

## AUTOMATIC RIVER NETWORK EXTRACTION FROM LIDAR DATA

E.N. Maderal<sup>a</sup>, N. Valcarcel<sup>a</sup>, J. Delgado<sup>a</sup>, C. Sevilla<sup>a</sup>, J.C. Ojeda<sup>a</sup>

<sup>a</sup> National Mapping Agency of Spain (IGN-ES) (enmaderal, nvalcarcel, jdhernandez, cssanchez, jcojeda)<sup>a</sup>@fomento.es

### Commission WG VIII/4 - Water Resources

**KEY WORDS:** Lidar, DTM, water resources, fundamental datasets, IGN-ES, INSPIRE.

### ABSTRACT

National Geographic Institute of Spain (IGN-ES) has launched a new production system for automatic river network extraction for the Geospatial Reference Information (GRI) within hydrography theme. The goal is to get an accurate and updated river network, automatically extracted as possible. For this, IGN-ES has full LiDAR coverage for the whole Spanish territory with a density of 0.5 points per square meter. To implement this work, it has been validated the technical feasibility, developed a methodology to automate each production phase: hydrological terrain models generation with 2 meter grid size and river network extraction combining hydrographic criteria (topographic network) and hydrological criteria (flow accumulation river network), and finally the production was launched. The key points of this work has been managing a big data environment, more than 160,000 Lidar data files, the infrastructure to store (up to 40 Tb between results and intermediate files), and process; using local virtualization and the Amazon Web Service (AWS), which allowed to obtain this automatic production within 6 months, it also has been important the software stability (TerraScan-TerraSolid, GlobalMapper-Blue Marble, FME-Safe, ArcGIS-Esri) and finally, the human resources managing. The results of this production has been an accurate automatic river network extraction for the whole country with a significant improvement for the altimetric component of the 3D linear vector.

This article presents the technical feasibility, the production methodology, the automatic river network extraction production and its advantages over traditional vector extraction systems.

### 1. INTRODUCTION

During the last 3 years, the National Geographic Institute of Spain (IGN-ES) has been working on the improvement of the production system of the hydrographic reference geoinformation, more specifically on the subject of the automatic capture of the stream vector network, which traditionally had been produced by manual methods using photogrammetric restitution techniques. The main reason for this change is the current need for high resolution 3-dimensional geoinformation in the hydrographic theme, mainly promoted by the completion of the LiDAR coverage for the whole Spanish territory, with an average density of 0.5 pulses/m<sup>2</sup>. Additionally, the current state of the art in this research provides some tools and very helpful previous experiences of automatic stream network extraction. The main challenges for IGN-ES is on one hand the size of the country, with 504.000 km<sup>2</sup>, and in the other hand, the need of updated high resolution hydrographic geoinformation, keeping the current planimetric resolution, and with significant improvement in the altimetric resolution, according to the possibilities provided by the Lidar data. This improved geoinformation must also be produced fully coherent with the altimetric data obtained from the LiDAR data, high resolution tridimensional stream network (according to the existing Lidar coverage), hydrological and flow direction DTM able to be integrated in hydrological tools for analysis, planning and management, delimitation of physical contours of drainage basins, codification of hydrographic stream network, etc. Several studies have concluded that LiDAR derived DEMs have suitable horizontal and vertical resolutions for mapping surface water flows (James, 2010).

In order to start with this line of work, at the end of 2013 and during 2014 a first methodological approach is developed using the existing DTM with 5 m grid size (DTM05), derived from LiDAR data; the main reason is that IGN-ES had already generated and reviewed this DTM, and its size seemed to be a first compromise between accuracy and volume of information for geoprocessing. After determining the optimum area unit to divide the territory for processing, a number of corrections to the mosaics were built and corrected in several steps, mainly to remove obstacles from transport network and then a hydrological calculation was done, based on accumulation, using an average basin size based on the contrast with the existing cartography (number of river streams and channel heads). The results using this first approach showed that additional large manual editing was needed, and although the Z accuracy was good and consistent, planimetric results did not improve the current existing data available in the IGN-ES (national topographic base with accuracy in XY < 5 meters 99% level of confidence).

After this first stage, IGN-ES modifies the initial methodological strategy, generating a new DTM with 2 m grid size (DTM02) from the same PNOA LiDAR data, analysing the viability of this approach with several test sites, selected according to specific hydrological and topographic characteristics, to check the results of the applied methodology. Positional accuracy of automatic results are measured objectively and time spent during the different productive steps are also evaluated, with satisfactory conclusions, so IGN-ES decides to apply the validated procedures to produce the river network for whole Spanish territory.

To achieve this goal, a big data processing infrastructure is implemented, with both local and cloud servers, in order to finish the automatic extraction of the stream network during the first semester of 2016.

