, vegetation as v
11  where ST_intersects(b.geom,v.geom)=true
12  group by 1,2,3,9
```

Lines 1 to 4 start the 'select' calling the columns existing in the blocks table. Lines 5 and 6 presented the value of the area of intersection between vegetation geometry "v.geom" and the respective block geometry 'b.geom'. The first function called 'ST_Intersection( )' created the geometry resulting from the intersection of the two initial geometries. The function 'ST_Area( )' calculates the area of the intersection geometry. Line 6 creates columns 'aream2_veg', containing the vegetation area value input into the respective block.

Lines 7 and 8 use the same functions of intersection and area for the calculation of the vegetation volumes using also the columns 'v.max_alt' and "v.min_alt", from the vegetation table.

Line 9 calls the blocks original geometry column in which the values are presented in form of thematic maps.

Line 10 calls the blocks and vegetation tables running a 'cross join' crossing all the elements from the original tables. Line 11 makes sure the cross is only for the records whose geometries are intercepted "ST_Intersects( )=true".

In line 12 the function 'group by' was used to aggregate the values for each block in the region.

At the end, the data from the database were extracted in a shapefile format by the *ETL* tool to distribute the data among students and The Municipal Government of Belo Horizonte (Figure 18).

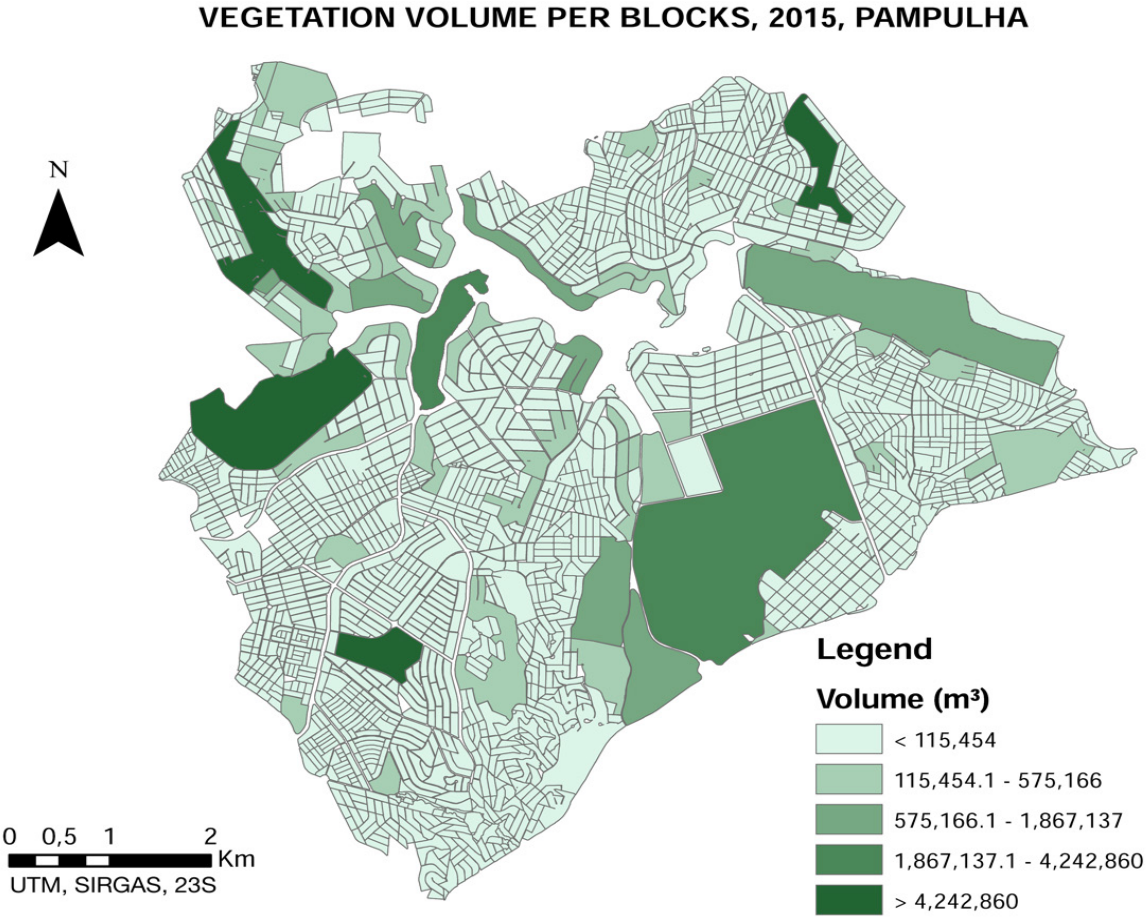


Fig. 18 – Vegetation volume per block.