Several tests were made in different steps of the works to select the best software and tools to be used in the production chain (DTM generation, mosaic production and hydrological DTM geoprocessing for deriving the vector stream network, considering the advantages and disadvantages of its integration in the automatic production, and more specifically, evaluating the quality of the results in each case.

## 2. GEOSPATIAL REFERENCE DATA PRODUCTION OF HYDROGRAPHY IN IGN-ES

Continuous users' needs linked to technological advances, current legislation and global context in terms of geographic information have led to promote a change of the productive systems. In 2014, the National Geographic Institute of Spain (IGN) launched a new production system of Geographic Reference Information (GRI) of high resolution and INSPIRE compliant (2007/2/CE Directive), in line with the United Nations decisions about Geospatial Reference Information Management and also fitting user requirements at national, European and global level.

The first step was to collect and analyse user requirements and needs by means of meetings with the relevant stakeholders and users: hydrography managers, data producers, data providers and relevant users. It was also necessary to analyse their data and product specifications in order to fix the GRI-HI data content and structure. In the case of IGR of hydrography, there is a need of providing an accurate and efficient information for mapping, a topological network for all kinds of applications and hydrological simulations, and also, digital hydrological models and flow direction and accumulation models (Heine, 2004). The second step was to analyse the existing legislation, the INSPIRE specifications in hydrography and other documents in parallel thematic areas together with the common documentation of conceptual models and INSPIRE Network. The objective was to develop the IGR-HI data product specifications including the feature catalogue. The application schema is INSPIRE compliant and is based in the two use cases developed in the INSPIRE specifications: Physical Waters and Network model.

The next process was to develop a methodology to ensure the most accurate, objective, interoperable and updated GRI hydrography, obtained as automated as possible.

The data set resulting (IGR hydrography) will be published through web services compliant with the INSPIRE Implementing Rules. Specifically.

## 3. SPANISH LIDAR DATA PROJECT

The Spanish National Program for Land Observation (PNOT) addresses the need of information about the territory, and its evolution over time. It provides the basic geographic information, needed for a sustainable management of our environment, agriculture, infrastructures, emergencies and security, among other activities. PNOT is composed by three big National Programs: PNOA (Plan Nacional de Ortofotografía Aérea), the National Program for Aerial Orthophoto; PNT, the National Remote Sensing Program, which coordinates continuous image coverages from Earth Observation satellites, and SIOSE, Spanish Land Cover and Land Use Information

System, that makes use of the data acquired by the other two programs, to extract and describe land cover & use features.

PNOA program, started in 2004, is a collaborative program between Spanish Public Administrations. Initially, digital Elevation Models (DTMs) needed for orthophoto production were obtained by automatic image correlation and stereoscopic edition, but due to the increasing need for more accurate altimetric resolution, PNOA LiDAR subprogram started in 2009 with the objective of covering the whole Spanish territory with aerial LiDAR data in order to get a better accuracy and detail in DTMs (Digital Terrain Models, Digital Surface Models, etc.) and also for many other uses, such as flood areas mapping, vegetation monitoring, etc. Technical specifications of PNOA LiDAR flights set an average density of 0.5 pulses/m<sup>2</sup> and an average updating period of six years.

In 2015, the first coverage was finished and the second coverage has already started, with improved specifications, providing LAS files that contain a set of fields for each point as: coordinates in ETRS89 GRS, ellipsoidal heights, intensity, scan angle, return number, etc., dividing the Spanish territory 2 x 2 km tiles, with an average LAS file size of 200 MB, containing between 5 and 10 million points, depending on the area. This raw data is then post-processed in several ways: convert ellipsoidal heights into orthometric heights using the official Spanish Geoid model, automatic classification of point clouds, manual edition of classification, extraction of DTM from ground class, etc. PNOA LiDAR products are available for free download and use.

## 4. METHODOLOGY FOR AUTOMATIC RIVER NETWORK EXTRACTION FROM LIDAR DATA

### 4.1 Technical Criteria

The technical criteria for this work are summarized in the following points:

- Geographic feature to be captured: natural river stream network (Surface water)
- Data source: automatic classification of LiDAR data (0,5,p/m<sup>2</sup>)
- Digital Terrain Model: automatically generated, with 2 m grid size.
- Hydrological processing: D8 algorithm (O'Callaghan y Mark, 1984), (Olaya, 2004).
- Positional quality: around 1 meter in XY and better than 50 cm in Z ; it must be continuous and topologically and hydrologically correct in 3D
- Degree of automation: over 60%
- Required products: hydrological digital terrain model, and direction model, together with the 3D river network.
- Verification of results: field works are not considered in this methodological development.

It is also necessary to summarise here the characteristics of the reference data used in this work:

- Aerial images of the national aerial orthophoto program (PNOA): reference date=2012-2014, GSD=50 cm, mseXY<=1meter, mseZ<=1.5meters
- LiDAR: reference date=2009-2014, resolution=0.5points per square meter, maximum error<=0.4 m (95%), mseXY<=0.3 meter (nadir), mseZ<=0.2 m
- Spanish cartography: reference data=1994-2013, accuracyXY<=5 meters, accuracyZ<=10 meters, mseXY=3 meters, mseZ=5 meters.

## 4.2 Viability Analysis

In order to establish a methodology and evaluate the feasibility of whether to proceed to the automatic extraction of river network throughout Spain, this lines of work were established:

**MDT generation:** the hydrography datasets calculus comes from digital terrain models (DTM) obtained from Lidar data. The density of LiDAR used allows different range of DTM pixel resolutions depending of the expected accuracy for derived information calculated over it. This subject motivated to work with DTM resolution of 5m and 2m, and test different commercial tools.

**Hydrological procedures:** the main question for viability was to assess which automated processes would be required to obtain a hydrologically corrected DTM and its derived vector network . For this, several singular cases were considered and different types of processes were developed: for surface water modelling, for integrate buildings as obstacles or to remove different types of obstacles (transport network, no soil LiDAR misclassification, lift areas). For these developments ArcGIS software were selected.

**Procedure for the evaluation of results:** in order to provide the more objective conclusions. A visual comparison by a photo interpreter provide a first overview of 3D networks comparison but the main element was the positional XYZ control design.. This procedure was defined according to the following steps:

1. Determination of the point population for the study and representative statistical sample: based on a sample of selected vertex from existing cartography
2. Capture of control points: obtained in a stereoscopic system and used as ground control points, based on identification of sample points.
3. Automatic accuracy XYZ measurement: based on a proximity tool to find the homologous network point to the control point. Average distance and average difference in the Z dimension are calculated in order to evaluate the quality of the results.

To evaluate results, several small areas (from 15 to 163 Ha) were selected considering as representatives for their singular hydrological and morphological characteristics. These sites were located in the Guadalquivir basin district (Figure 1).

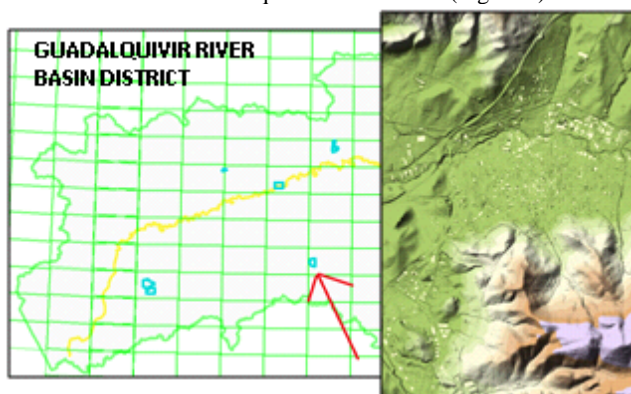


Figure 1. Urban test area (Lucena urban area in Córdoba, Spain).

For each test areas, MDT05 and MDT02 were produced with automatic network calculation (including MDT corrections, and hydrological calculation with an average accumulated threshold of 15 ha), processing time were registered, and last, results were analysed and compared with the positional control XYZ procedure.

Results of the viability analysis are summarized in the following table:

Dataset (hydro network)	XY error evaluated	Z error estimate d	Estimated processing time for Spain (1 computer)
IGN-ES cartography 1:25.000	2-3m	< 5 m	
2m DTM (ARNE02)	2-3m	< 0.5 m	29,9 months
5m DTM, (ARNE05)	3-4 m	< 0.5 m	12,8 months

Table 1. Results from viability analysis.

Resolution of 2m was selected for the Automatic River Network Extraction methodology (ARNE) due to their optimal DTM appearance and sufficient precise results of hydrographic data derived from it, although the solution ARNE05 was seriously considered due to the fact that meant a reduction of 50% of the time for the whole processing (it is not include DTM time processing because already exist), but ARNE02 time consuming was planned to be solved using more servers for processing, to finish the automatic extraction of the network on planned time. So the final selected option was ARNE02, which provides similar results in XY accuracy to the existing cartography, but has the benefit of a big increase of the vertical accuracy, and ensuring the coherence between the LiDAR DTMs

It is necessary to stress that the quality control procedure does not provide absolute measurements of the vertical error in Z (is estimated as show in Table 1), so it had to be complemented with a few but very accurate control points coming from precise GPS field points.

Also it was concluded that the best results for the DTM generation and subsequent network calculation were provided by the TerraScan software.

## 4.3 Methodology for the production of automatic river network extraction

The aim is to establish the steps and implement tools to automate mass production for the whole territory.

The production of automatic river network extraction workflow has been designed as shown in Figure 2.