4. DISCUSSIONS AND CONCLUSIONS

Now, the use of new tools is focused, mostly, on actions aiming the alphanumeric information type. It is noticed these logics, when duly adopted by the professionals working in the geoprocessing area, can represent gains not only regarding the geoprocessing capability, but also in the manner such spatial analysis tools can be used for the development of dynamic processes.

In the case study presented, even fragmenting the information from LiDAR data, they present over 1.5 million of altimetry information points. Undertaking spatial analysis from such large base is almost impossible using desktop software of free domain which, mostly, are limited regarding the total volume of data to be processed.

The alternative for the professionals, in such cases, is to use commercial software which financial cost may get to prevent the projects development or are simply out of reach for the most part of the users. Even though, the work is extremely slow and demands the action of the technician in control and in the management of all the steps of the process.

The two tools used in this case, *ETL* and Geographical Database, are presented in both versions: paid and free domain ones, but we used just the free one. The free software used were *Geokettle* (*ETL*) and *Postgres/PostGIS* (database). The differences between the market solutions for such tools are occasional, enabling the professionals to access a powerful set of software capable of using the most different spatial analysis.

The use of such kind of solution for the spatial processes is capable to change the evolution process of the spatial analysis tool, for being able to replace a different set of geoinformation tools, especially those related to the Geographic Information System and to the Specialist Systems. Within this perspective, it is proposed a new approach on the technologies of geoinformation development (Figure 19).

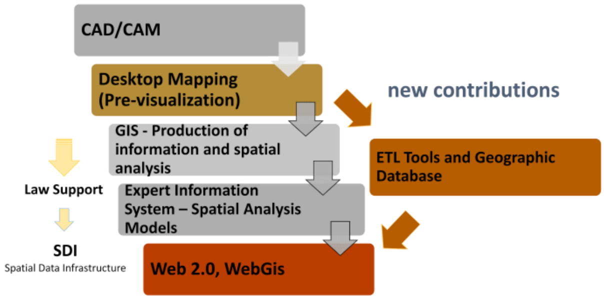


Figure 19 – Table of current development of technologies of geoinformation

By this new approach, the mapping desktop tools are intended to the pre-visualization of information, as for the post-visualization, aiming the cartography production and the composition of communication mechanisms among the users. With the advent of the Web 2.0, even the final layouts of the classical digital cartography presentation are becoming obsolete.

The gain of incorporating the geographic database into the spatial analysis is mostly related to the automation capability and the dynamic data processing.

In the case study, the decision regarding the use of database results from the need of performing the spatial cross of 360 files organized in an analysis grid. The repetition of the analysis, in addition to representing a waste of time and technical efforts, may result in errors from inattention ad usually means redundancy (unnecessary files creation) for each phase.

For the total information load into the database there were spent 24 consecutive hours in an automated process using the ETL tool. Such load is responsible for the input of 360 geographic tables into the database with a set of over 1 million surfaces related to the vegetation spots and over 280 millions of altimetry points. During the loading, information holding projection information were standardized.

To execute the spatial calculations to recover the maximum height of the vegetation spots there were spent approximately over 24 consecutive hours. At the end, the full process to the final calculation of the vegetation volume was performed in 48 hours.

For the analysis of the gain in using geographic database in comparison to market GIS software, it was performed the processing of one grid cell in the two environments. Using the software *ArcGIS*, a powerful application, and with the use of a computer with dedicated 2MB graphics card, it took 10 minutes to recover the maximum height per vegetation spot. The same operation performed in the database for the same grid cell took only 16 seconds.

It must be considered that, for the use of tools like the database, specialized knowledge is required from the responsible technician. On the other side, the database enables, above all, enhancing the work with data and processes, once the tool is designed for the development of systems which results are to be quick and constantly updated. Vegetation cover or buildings are not elements that changes daily, as the values of their volumes, but if a public manager needs to work with information of significant temporal change, it is, at last, the basis for the use of dynamic cartography.

The possibility of setting systems with the client/server architecture from the database enables the focus on the development of even more friendly interfaces and in a web environment, giving or not the database all the intelligence on managing and processing of spatial data.

The database processing on demand enables the creation of spatial tools capable of producing results in an immediate manner, as the example of simulation of alternative scenarios. Together with the *ETL* tool, the gain at organizing and using the data, whether punctual or from the big data, structured or not, are promising for the creation of spatial data infrastructures that can provide intakes inputs? for the most different territorial analysis and management tools.

AKNOWLEDGMENTS

Contribution to the Project “Geodesign and Parametric Modeling of Territorial Occupation: new resources of geo-technologies to landscape management of Pampulha Region, Belo Horizonte”, with the support of CNPq, Call MCTI/CNPQ/MEC/CAPES Nº 22/2014, Process: 471089/2014-1, to the project “Geodesign e Modelagem Paramétrica da Ocupação Territorial: Geoprocessamento para a proposição de um Plano Diretor da Paisagem para a região do Quadrilátero Ferrífero-MG” Process 401066/2016-9, Edital Universal 01/2016, and to the Project "Programa Pesquisador Mineiro – PPM IX", Process TEC – PPM – 00059-15.