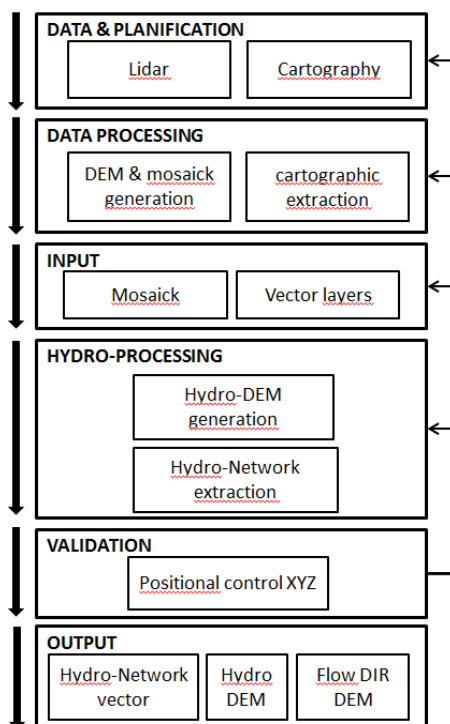


Figure 2. Flux diagram of the methodology for automatic river network extraction from LiDAR data

#### 4.3.1 Data and planning

One of the key points is to work with large amounts of data, more than 160,000 LiDAR files, which generate an equal number of MDT files.

The hydrographic calculation is organized in production units to be able to process the 2 meter grid size DTM, corresponding with planned river basins areas (sub-zones oriented to computer processing), these sub-zones are aggregated into zones (oriented to human resources distribution) and these zones compound the whole river basin district from Water General Directorate (oriented to project management); thus three levels of hydrological and management organization were established (see Figure 3).



Figure 3. Planned zoning for the Guadalquivir river basin district: 5 zones and 24 sub-zones.

#### 4.3.2 Data processing

##### 4.3.2.1 DTM and mosaicing generation

Steps in the DTM generation and mosaicking are:

- DTM generation for isolated Lidar files
- Selection DTM files from planned areas (sub-zones)
- Mosaicking all DTM files for work planned
- Correction and verification actions

##### DTM generation

DTM generation is driven by the previous Lidar automatic classification, using ground class (ASPRS standard Lidar class n°2) and filling the holes by interpolation to produce the master DTM to work, and ancillary DTMs for only ground (ASPRS standard Lidar class n°2) and buildings (ASPRS standard Lidar class n°6) to help in next production phases.

DTM generation by isolated LAS files can provoke discontinuities in the terrain between neighbour files. To avoid this problem, during each DTM generation are used also Lidar points from neighbour LAS files generating over extended DTMs. A useful over extension is 40m by side. Later the resulting DTM is cut by its actual extension.

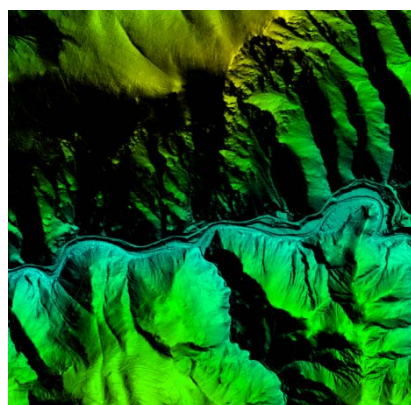


Figure 4. Example of generated DTM file of 2m grid resolution.

##### Selection DTM files from planned sub-zones

To generate a unique mosaic for each sub-zone, firstly it is necessary to select the adequate individual DTM files that are inside to each sub-zone. This selection is automatic by geographic comparisons of sub-zones and DTM extensions including a sub-zone overlap of 1000 m. to guarantee selection of required DTM files. Safe software FME was used to organize the mosaicing production.

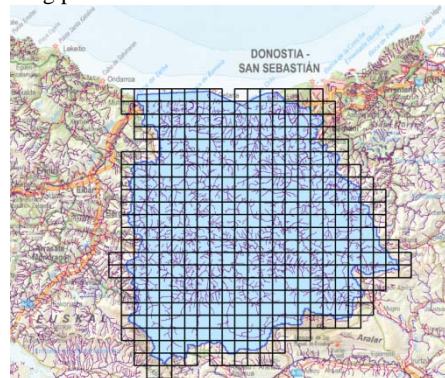


Figure 5. Example DTM files selection (in black) per one working unit (in blue).

### Mosaicking DTM files

All DTM selected files per sub-zone generate a unique and continuous DTM file per entire working unit. Small gaps of data and irregularities are also corrected in this step. Blue Marble Geographic's Global Mapper software was used for mosaicking.

The selected format for DTM files was ESRI ASCII GRID for its interoperability with many platforms. Terrasolid TerraScan software working under MicroStation environment was used to generate DTM files from Lidar data.

### Correction and verification actions

Some actions are required to provide a useful DTM for hydrographic calculus. Safe software FME was used for correcting and verification actions like omission or commission files, DTM gaps (i.e. water, lost signal points, scant density, etc.), these areas are recuperated from other DTM origin and put together with Lidar DTM, and mosaicked them together. In coastal areas it is also recommended to cut Lidar DTM by the coastline, and finally, after mosaicking it is also recommended to contrast the possible offset of pixels. Pixels can experiment a displacement after commented steps (i.e. 1/2 of pixel resolution) and vary the georeferentiation of the whole dataset. Automatic verification of pixel displacement was carried out to avoid this problem.

#### 4.3.2.2 Cartographic extraction

For preparing the cartographic layers an ETL tool was developed using FME software, which automate the extraction of the original data, and the load in the projects. Vector layers are the rivers and their headwaters, water bodies (reservoirs, rivers and lakes), dams, transport network (roads and railways), bridges-like constructions, sinks, national boundaries and cadastral blocks.

#### 4.3.3 Hydro processing

This part of the methodology are based in ESRI tools and taking into account issues such file management and software performance, it was decided first to import all the initial files for each sub-zone to an ESRI geodatabase.

##### 4.3.3.1 Hydro-DTM generation

This phase is designed to gradually refined the DTM to be hydrologically correct. Water surfaces are modelled, DTM obstacles are corrected with several techniques and adjustments are made in filling areas. Methodology is based in the calculation of the derived network, at the end of each process, which serves as a starting point for the next one. Network is automatically obtained from headwaters, and following the flow direction model applying cost-path processing. Applied processes are explained below.

### Hydrological surfaces correction

The answer provided by LiDAR in those areas with significant surface covered by water is practically zero, so the DTM calculated from LiDAR in these areas has a lot of holes or unreliable information. On the other hand, the axis of the hydrographic network in these areas is going to be fictitious. To identify the areas to be correct, some pre-existing cartographic elements with surface geometry are used: reservoirs, rivers and lagoons.

The aim of this process is the hydrological modeling of the surface water bodies, correcting the initial DTM. The following

figure shows what is achieved in areas with surface water bodies.

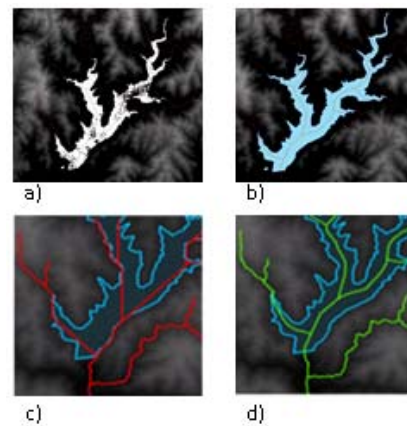


Figure 6. Surface hydrological modelling of water bodies: a) original DTM obtained from LiDAR data, b) water body and fictitious river network from cartography, c) automatic feature extracción without modelling only filling gaps, d) automatic feature extraction with surface hydrological modelling

Figure 7 resumes the procedure: get the DTM information out of the water body (PointElevation) using a exterior buffer from the cartographic water body, and together with the fictitious axe and elevation of points of the axe (Stream&PointElevation), those point are incorporated to the hydrologic interpolation to model the DTM. To determine the elevation points from the fictitious axe, minimum and maximum elevation are obtained, and then the rest of the nodes are interpolate over the axe, with a constant decreasing parameter, to force interpolation within the axe. Dams is also integrated, to help the modeling exercise. This process is done using the ARGIS TopoToRaster tool based on ANUDTM (Hutchinson, 2011)), which is a interpolation method specifically designed for creating hydrological DTM.

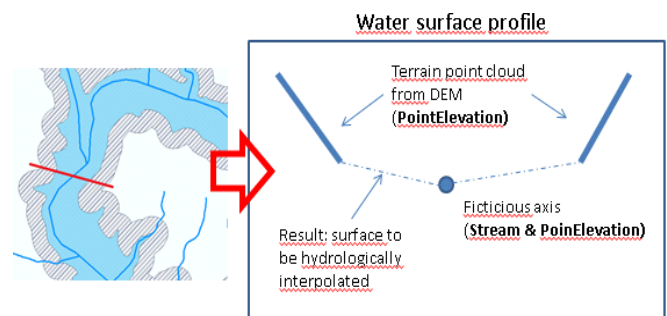


Figure 7. Hydrological modelling of water surface. Top image: an area of a water surface from a reservoir, bottom image shows the indicated profile where both sides (PointElevation from DTM) comes from an outside buffer made over the water surface (grey zone at top image)

### DEM buildings integration

The objective is to integrate buildings in the DTM to act as obstacles to the path of the river. This integration is made from two sources: the automatic classification of LiDAR points classified as buildings (ASPRS Lidar class standard No. 6) and the urban blocks of the Cadastral cartography.