We kindly thank PBH-Prodabel for the use of cadastral data.

REFERENCES

AGRAWAL, D., ABBADI, A. E., EMEKCI, F., METWALLY, A. Database management as a service: Challenges and opportunities. **Data Engineering, International Conference**, 2009. p.1709–1716.

AGRAWAL, D., DAS, S., ABBADI, A. E. Big data and cloud computing: New wine or just new bottles? **PVLDB**, 3(2), 2010, p.1647–1648.

ANDRIENKO, G., ANDRIENKO, N., KEIM, D., MACEACHREN, A. M., WROBEL, S. Challenging problems of geospatial visual analytics. **Journal of Visual Languages and Computing**. n.22(4), 2011, p. 251-256.

BAO, J., ZHENG, Y., WILKIE, D., MOKBEL, M. F. **Geoinformatica** 19(3), 2015, p. 525-565.

BORGES, J. L. **O Aleph**. Buenos Aires, Losada, 1949. 150 p.

CAMPAGNA, M. Sistemas de Suporte ao Planejamento (Planning Support Systems): Retrospectivas e Prospectivas. In.: MOURA, A. C. M. (ed.) **Tecnologias da Geoinformação para Representar e Planejar o Território Urbano**. Rio de Janeiro, Interciência, 2016. p. 217-252.

DAVIS JR., C. A., MORO, M. M., MATEVELI, G. V., MACHADO, N. G. Contribuições voluntárias: impactos potenciais dos cidadãos online and seus dispositivos móveis. In.: MOURA, A. C. M. (org.) **Tecnologias da Geoinformação para Representar e Planejar o Território Urbano.** Rio de Janeiro, Interciência, 2016. p. 23-34.

GOODCHILD, M. F. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. **International Journal of Spatial Data Infrastructures Research**, 2, 2007, p. 24-32.

KINGSTON, R. Public participation in local policy decision-making: the role of web-based mapping. **The Cartographic Journal**, n.44(2), 2007, 138-144.

MACEACHREN, A. M., GAHEGAN, M., PIKE, W., BREWER, I., LENGERICH, E., HARDISTRY, F. Geovisualization for knowledge construction and decision-support. **Computer Graphics & Applications**, n. 24(1), 2004, p. 13-17.

MAGUIRE, D. J. GeoWeb 2.0 and Volunteered GI. **Workshop on Volunteered Geographic Information**, Santa Barbara, California, USA. 2007.

MANOVICH, L. Visualização de data como uma nova abstração and anti-sublime. In: Leão, Lúcia (org.) **Derivas: cartografias do ciberespaço**. São Paulo: Annablume. 2004.

MCCORMICK, B. H. DE FANTI, T. A., BROWN, M. D. Visualization in Scientific Computing. **Computer Graphics**, ACM Siggraph, 21(6), 1987. 81 p.

MOURA, A. C. M. Landscape design or parameterization? Recent tendencies in geo-technologies for representing and planning urban territory. Bologna, **Disegnarecon**, 6(11), 2013. p. 3-10.

ONSRUD, H., CÂMARA, G., CAMPBELL, J., CHAKRAVARTHY, N. S. Public Commons of Geographic Data: research and development challenges. In.: EGENHOFE, M.; C. FREKSA C., MILLER, H. J. (ed). **GIScience 2004**. Berlin/ Heidelberg, Springer-Verlag, 2004, p. 223-238.

PENSA, S., MASALA, E., OGNEN, M. E se la forma seguisse la funzione? L'esplorazione della desiderabilità nella città di Skopje. Disegnarecon, 6(11), 2013.p. 141-148.

PENSA, S., MASALA, E. O papel da Visualização no Planejamento Urbano: uma abordagem a partir dos conceitos por trás da imagem espacial. In: MOURA, A. C. M. (org.). **Tecnologias da Geoinformação para Representar e Planejar o Território Urbano**. Rio de Janeiro, Interciência, 2016. p. 36-60.

ROCHA, N. A., SENA, Í. S., FONSECA, B. M., MOURA, A. C. M. Association between a spectral index and a landscape index for mapping and analysis of urban vegetation cover. In: INPUT 2016, 2016, Torino. **9th International Conference on Innovation in Urban and Regional Planning e-agorà/e-άγορά for the transition toward resilient communities**. Torino, Itália: SiTI, 2016. v.1. p.526 – 531.

STEINITZ, C. **A Framework for Geodesign: Changing Geography by Design**. Redlands, California: Esri Press. 2012. 224 p.

XAVIER-DA-SILVA, J. (Ed.). **Geoprocessamento para análise ambiental**. Rio de Janeiro, J. Xavier-da-Silva, 2001. 228 p.