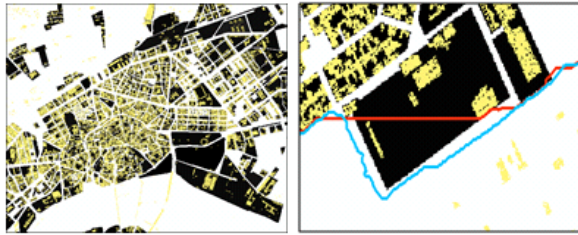


Figure 8. DEM buildings integration (Cadastré blocks in black and LiDAR buildings in yellow). River extraction only with LiDAR data (red) and with cadastre data integration (blue)

#### Obstacles correction 1 (bridges from cartography)

The objective is to identify intersections with the transport network, to correct the related obstacles to the water flow, using points before and after the intersection, by interpolating the drainage



Figure 9. Obstacles correction 1 (bridges from cartography)

#### Obstacles correction 2 (bridges from LiDAR)

The objective is to identify intersections with the transport network due to a misclassification of LiDAR ground points, in a similar way as the previous case, but it is only done when deviation between the network and cartography is higher than 10m.

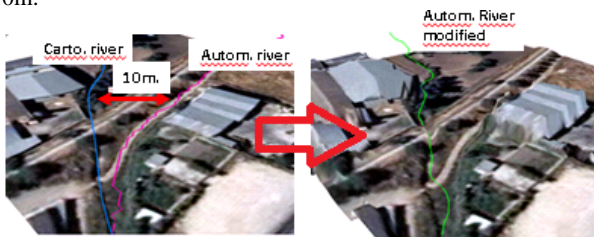


Figure 10. Obstacles correction 2 (bridges from LiDAR)

#### Lift correction

The objective of this process is to correct again new potential obstacles due to errors in the ground classification of LiDAR points, for instance when vegetation is assigned to ground class, generating a non existing obstacle as little lift in the DTM, and provoking lateral deviation in the path of the calculated stream.

Lift areas are detected (by deviation from cartography), maximum elevation level to overcome is determined in each one, and then elevation value is reduced in every cell exceeding this level.

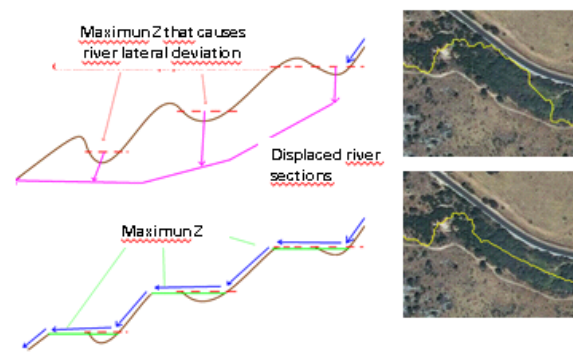


Figure 11. Lift areas correction. Top images shows original situation. Bottom images the result of the correction.

#### Cartographic burning-in correction

In this phase, water stream is forced to follow the trace of the existing hydrographic cartography. Sections to be corrected are those in which a deviation higher than 10m exists; for these sections, elevation values of cells are reduced (burning-id) following the trace of the existing cartography. This procedure is also iterative (2 iterations are applied at least). ESRI Arc Hydro tools contain options to burn in stream networks to raster elevation data by lowering the elevation of the stream network cells, changing elevation values in burned locations (Poppenga, 2013).

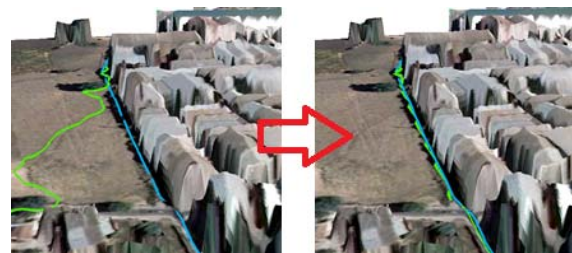


Figure 12. Cartographic burning-in correction.

#### Fill areas readjustment correction

Finally a small readjustment is made in fill areas to refine the network path to the real terrain. In this process, automatically filled zones greater than 100m<sup>2</sup> and near to the water course are identified. In order to obtain the drainage trace in these zones, difference between the filled DTM and the original DTM is evaluated, and a new DTM is obtained with the inverse relief in those zones; then change slope lines are generated, corresponding with the corrected drainage trace, and on this trace, elevation value is reduced in 0,1 m in the matching cells with the network to be readjusted.

#### Results of hydrological DTM generation(hDTM)

Final results for this phase are: hydrologically corrected DTM (hDTM), its associated flow direction model, and the set of derived stream of each correction process.

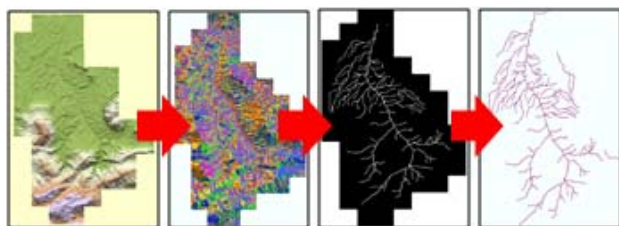


Figure 13. Phases and results of hydro-DTM generation. First hydrological generated DTM, next the flow direction raster model, next the least-cost path stream from river headers and finally the derivate stream extraction.

#### 4.3.3.2 Hydro-Network generation

The objective of this phase is to obtain the automatic 3D network, together with the catchments and downstream junctions, which later will allow joining all the production units and composing the drainage basic.

#### Flow accumulation raster and statistics for stream definition

Flow accumulation raster is calculated for each sub-zone, using a 64 bits pixel size (to load big numerical values) and from this raster; those accumulation values homologous to the headwater nodes are identified. Average value is used as stream threshold to initiate the stream by accumulation in the next process. This value will be variable depending on the territory.

#### 2D Hydro-network generation

This process calculates the whole 2D network for each sub-zone, generating the following layers.

- 2D network extraction: resulting of the combination of the derived stream from the cost-path calculation using the cartographic headwaters, and accumulation network by using the stream threshold. This combination is made in raster format and them is vectorised (stream to feature).
- Catchment definition: physical drainage basins generation.
- Downstream junctions: each drainage basin has an downstream junction layer.
- Section attributes: each section is characterized with the typology of network (derived, accumulated), slope, length and type of correction made in the DTM.

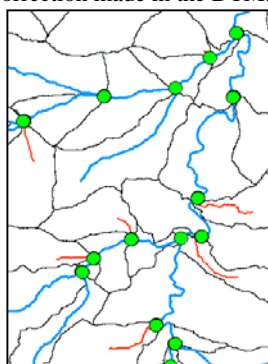


Figure 14. 2D network generation. Derived network appear in blue, accumulation network in red, catchments in black and downstream junctions for each catchment in green.

#### 2D Watershed processing

So far, all the work has been made in the DTM spatial scope. From here, calculation will be made in the spatial scope of the drainage basin; for this, previously a selection of the calculated downstream junctions is performed, so that if the sub-zone is a header, then only downstream junctions must be selected but if the sub-zone is intermediate, then a upstream junction related

with the precedent basin in hydrological sense must be also selected. This operation is the only manual input throughout the full procedure and requires some skill and practice for proper selection, because conditions the quality of the hydrologic union between production areas.

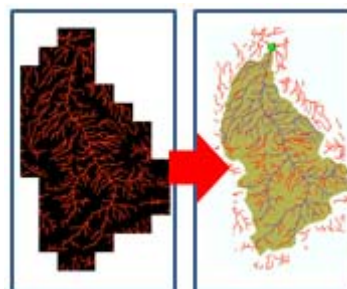


Figure 15. Watershed processing. Right, the extracted network for the DTM, left the result for the calculated catchment.

#### 3D Hydro-network generation

Last process of the automatic extraction consists on performing a simplification of the number of vertex of the network, to facilitate its subsequent manual edition, and including the vertical dimension using the hDTM, finally obtaining the final 3D hydro-network.

#### 4.3.4 Validation of results: automatic positional control XYZ

For this validation, the procedure described in point 4.2 was used. But in this case, instead of manual capture of control points, it was decided to use the point sample from the cartographic network, considering the fact that its positional accuracy in XYZ was well known (2 to 3 m). These points are the considered as control points, and so the automatically generated network is tested very efficiently. Z coordinated was also tested but only to detect gross errors.

### 5. PRODUCTION OF AUTOMATIC RIVER NETWORK EXTRACTION

#### 5.1 Planning and resources

The whole area to be processed was of approximately 600,000 km<sup>2</sup> (504,000 km<sup>2</sup> plus a percentage of overlap between production areas). The length of river network to be extracted was more than 654,000 km (length of the current network mapping at 1: 25,000). Using manual photogrammetric capture techniques, production time per kilometer of network it is estimated at about 3 to 4 minutes. In this case, 30 months were estimated to produce automatic network with a team dedicated to hydrological processes with ESRI tools. This part is about 65% of the production; the rest is divided between time planning and organization, and data preparation and load, especially DTM02 and mosaics. To produce the entire network within a period of 4-6 months is necessary to have a large number of geo-processing machines, and also technicians coordinating and monitoring the processes.

In summary, 15 people have been involved with full or part time. A intranet dedicated infrastructure was been configured, consisting of a storage server with 40TB plus and its corresponding back-up, and two virtualization environments Xen Server 6.2 SP1 with 22 machines prepared for geo-processes (Windows Server 4 cores, 32 Gb RAM, 200Gb HD). Processing capacity was also complemented with Amazon Web Services providing other ten geo-processing machines in the cloud (type c4, Windows Server 4-core, 7.5 Gb RAM, 200Gb HD). Works have been done using 4 licenses for TerraScan, 2

Licensing Global Mapper, some 40 dedicated ArcGIS licenses 10.3.1. 32 bits (Spatial Analyst and 3D Analyst).

## 5.2 Production

In Spain there are 24 River basin districts or demarcations (Figure 16), which have been used as production blocks. The planned units (sub-zones) for producing each demarcation were calculated taking into account that the surface should be in the range of 1,500-2,000 km<sup>2</sup> to balance properly processing times with the available resources. In the first defined sub-zones, with larger areas (some greater than 3,000km<sup>2</sup>), it was necessary to subdivide some of the sub-zones because of the limitations of the hardware-software environment for massive processing. The number of sub-zones for all Spain is approximately 300 and in each of these sub-zones, ArcGIS processes have lasted an average time of six days.



Figure 16. River basin district in Spain: areas of production management (at bottom right is Canarias in fictitious location)

Massive production started in November 2015, after the stress test in the Guadalquivir demarcation; processes were fully optimized in January 2016. Despite this, in early May 2016 the automatic production is complete, investing a total of 6 months. The following table shows processing times in the most significant phases of work.

FASE	Processing units	Execution time(months)
DTM02	160.000 (files)	3,5
Mosaic	300 (sub-zones)	2,3
Hydro-Processing	300 (sub-zones)	6

Table 2. Processing cost time.

Results for each sub-zone have been validated by the positional quality control at the end of each phase, being the expected: average discrepancy between 2 and 3 meters in XY. In some cases discrepancies were observed but were due to errors in the processes that were solved.

## 6. RESULTS AND CONCLUSIONS

A new methodology has been established that take full advantage of the existing LiDAR data available in IGN-ES. The automation achieved with this methodology achieves approximately a 70% success, and even considering the manual edition of the remaining 30%, this methodology represents an estimated savings of 60% of the capture time compared to manual capture techniques. Automatic extraction provides an objective and homogeneous result and provides a significant improvement in altimetric accuracy over manual capture, done

by applying criteria based on individual knowledge, skill and experience of the human operator. However, it is noteworthy that without the pre-existing cartography, it would not have been possible to reach this level of automation and quality of results.

Viability analysis has allowed justifying and taking the decision to produce according to this new methodology, highlighting the savings in time and the significant improvement in accuracy in the vertical dimension.

This methodology provides vector network, hDTM and flow direction model but within the same work there are some other intermediate products that are interesting for users, such as the drainage basin boundaries (limits of physical basins), necessary to define the accurate limits of the demarcations. This production was done according to the defined deadlines and has been made possible by the effective involvement of resources, both human and material. Also it has been a very positive experience with cloud processing services due to its flexibility, secure storage and high capacity of processing.

Finally, this work is considered as a first version, the beginning of a shift of the production system, using mainly geospatial data (altimetry, mapping). Next versions have to go a step further, to include climate and hydro-geological parameters and data that complement hydrological modeling.

## REFERENCES

- Heine, R.A., Lant, C.L., and Sengupta, R.R. 2004a. Development and comparison of approaches for automated mapping of stream channel networks. *Annals of the Association of American Geographers* 94(3), pp.477–490.
- Hutchinson MF, 2011. ANUDEM version 5.3, USER GUIDE, [http://fennerschool.anu.edu.au/files/useDEM53\\_pdf\\_16552.pdf](http://fennerschool.anu.edu.au/files/useDEM53_pdf_16552.pdf).
- James, L.A. and K. Hunt, 2010a. The LiDAR-Side of Headwater Streams: Mapping Channel Networks with High-Resolution Topographic Data. *Southeastern Geographer* 50(4), pp. 23-539.
- O'Callaghan, J. F. y Mark D.M, 1984a. *The extraction of drainage networks from digital elevation data. ComputerVision, Graphics and Image Processing* 28, pp. 323–44
- Olaya, V., 2004b. *Hidrología Computacional y Modelos Digitales del Terreno —Teoría, práctica y filosofía de una nueva forma de análisis hidrológico*, pp. 62-64. <http://heart.sf.net/textos.html>
- Poppenga, S. K., Gesch, D. B. and Worstell, B. B. 2013a. Hydrography Change Detection: The Usefulness of Surface Channels Derived From LiDAR DEMs for Updating Mapped Hydrography. *JAWRA Journal of the American Water Resources Association*, 49, pp. 371–389.